

Improving Credibility of tweets for effective crisis management Using Twitter analysis of Chennai Flood – A Case study

¹Ms. Arathi .E , ²Dr. D. Pugazhenthii , ³Mr.Abinandan Arulprakash, ⁴Arihant Devapal

¹Dept. of Computer Science, SRM University, Kanchipuram, India

^{2,3,4}PG & Research Department of Computer Science, Quaid-e-Millath College for Women, Chennai, India

ABSTRACT :- Nowadays social media plays an important role in information propagation among the people. In the crisis, twitter is the commonly used social media that brings millions of users together. Social media service plays a significant role in raising awareness, social mention of relentless social workers and coordinating relief work in the time of crisis. This article explores the role of retweet in emergency situations. We analyze how the information had propagated, and the actions were taken based on the source. The feature retweet was considered as a measure of analysis to enhance the reliability of the spread information. Twitter analysis of the flood in Chennai due to heavy torrential rain for two days in December 2015 is taken as the study.

Keywords: Respondent, Retweet, Crisis, Tweets.

I. INTRODUCTION

A Natural disaster is an event that alters ecological environment between human and nature. In this era of global warming, natural disasters are linked to extreme weather conditions which affect millions of people around the world. In 2004, a massive tsunami killed more than 200000 people in South East Asia. In 2014, Nepal was devastated by an earthquake which affected around 90000 people. Communication plays a vital role during the time of crisis and natural disasters. Especially social media have become an important channel of communication during the crisis, and this compliments the information propagation by the traditional media. In 2013 [1] twitter offered a new service called twitter alerts to prioritize information from trusted sources during emergencies. Social media provides communication between affected area and the rest of the world. They provide rapid detection of natural disaster, situational awareness as well as relief coordination. Many individuals engage themselves as active producers and disseminator of information.

The social media messages during emergency convey timely actionable first-hand information. [2] Processing social media texts involves filtering, classifying, ranking, aggregating, extracting and summarizing using traditional information processing methods. Fig -1 shows the phases of disaster management life cycle are preparation, impact, response and recovery. Social media communications are the valid and useful source of information, among many stakeholders like the general public, organizations, and response agencies to take needed actions in disaster situations.

Critical tasks that can be implemented by social media:

- Alert citizens in areas likely to be affected by a catastrophe;
- massively spread the real-time information both to the affected areas and to the people willing to help.
- Receive real-time data from affected areas;
- Mobilize and coordinate the immediate relief efforts and
- Optimize recovery activities.

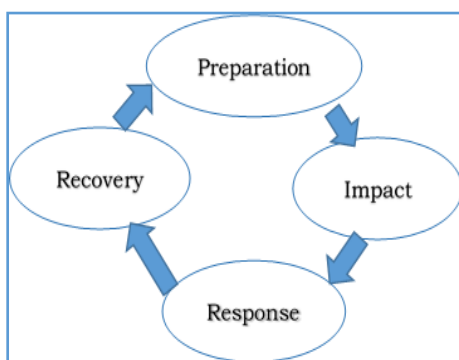


Figure. 1. Life cycle of disaster management

Twitter is a social media network where the user registers him to share his thought or ideas using small text messages of the maximum character of 140 character of length. These short messages called tweets are the social interaction between users. This interaction creates a mesh of networks as twitter network. The user can post their tweets or repost tweets of another user. The reposted tweets are retweets.

User can refer the subject they tweet by # (hashtag) for example # <subject>.

To particularly specify a user: keyword @<user name>.

The user can follow another user to get updated info about the particular topic spread information. [3] Twitter is used for constructing collective wisdom which can prompt accurate and precise information which shall be utilized in the crisis management. The hashtags are created for a group or an individual for the emergency situations. It can update the other members of the group about the conditions, feedbacks, and crowd sourced information of disaster. As the part of minimizing the effect of the disaster, the enhancement of scientific and intelligent disaster management system through the integrated analysis of unstructured data like social media is much needed. Also, the credibility of the information propagation shall be visualized

The remaining work is organized as follows. Section 2 present background. Section 3 presents the analysis of tweets in a disaster scenario. Section 4 presents the twitter analysis of the Chennai flood happened in Dec 2015 as a case study. Finally, Section 6 has Conclusion and Future work.

2. BACKGROUND

During the disaster, the role of twitter is very prominent. The author emphasizes twitter analysis to model the activity and importance of user. [5] Twitter official API enables data collection. Its functionalities, allow developers to search for, and store user profile information, user connection, tweets, and retweets. Location can be found if geo latitudes are attached with tweets. Twitter API is based on the REST architecture (Representational State Transfer), defines the resource and method for accessing real-time data. OAuth is an open standard which provides secured delegated access from the server. API has limitations as

- (1) the number of request per time window .
- (2) The window time limit has 15 minutes.
- (3) The search is limited to 180 requests.

The virtualization of the data is done by many tools available now. Graph analysis shows the maximum tweets, peak hours of twitter user and also analytics about user links and role within the particular group.

Takeshi et al describe the real-time nature of twitter which is used in solving natural disasters. This social media aid us in event detection by the user as virtual sensors and the tweets as sensory information. [6] To detect the event, semantic analysis is done on tweets. The tweets are classified as the positive or negative class using support vector machine algorithm. The event location estimation can be done using temporal model, spatial model or particle filters. Location estimation is done using particle filters. The application notifies people about the earthquake.

3. ANALYSIS OF TWITTER IN EARTH QUAKE SCENARIO

Marcelo et al explore the behavior of twitter in an emergency situation [4]. The analysis is about the trustworthiness of the information passed through the tweets. Their study concluded that it is possible to find the source of information, based on the source, the information can be categorized as trusted one or rumor. To determine the trustworthiness of information seven confirmed truth and seven false news was listed for the tsunami occurred in the location of Iloca and Duao. The tweets were filtered by the place of its origin, i.e., in and around the disaster area. The tweets were compared with the list of truth and rumor, then classified, labeled as four categories affirms, denied, queries and unrelated. The retweets of the tweet automatically got the same label. The result shows the propagation of information and needs to check the reliability of the tweets.

In recent times when a disaster occurs the victims and people in the scenario gives us the ground information about the crisis. For decades in the case of emergency, relief or rescue operations were carried out by the specified organization with a centralized command structure with standard operating procedures. It was usually slow when compared with rescue operations based on online social media information. In twitter the request for help, the volunteers, people willing to donate connects easily, the victims were reached quickly, and rescue operations are efficiently organized. In recent times, the involvement of the public in an emergency scenario is the reason for the information overloads in twitter during the disaster. Loads of information after cleansing, classifying the related data are then analyzed. Polluted, false rumor and confirmed truth have to be identified and classified to avoid delay in helping the victims in the critical situation. The unreliable and false news have to be removed for effective crisis management to enhance the trustworthiness of tweet.

IV. CASE STUDY

4.1 Role of Twitter during Flood in Chennai in December 2015.

Chennai is being pounded by the worst rains in nearly a century, with a fresh bout of heavy downpour which three days were leaving most parts of the city under water. Several areas of the state of Tamil Nadu are also being lashed by torrential rains, with the number of deaths rising to 188. Chennai floods unleashed the power of social media, especially Twitter. For three days, when all else failed this Social Media platform connected the flood-affected with their saviors. Twitter was a life saver, starting December 1, it started buzzing with volunteers offering help, and the marooned tweeting for help.

A concerned commuter tweeted the effect of flooding of railway tracks and stations shown in Fig 2a. The user passed a valuable information to the people to avoid train service. Making other situation awareness of disaster.

Offers of help: There were good Samaritans who immediately took to twitter to rescue the marooned. Figure 2b shows the tweets that were put out offering help. The social media site was used by worried friends and relatives to get information about their loved ones. Fig 2c shows the concerned relative asking help for finding their family in the affected area.

Figure 2d shows the people using twitter as media to appeals for food, water, material, emergency supplies. Drinking water was a huge problem, reading need for water tweets, many volunteered free water to the needy. Many tweets and retweets were about medical supplies.

Twitter became a part of the choice to seek help. Fig 2e is the twitter by a volunteer to boost the spirit of the people of Chennai. Volunteers used twitter to find transport, more hands, food, medicine, water, etc. Several images and videos captured how people stood shoulder by shoulder and helped each other.

Retweets confusion: Several good hearted citizens retweeted each tweet for help. Within few hours, the twitter was overloaded with the tweets and retweets about Chennai rain and flood information. The overload of information hampered the relief operations. Many volunteers tweeted #verified tag to avoid confusion in relief work. Due to retweets of outdated help needed texts, volunteers would visit areas where help had already was provided. The old and false retweets created confusion and wasted the effort and time of the volunteers and delayed the relief operations. Request on twitter to remove all those tweets that had already been broadcasted to all twitter users who are helping for cause as in fig 2f.



Figure. 2a. Tweet about train service

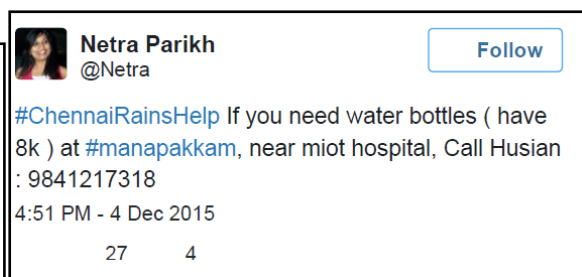


Figure. 2b. – Offers of help



Figure. 2c. Offers of help



Figure.2d. Appeal for food and water

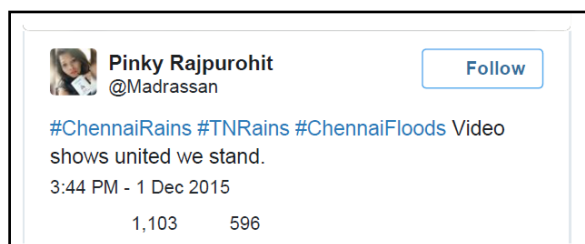


Figure. 2e. lift the spirits up of people of Chennai

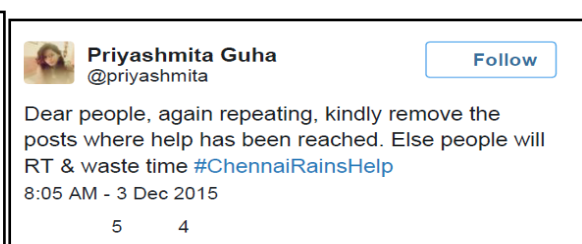


Figure. 2f. retweet confusion.

4.2. Trustworthiness of Tweets:

Not an easy task. The overload of aid messages sent out on Twitter, makes the people difficult to find which ones are the most recent and which ones are now out of date. The volunteers were requested to scan through the messages on #ChennaiRainsHelp and if you come across any tweets which are more than two hours old, they were asked to call the number and verify the condition, then to retweet the tweet with the words "verified." The volunteers were requested to help clean up the hashtag by checking the timelines of those who are offering help, and if there has been an update to a particular aid message, you are requested to approach the Twitter user and ask them to delete their out-of-date message to avoid any confusion.

In the case of Chennai flood, the information overload, and retweeting old help messages slowed down the relief work, also discouraged the volunteers, the integral part of relief work is the worst-case scenario of using social media in disaster situations. The reliability of real-time tweets shall be checked as monitoring of relief operations. Unreliable tweets and outdated tweets shall be removed, and the least retweeted request shall be highlighted shall enable the efficient disaster management system.

The classical method of machine learning are applied to classification of twitter datasets. Some researchers had been done using classifiers like naïve Bayes, decision tree, and SVM to identify spam, phishing and unreliable data on twitter information. Ranking algorithms with trust related queries are applied to obtain reliable information. Castillo et al shared that automated classification technique can be used to detect the news topics from conversational topics and assessed their credibility based on various features of twitter.[2], then they were evaluated with confirmed truth. The features of twitter used were the content of the tweet, user, hashtag and propagation based features. One of their observation were tweets without URL be likely to be non-related information. Negative sentiment content tweets were related to the event and classified as positive tweets.

The features of tweets were analyzed to find the reliability of information passed in the twitter are [7] topic specific, credibility annotations, retweet chains and dyadic pairs as shown in table 1. Topic extraction algorithms are used to extract the topic related tweets, and the retweet feature helps us to find the maximum retweeted message Retweet is a way of acknowledgment of information by public and broadcasting to other people. The dyadic pairs are the #hashtag and @reply pairs. This feature effectively eliminates the fake tweets.

TABLE 1. Twitter Features

<i>Features</i>	<i>Context</i>	<i>Hashtag(#)</i>
<i>Diverse Topics</i>	<i>Different subjects in twitter</i>	<i>N different topics</i>
<i>Trustworthiness</i>	<i>Assessment of tweets</i>	<i>Reliable or unreliable</i>
<i>Retweet Chain</i>	<i>Classification based on retweets</i>	<i>No of counts – long or short</i>
<i>Dyadic pairs</i>	<i>Interaction between hashtags</i>	<i>Dyadic or non-dyadic</i>

4.3 To enhance the trustworthiness of twitter information the retweet feature is taken as study.

Phase I: Latest information about Chennai flood was collected using twitter API. The collected data were cleansed and needed information shall be filtered with timeline and geo-location and stored as dataset.

Phase II: The dataset shall be applied with machine learning algorithms with keywords related to Chennai flood. They are #tag as Chennairain, Chennaimicro, Chennaiconnect, Chennai, flood, need, help, food, shelter, water, inundations, transport, medicine.

Phase III: Various visualization tools are in the market to visualize the data for analytical purpose. Tableau 9.2 is the visualization tool used for creating graphs and dashboard for visualizing the data.

The above phases are given as model in fig.3.

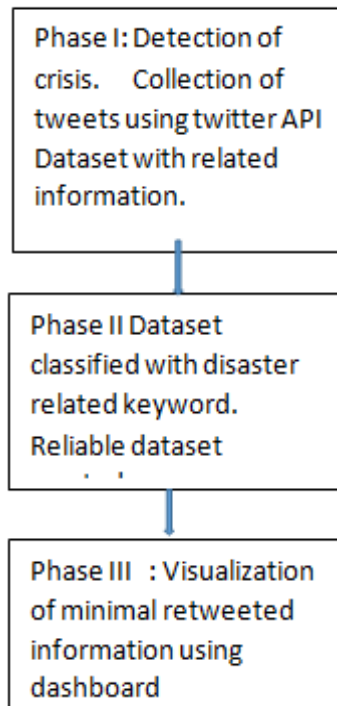


Figure 3. Phases of analysis of twitter

Using the dataset the peak hour of tweets is displayed as graph This graph shows us the time at which maximum tweets are sent. The time at which more people use twitter in Fig. 4.

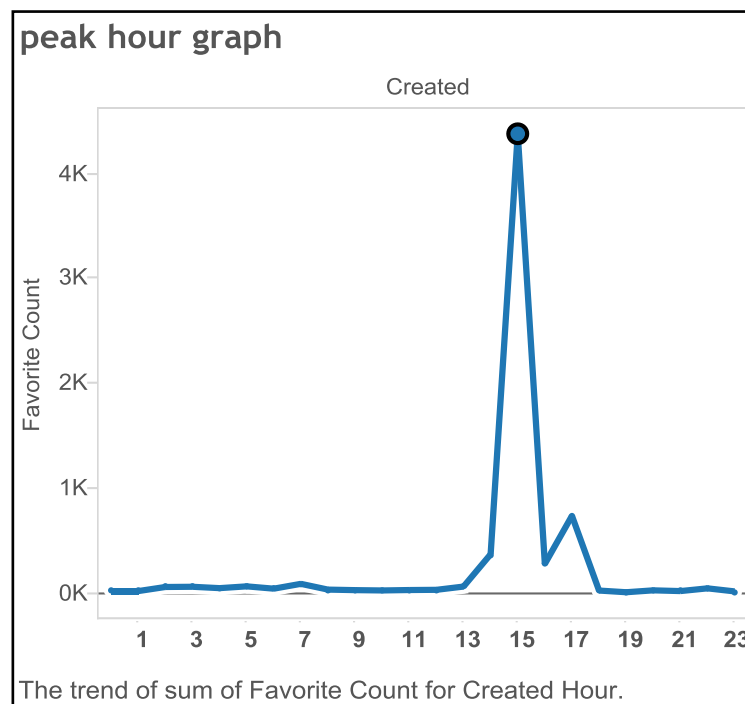


Figure. 4. Peak hour for tweets.

Figure. 5 shows the average retweet with total tweet count .This clearly shows us that almost each tweets is retweeted in three folds.

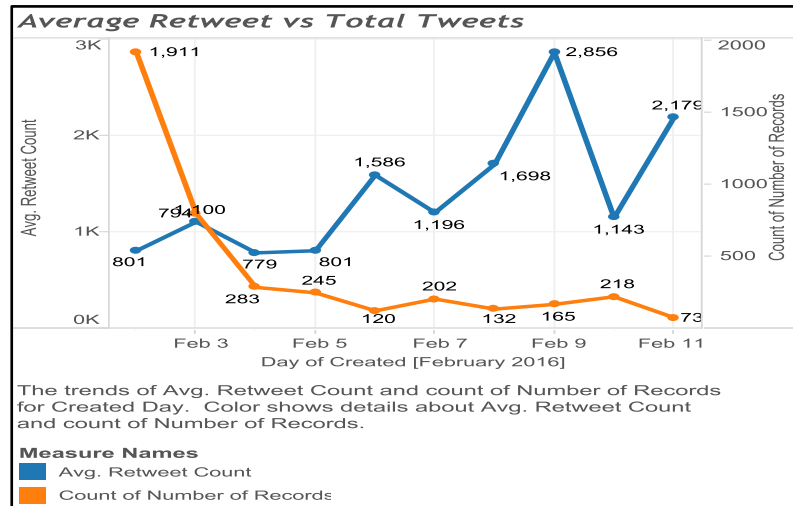
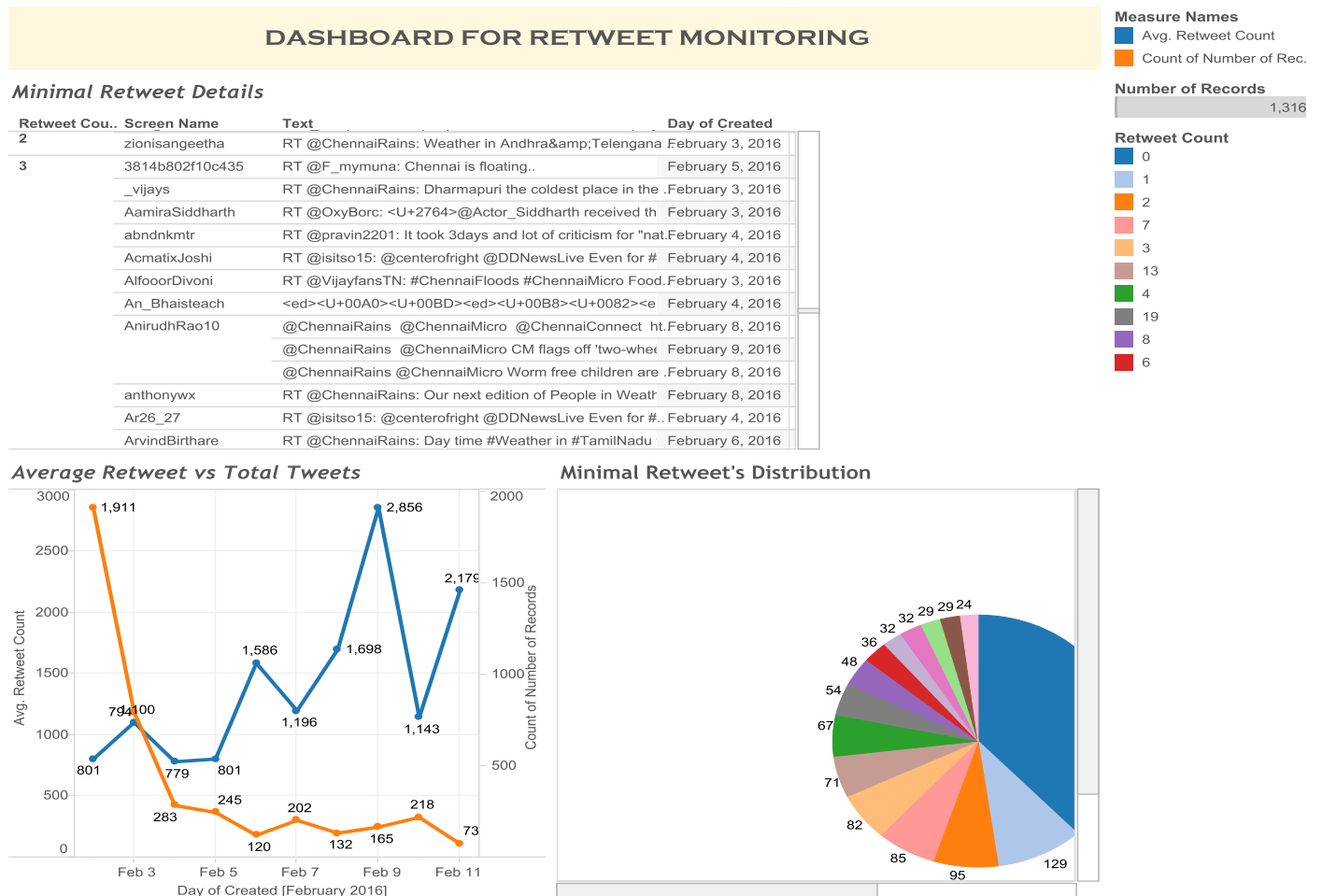


Figure.5. Average retweets Vs total tweets

The pie chart in the dash board shows the minimal retweet distribution. The pie chart clearly visualize the distribution of tweets which are not retweeted more than 20 users. If the reliable request messages are not passed to the respondent, the victims may not be rescued. The delay is very critical in disaster scenario, the pie chart shall be helpful to monitor the retweets and help the unreachable victims during disaster.

Dashboard about retweeted information is created as in Fig. 6. In the dashboard visualization of reliable minimally retweeted data is displayed in two different ways. The line graph shows the average of retweets to the total tweets. The pie chart shows the minimally retweeted message distribution. The minimal retweeted message shall be displayed in the table to provide faster relief actions by the respondent. The average of retweets and the total tweets in graphical form. This will help the organization to monitor and carry out the relied operation overloads.

Figure 6 : Dashboard for retweet monitoring.



5. CONCLUSION AND FUTURE WORK

Twitter is the social media, connected easily by the people during any crisis or disaster scenario. The present study proposes a model to enhance the reliability of tweets by using retweet features of twitter. In the crisis, retweets plays a vital role in connecting the affected with the respondent. Retweet's source and geo-location is identified and the request outside the crisis area can be eliminated. Then the tweets are classified using common keyword used in disaster scenario. The analyzed minimum retweeted messages are visualized using a dashboard. Such that the respondent for the relief operations can monitor the victims request and respond them with relief measures. In future the dyadic pairs would be identified and the contents would be compared for #hashtag and relevant @reply to identify the victims who were rescued and avoid out dated tweets and help responded tweets.

REFERENCES

1. Bruno Takahashi, Edson C. Tandoc Jr. , Christine Carnichael, "Communication on Twitter during a disaster : An analysis of tweets during Typhoon Haiyan in the Philippines", Elsevier journal of Computer in Human behaviour, vol.0747-5632, pp. 392-398, 2015.
2. Muhammed Imran and Carlos Castillo, Fernando Diaz, Sarah Viweg, " Processing Social Mass Messages in Mass Emergency, ACM Computing Survey, vol 47, pp. 67.1 – 67.38, June 2015.
3. Seonhw A Chow, "The Real-time Monitoring System of Social Big Data for Disaster Management", Advances in automatic control, pp. 110 – 114, 2012
4. Marcelo Menoza, Barbara Poblete, Carlos Castillo, " Twitter under Crisis : can we trust what we RT?", workshop on social media analytics , ACM, pp. 71 – 79, July 2010.
5. Silvia Ciotec, MihaiDascalu, StefanTrausan – Mattu, POSTDRU, 2012.
6. Takeshi Sakaki, Makoto Qkazaki and Yuaka Matsuo, " Tweet Analysis for Real-Time Event Detection and Earthquake reporting system development", IEEE , vol 25, pp. 919-931, April 2013.
7. Starr Roxanne Hiltz, Linda Plotnick, "Dealing with information overload when using social media for emergency management: Emerging solutions. " , Proceedings of 10th International ISCRAM Conference, pp. 823-827, May 2013.
8. John O'Donovan, Byungkyu Kang, Greg Meyer, Tobias Hollerer, SibelAdah," Credibility in Context : An Analysis of Feature Distributions in Twitter", IEEE International Conference on Privacy, Security, Risk and Trust, pp. 293 – 301, 2012
9. Andrea . H. Tapia , Kathleen. A. Moore , Nicolas . J. Johnson , "Beyond the Trustworthy Tweet:A Deeper understanding of Microblogged Data Use byDisaster Response and Humanitarian Relief Organizations. " , Proceedings of the 10th International ISCRAM Conference, pp. 770- 779, May 2013.
10. Twitter :www.twitter.org for twitter API
11. <http://www.easymedia.in/how-twitter-became-the-bridge-in-flooded-chennai/> for tweet images.
12. <http://www.ibtimes.co.uk/chennai-floods-ten-ways-help-victims-without-spending-any-money>