ISSN: 2349-6495

Natural Tragedy Commendation Hasty Alert Using Tweet Events Over Distributed Processing Framework

R.Pushpanathan, Dr.T.Nirmalraj, S.Rajalakshmi

Department of Computer Science, SCSVMV University Enathur, TamilNadu, India

Abstract— *An Event processing is the scheme of streams that related with information (data) about things that happen (events), and deriving a conclusion from tweet in real time. Twitter is a social network platform that consists of billions of users all over the world where people collaborate and Share information related to real world events. An important characteristic of Twitter is its real-time nature and also investigate the real-time interaction of events such as cyclones in Twitter and propose a framework to monitor tweets to detect a target event. These large scales tweet data processing are done by placing those tweet events in a distributed system. The server processes the tweet queue and executes the operations based on it. An devise classifier of tweets based on features such as the keywords in a tweet, the number of character, the number of words, and their context. The status update which almost pinpoints what is happening in and around an individual user and also tracks the user location. This small content with real world information when processed with some statistical tool may assist us to predict a live occurring event (e.g. cyclone) and regard each twitter user as a feeler and apply particle filtering, which are widely used for location estimation. Tweet in the message queue is done by Apache Kafka which is a distributed publish-subscribe messaging queue system. These frameworks will parallelize our computations over a cluster of machines.*
Keywords— Event, Distributed System, Apache Kafka, Cluster.

I. INTRODUCTION

Micro blogging sites allow instantaneous sharing of real-time data, messages, information. The main use of micro blogging sites is to provide useful information to the people. It is experimental that most of the knowledge is gained from social networks and many useful results are obtained. The traditional collaborative filtering does not consider social relations. In Twitter, a well-liked micro blogging service has become a new information channel for users to share information. It's an online social network used by millions of people all around the world. Twitter used as bridge to stay connected to their friends [1], family members and co-workers through their computers and mobile phones. Every day, nearly 170 million tweets are created and re-distributed by millions of active users which are not made. Micro blogging and plurk has a timeline view integrating video and picture sharing. We particularly observe Twitter because of its fame and data of information volume. An important common characteristic among micro blogging services is the live event occurrence of nature. Although blog users typically update their blogs once every several days, Twitter users write tweets a number of times in a single day. Users can know how other users are doing and often what they are thinking, view, wisdom, opinion about now, users repeatedly return to the site and check to see what other people are doing Vic vera. The huge volume updates results in numerous reports related to events. They include social events such as parties, sports, awareness programs, and natural disasters. It includes disastrous events such as storm, fire, traffic jam, riots, heavy rainfall, and cyclones. Actually, Twitter is used for various real time notifications such as that essential for help during an important emergency and live traffic updates.

This post well represents the drive of our study. The goal of this research study, "can we detect such live event occurrence in instantaneous by monitoring tweets?" This paper presents an analysis the real-time nature of twitter and proposes an event warning system / that monitors tweets and delivers notice on time. To obtain tweets on the target event closely, we concern on the semantic study of a tweet: For e.g., users might make tweets such as "cyclone" thus "cyclone" or "Earthquake" or "volcanic eruption" could be the keyword, but user might also tweet such as I'm in "storm of depression" or "cyclone in my location". We systematize the preparation of data that is training data and devise a classifier using a sustain vector machine based on features such as keywords in a tweet, the no. of words, and the context of target event words. Subsequently, we would like to make a probabilistic spatiotemporal model of an event. We make a crucial

ISSN: 2349-6495

possibility for each twitter user and it is gaze at as a sensor and each tweet as sensory information. These near sensors, which we identify social sensors, are of a huge variety and have various individuality: some sensors are very active; others are not. A sensor could end up in in-operable or malfunctioning sometimes (e.g., a user is driving or doing some sort of work). Consequently, social sensors are very noisy compared to ordinal physical sensors. Concerning a twitter user as a sensor, the event recognition problem can be abridged into the object detection and location outlook of the problem in a ubiquitous/pervasive computing surroundings in which we have numerous location sensors. Account holder has a mobile device or an active badge in an environment where sensors are placed. Through infra ray's message or a Wi-Fi signal, the user location is predictable as providing location-based services such as navigation and museum guides.

II. IMPLEMENTATION

2.1 EVENT DETECTION

We target event occurrence/detection. An event is an random categorization of a space/time region. An event might have actively contributed agents, passive factors, products, and a location in space/time. We target events such as cyclones, typhoons, and traffic jams & other normal happening which are able to be seen through tweets [6]. These events have several properties:

i) They are of enormous level (many users experience the event).

ii) They particularly influence people's daily life (for that reason, they are induced to tweet about it).

iii) They have the spatial and temporal regions so that real-time location estimation would be promising. Such events contain social events such as huge social gathering, sports events, exhibitions, accidents, and political campaigns. They also contain usual events such as storms, heavy rainfall, tornadoes, typhoons/hurricanes/cyclones, and cyclones. We assign an event we would like to detect using Twitter as a target event. material on each page should fit within a rectangle of 18 x 23.5 cm (7" x 9.25"), centered on the page, beginning 2.54 cm (1") from the top of the page and ending with 2.54 cm (1") from the bottom. The right and left margins should be 1.9 cm (.75"). The text should be in two 8.45 cm (3.33") columns with a .83 cm (.33") gutter.

2.2 Semantic Analysis on Tweet

To identify and detect a target event from Twitter micro blogging, we have to search from Twitter and find useful tweets. Tweets strength includes state of the target event. For e.g. "Users might tweet such as" cyclone!" or storm depression or cyclone weather approaching". Moreover, even if a tweet is referring to the target event, it might not be appropriate as an event report; for e.g. a user makes tweets such as "The cyclone yesterday was scaring and pulled us into the dark side", or "Three cyclone in four days. Japan scares me." These tweets without any doubt can be declared as truthfully to the mentions of the target event, but they are not real-time reports of the events. Therefore, it is essential to make clear that a tweet is actually referring to a real cyclone occurrence, which is denoted as a positive class. To categorize a tweet into a positive class or a negative class, we use a support vector machine (SVM) [14], which is a widely used machine-learning algorithm. By preparing positive and negative examples as a training set, we can produce a model to classify tweets automatically into positive and negative categories. We prepare three groups of features for each tweet as follows:

Features A (statistical features) the number of words in a tweet message, and the position of the query word within a tweet.

Features B (keyword features) the words in a tweet.

Features C (word context features) the words before and after the query word.

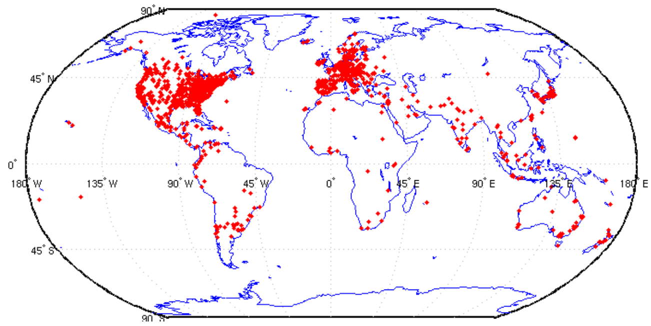


Fig.2.2: Twitter user map.

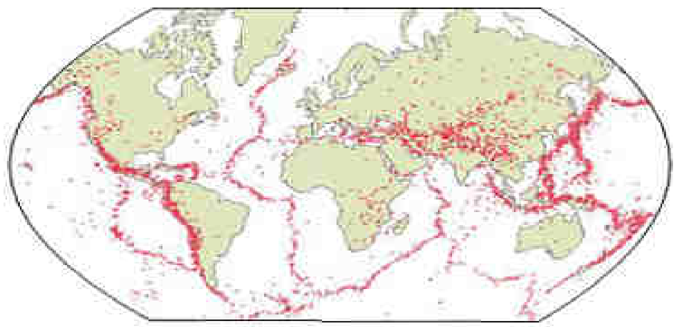


Fig.2.2: Twitter Cyclone map.

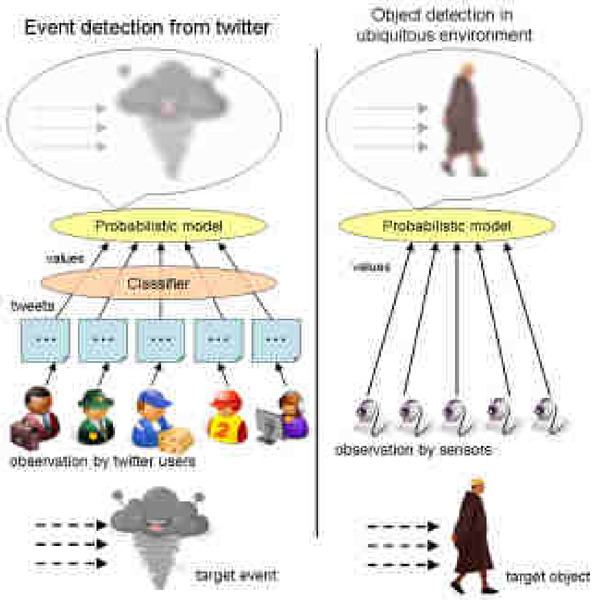
2.3 Tweet as a Sensory Value

We can look for the tweet and categorize it into a positive class if a user composes a tweet on a target event. If he/she makes a tweet about an cyclone occurrence, then it can be considered that he/she, as an "cyclone sensor", returns a positive value. A tweet can therefore be considered as a sensor reading [12].

ISSN: 2349-6495

Assumption 2.1 Each Twitter user is regarded as a sensor. A sensor detects a target event and makes a report probabilistically. The virtual sensors (or social sensors) have various characteristics: some sensors are activated (i.e. make tweets) only about specific events, although others are activated to a wider range of events. The number of sensors is large; there are more than 40 million sensors worldwide. A sensor might be inoperable or operating incorrectly sometimes (which means a user is not online, sleeping, or is busy doing something). Therefore, this social sensor is noisier than ordinal physical sensors such as location sensors, thermal sensors, and motion sensors. A tweet can be associated with a time and location: each tweet has its post time, which is obtainable using a search API. In fact, GPS data are attached to a tweet sometimes, e.g. when a user is using an iPhone. Alternatively, each Twitter user makes a registration on their location in the user profile. The registered location might not be the current location of a tweet; however, we think it is probable that a person is near the registered location. In this study, we use GPS data and the registered location of a user. We do not use the tweet for spatial analysis if the location is not available (We use the tweet information for temporal analyses.).

Assumption 2.2 Each tweet is associated with a time and location, which is a set of latitude and longitude. By regarding a tweet as a sensory value associated with location information, the event detection problem is reduced to detecting an object and its location from sensor readings. Estimating an object's location is arguably the most fundamental sensing task in many ubiquitous and pervasive computing scenarios. Figure 3 presents an illustration of the correspondence between sensory data detection and tweet processing. The motivations are the same for both cases: to detect a target event. Observation by sensors corresponds to an observation by Twitter users. They are converted into values by a classifier. A probabilistic model is used to detect an event, as described in the next section.



an cyclone and a typhoon. It is apparent that spikes occur on the number of tweets. Each corresponds to an event occurrence. In the case of an cyclone, more than 10 cyclone occur during the period. In the case of typhoon, India's main population centers were hit by a large typhoon in October 2012(Named Nilam). The distribution is apparently an exponential distribution. The probability density function of the exponential distribution is $f(t; \lambda) = \lambda e{-}\lambda t$ where $t > 0$ and $\lambda > 0$. The exponential distribution occurs naturally when describing the lengths of the inter-arrival times in a homogeneous Poisson process is a crucial assumption, but it enables application of various methods related to sensory information.

In the Twitter case, we can infer that if a user detects an event at time 0, assume that the probability of his posting a tweet from t to Δt is fixed as λ. Then, the time to make a tweet can be considered as an exponential distribution. Even if a user detects an event, therefore, she might not make a tweet right away if she is not online or doing something. She might make a post only after such problems are resolved. Therefore, it is reasonable that the distribution of the number of tweets follows an exponential distribution. Actually the data fits very well to an exponential distribution; we get $\lambda = 0.34$ with $R2 = 0.87$onaverage. To assess an alarm, we must calculate the reliability of multiple sensor values. For example, a user might make a false alarm by writing a tweet. It is also possible that the classifier misclassifies a tweet into a positive class. We can design.

2.4 Spatial model Tweet

Each tweet is made associated with a location based tracking process. If the chance given by the temporal model is larger than the threshold, the next steps are to determine the event location. We obtained the pinpoint

ISSN: 2349-6495

location information of each tweet using its associated Global Position Service (GPS) data or the registered location. Later that we relate particle filter to all set of tweet to get hold of the event location. To compute the location of a user where actual event is predicted several methods of Bayesian filters are proposed such as Kalman Filters ,multi-hypothesis tracking, grid-based and topological move toward and particle filters. For this study we use particle filter to estimate location of an event.

2.5 Particle Filters

Particle filter is a probabilistic approximation algorithm implementing a Bayes filter, and a member of the family of sequential Monte Carlo methods[11]. The particle filter works better than other comparable methods for estimating the locations of target events. The whole algorithm is given as follows:

- Put a query Q using search API every s seconds and obtain tweets T.
- For each tweet t belongs to T, obtain statistical, keyword, word context features, apply the classification to obtain value 0 and 1.
- If the tweets reached Threshold Limit then proceed step 4.
- For each tweet t € T, we obtain the latitude and the longitude using 1)GPS location 2)making a query to Google Map for the registered user .
- Calculate the estimated location of the event using normal particle filtering.
- Send alert E-mail and message to registered user as well as to nearest rescue team.

III. DISTRIBUTED FRAMEWORK

3.1 Twitter Storm

- A stage for doing analysis on streams of data as they come in simultaneously, so you can react to data as it happens.
- An extremely dispersed real-time computation system. Provides general primitives to do real-time computation.
- To simplify operational with queues & workers. Scalable and fault-tolerant toolbar.

3.2 Apache Kafka

- A distributed publish-subscribe messaging system. Designed for processing of real time activity stream data (logs, metrics collections, social media streams.
- Initially developed at LinkedIn, now part of Apache Does not follow JMS Standards and does not use JMS API.
- Kafka maintains feeds of messages in topics.

3.3 Processing Frame Work

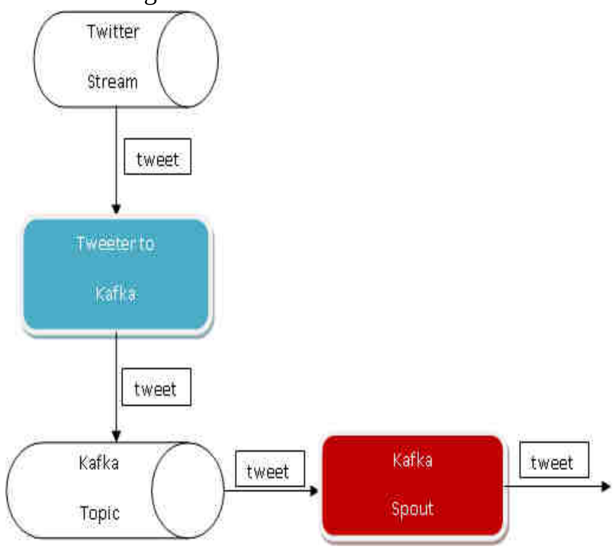


Fig.3.1:Kafka Topic Pipeline

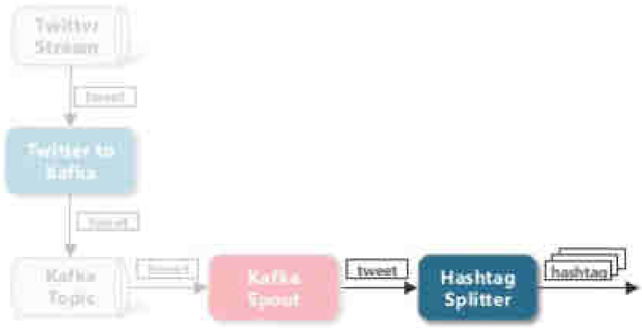


Fig.3.2:Kafka Hashtag Splitter

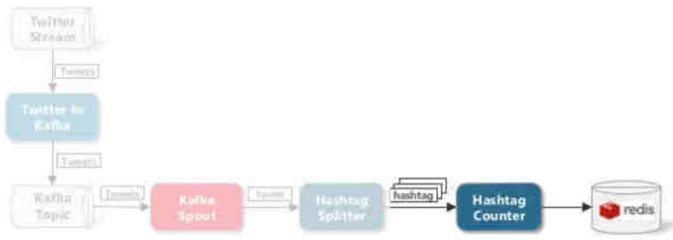
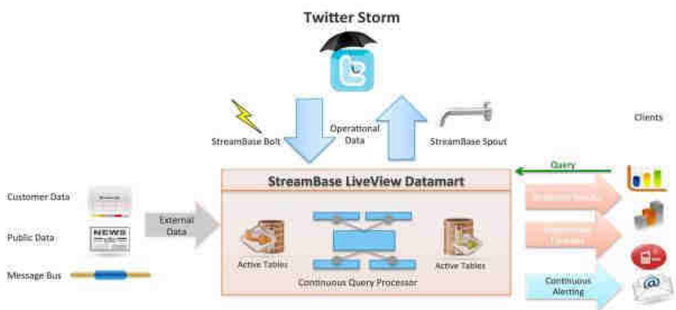


Fig.3.3:Kafka Hashtag Counter.



ISSN: 2349-6495

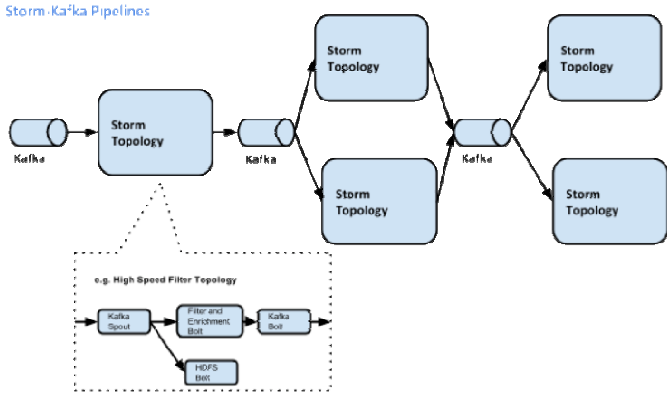


Fig.3.4:Distributed Process Framework

IV. SUMMARY AND DISCUSSION

In this paper, Twitter is a social network platform that consists of billions of users all over the world where people collaborate and Share information related to real world events. These large scales tweet data processing are done by placing those tweet events in a distributed system. Tweet in the message queue is done by Apache Kafka which is a distributed publish-subscribe messaging system. These frameworks will parallelize our computations over a cluster of machines.

Future Enhancements

The future work is to show the claimed, assumption, while strong, is quite logical considering the information collective in the twitter. Collecting these large data will impact processing to overcome it we can use Hadoop map-reduce concept for Big Data Processing via clustering.

V. SCREENSHOOTS

4.1 User Login Form

A Login is the process made by a User / Admin for connecting to a system or network service. Usually, a User must enter some Credentials, such as his User ID and Password, in order to successfully Login.

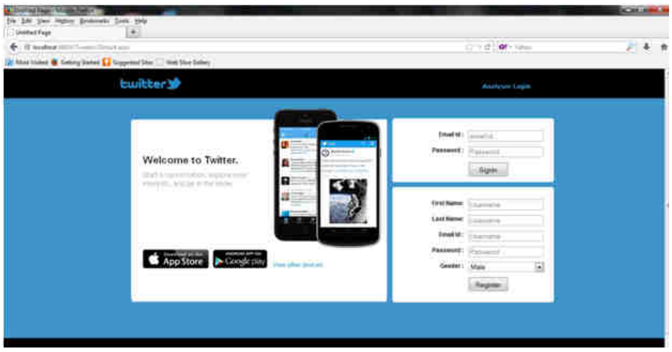


Fig.4.1:User Login Form

4.2 Registration

- New user has to register using the Registration Icon.
- The new user should be added into the database and it must be displayed in the user grid.
- User can able to easily delete or update their data's in the modules. Maintained under user profile.
- Maintain your profile
- Update your phone number and mailing address.
- Manage your account
- Set up Account Alerts.

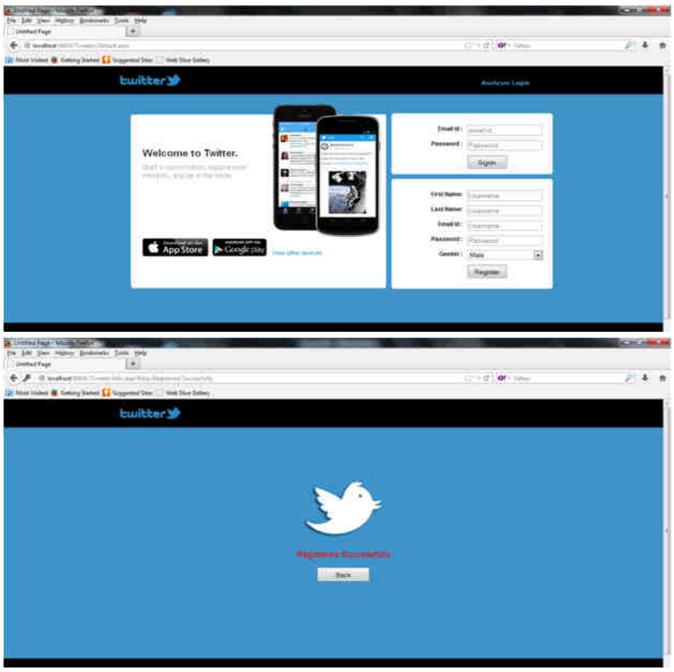


Fig.4.2:Registeration Form

4.3 Analyzer Login Form

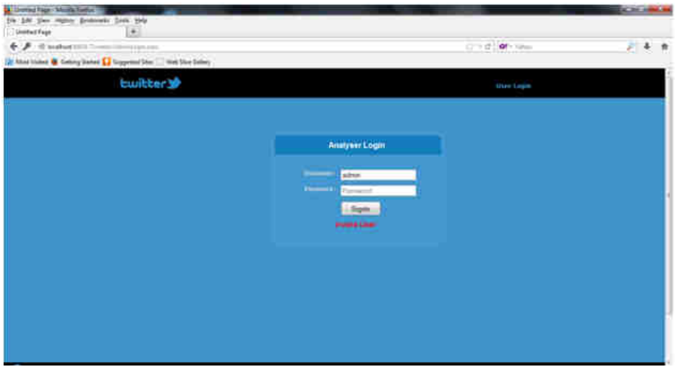


Fig.4.3:Analyzer Login Form

4.4 Tweet Message

Body of the tweet message is processed by kafka and based on topic message are Categorized

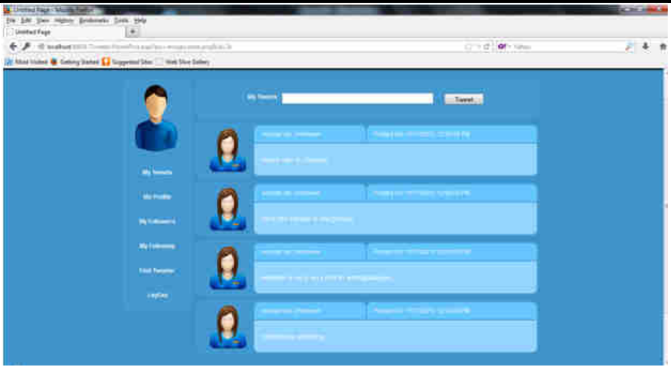


Fig.4.4:Tweet Message

4.5 General Crawl And Semantic Crawl

This form will analyze all incoming messages from kafka topic and crawl the entire system. Used to track real time event using the post tweets.

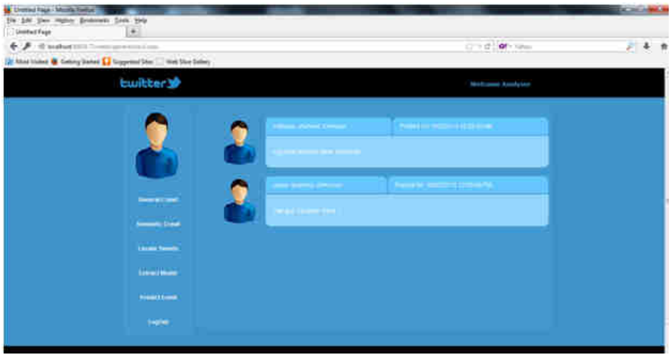


Fig.4.5.1:General Crawl

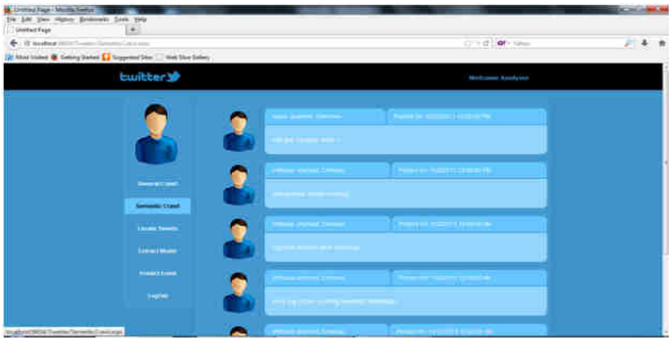
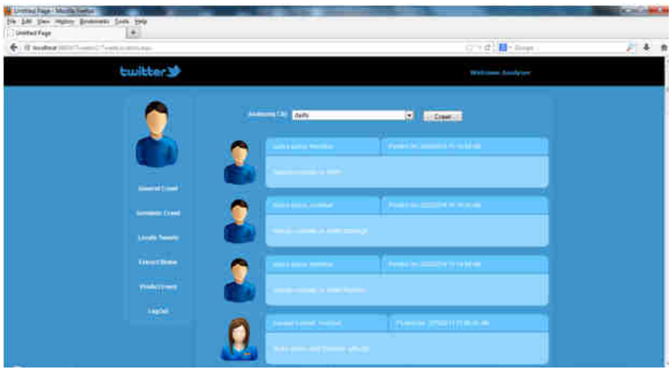


Fig.4.5.2:Semantic Crawl

4.6 Locate Tweet

Used to predict use location via post tweet message based on the current location we can identify the particular event occurrence.



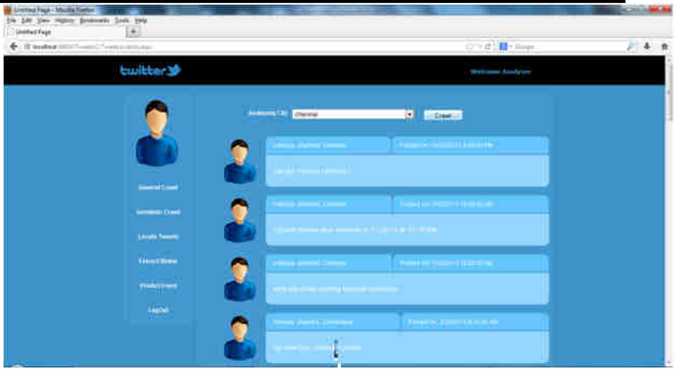


Fig.4.6:Locate Tweet

4.7 Predict Event

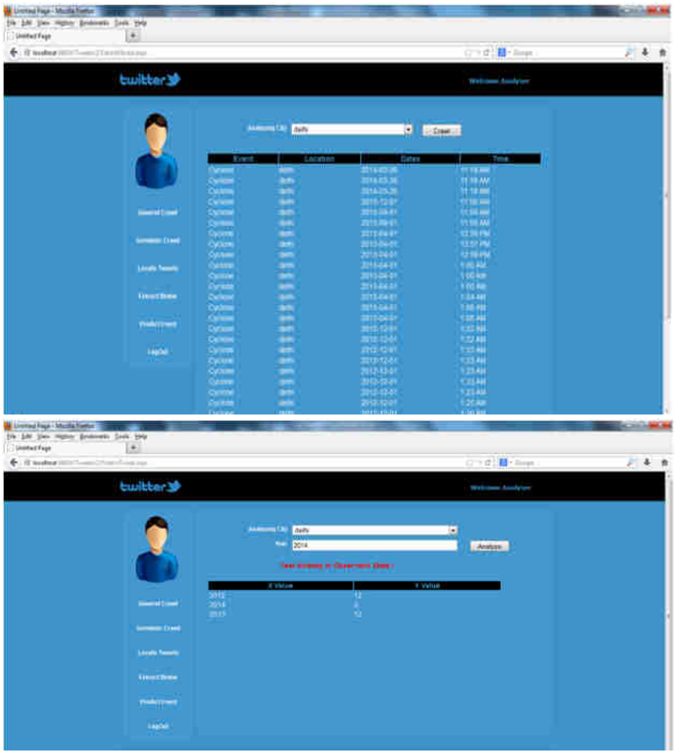


Fig.4.6: Predict Event

VI. ACKNOWLEDGEMENTS

I thank the ALMIGHTY GOD for enabling me to do this research work successfully.

I R. PUSHPANATHAN would like to express my gratitude to all who have helped me directly and indirectly during my research work.

I own a deep sense of gratitude and express my heartfelt and sincere thanks to SCSVMV UNIVERSITY.

I take this opportunity to thank my guide Dr. T. NIRMAL RAJ M.phil., PhD., Assistant Professor Department of Computer Science & Applications in SCSVMV UNIVERSITY her valuable guidance, continuous motivation and interaction while doing this research.

ISSN: 2349-6495

REFERENCES

[1] Takeshi Sakaki, Makoto Okazaki, and Yutaka Matsuo(2013) “Tweet Analysis for Real-Time Event Detection and Earthquake Reporting System Development”,IEEE

[2] K Sundharivel, Udaya Suriya Rajkumar(2014) “Real Time Natural Disasters of Data Analaysis Using Tweet” .

[3] R. Anitha, S. Selvakumar, V. Anusha Sowbarnika, and N. Varatharajan(2014) “Real-Time Event Detection and Earthquake”.

[4] Bala Kumari P1, Bercelin Rose Mary W2 and Devi Mareeswari(2014) Filtering Noisy Contents in Online Social Network by using Rule Based Filtering System.

[5] Ch. Pavan Kumar, B. Raghu Ram, and B. Hanmanthu(2013) “A Machine Learning Approach to Riddle Abused Scraps from OSN User Space”.

[6] Sandhya Rani Vemparala, Sayeed Yasin(2012): “Preventing Attacks in Multi-path Wireless Routing Protocols”.

[7] Dharnasi Rama Devi, Ms. Naga Aswani Puduru(2014) “A Probabilistic Model of Real Time Event Detection and Reporting”.

[8] G.L. Danah Boyd and S. Golder, “Tweet, Tweet, Retweet:Conversational Aspects of Retweeting on Twitter,” Proc. 43rd Hawaii Int’l Conf. System Sciences (HICSS-43), 2010.

[9] Takeshi Sakaki, Makoto Okazaki, and Yutaka Matsuo, “TweetAnalysis for Real-Time EventDetection and EarthquakeReportingSystem Development”.

[10] Vanithasri, K. and Tamilmani.G “Natural Disaster Recommendation Speedy Emergency Alert system using Twitter for social networks” [2012].

[11] S. Anand, and K. Narayana(2013). “Real-time Interaction of Events Detection Earthquakes”.

[12] Wenwen Dou, Xiaoyu Wang, William Ribarsky,Michelle Zhou “Event Detection in Social Media Data”.

[13] Amol jadhao, SweetyKumari, Firdos Khan, Shekh Sultan, Ruchita Khandge.(2015). “Real-Time Detection of Traff ic From Twitter Stream Analysis”.

[14] Thi Zin, Member, IAENG, Pyke Tin, Hiromitsu Hama and Takashi Toriu(2013). “Knowledge based Social Network Applications to Disaster Event Analysis”.

[15] Naveed Ahmad,(2010). “Flood Prediction and Disaster Risk Analysis using GIS based Wireless Sensor Networks”.

[16] Sandeep Panem, Manish Gupta and Vasudeva Varma. “Entity Tracking in Real-Time using Sub-Topic Detection on Twitter”.

[17] Mark Cameron, Bella Robinson, and Robert Power(2012). “Using Social Media to Enhance Emergency Situation Awareness Jie Yin, CSIRO ICT Centre Andrew Lampert, Palantir Technologies”.