Social Monitoring System for Dynamically Evolving Anomaly Over Text Stream

Yogesh Khedkar[1],Sayli Abhang[2] ,,Prakash Kshirsagar[3]
[1, 2, 3] Dept of Computer
[1, 2, 3] ZCOER, Savitribai Phule Pune University, India

Abstract- *A blog having a small group of members with their discussion. In a blog that is smaller than a traditional blog and contains very short entries is called as microblog. Using this real-time diffusion of information is going on. So in that so many abnormal discussion on microblog, this things are trending on the social network and be able to monitor their evolution and find related anomalies. So in this paper we proposed a system RING (real-time emerging anomaly monitoring system over microblog text streams). To find out and monitor this abnormal discussion, RING proposed a graph analytic approach such having some advantages like 1)RING is detect current trends emerging anomalies topics, 2)RING is among the first to discover emerging anomalies correlations in a streaming fashion, 3)RING works on the real-time data that work on the minutes to months data. RING is able to process big data to the entire Weibo or Twitter text stream with linear horizontal scalability.*

Keywords- RING, Twitter text stream, blog and microblogs, emerging, anomalies

I. INTRODUCTION

Social media are interactive Web-2.0 based applications like Facebook and Twitter etc. We are the social media having media having positive and negative impacts. Social media can help to improve individuals' sense of connectedness with real or online communities and social media can be an effective communication (or marketing) tool for corporations, entrepreneurs, non-profit organizations, including advocacy groups and political parties and governments. A popular component and feature of Twitter is retweeting. Twitter allows other people to keep up with important events, stay connected with their peers, and can contribute in various ways throughout social media. Retweeting is helpful strategy, which informs individuals on Twitter about popular trends, posts, and events. On the basis of these popular trends some abnormal things are happen so in this paper we are introducing RING System, that managing Real-Time Emerging Anomaly Monitoring System.

II. PROBLEM DEFINITION

Earlier works have expanded on mining and ranking existing routes from check-in data. But suggestions about your choice of place for traveling is not available even though you have enough information about place, here we are trying to resolve this problem.

III. OBJECTIVES

The present study in taken in hand keeping in mind the following objective :-

- The aim of this project is to detect the anomalies from current trends.
- This application can be used to detect anomalous event from text streams whenever some user post any tweets/comments related to their trending topic.
- In this, we are forming the module like, real-time event detection.
- We are applying our system on social media like Twitter, twitter have some current trends with his hashtags and current topics.

IV. SYSTEM DESIGN

Data Flow Diagram 0:

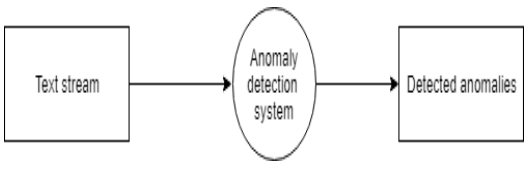


Fig 1: Dfd level 0

Data Flow Diagram 1:

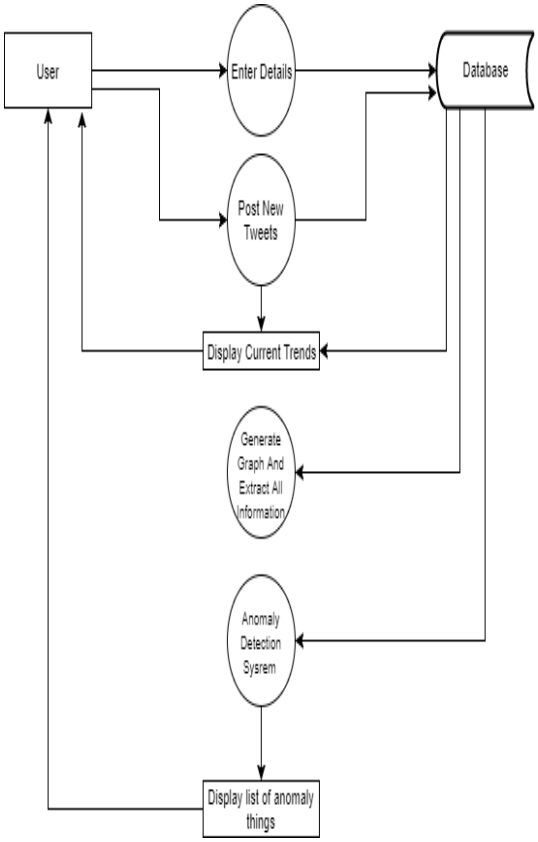


Fig 2 : Dfd level 1

Data Flow diagram 2:

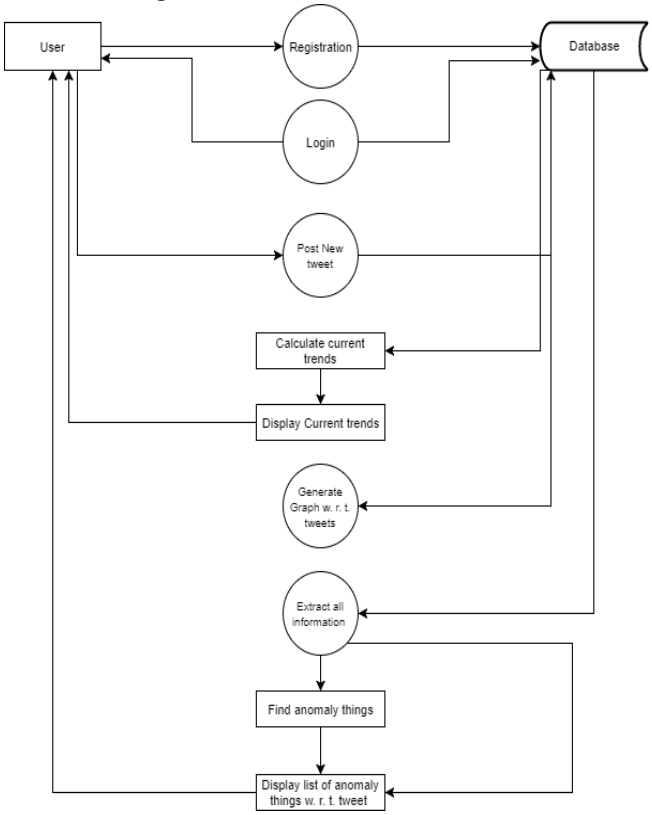
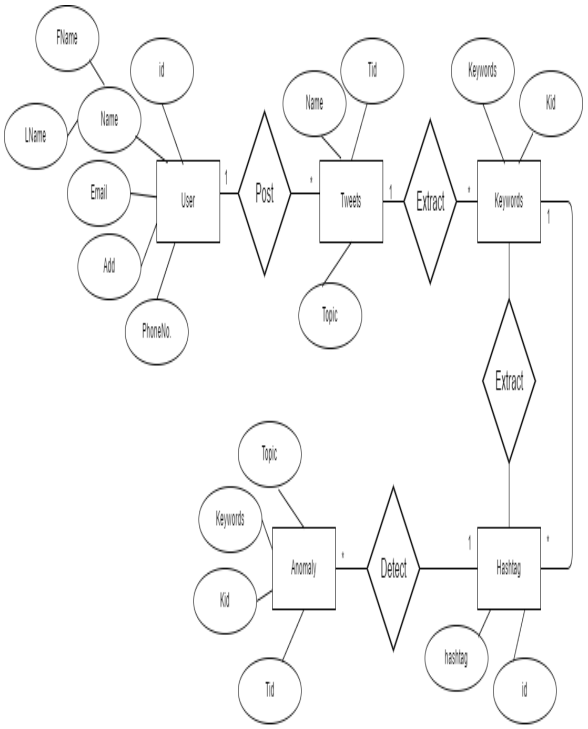
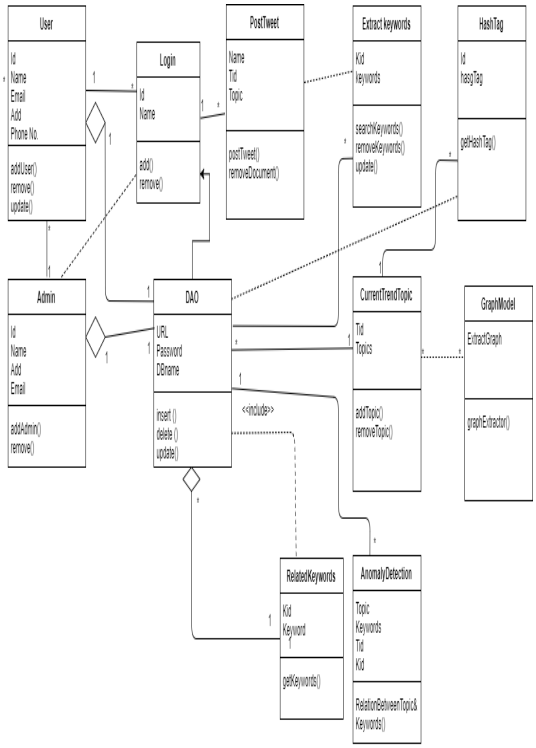


Fig 3 : Dfd level 2

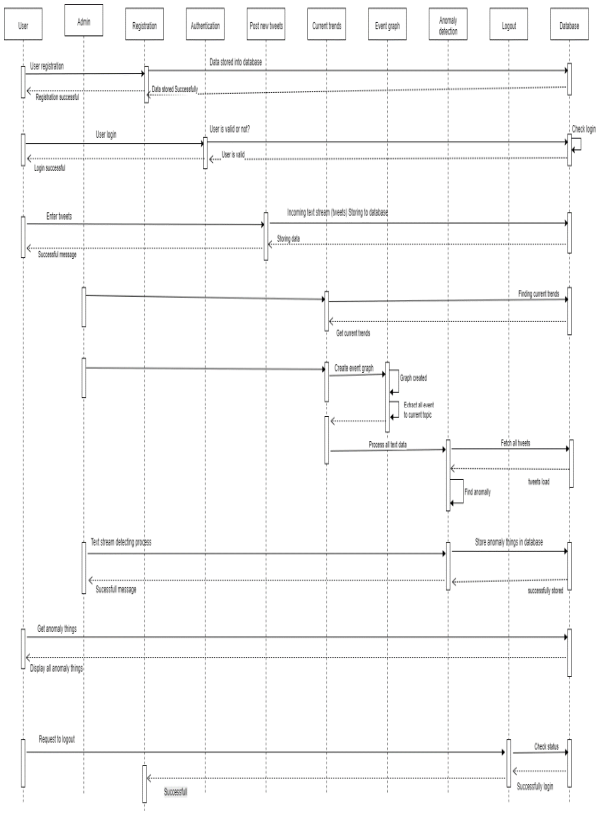
ER Diagram:



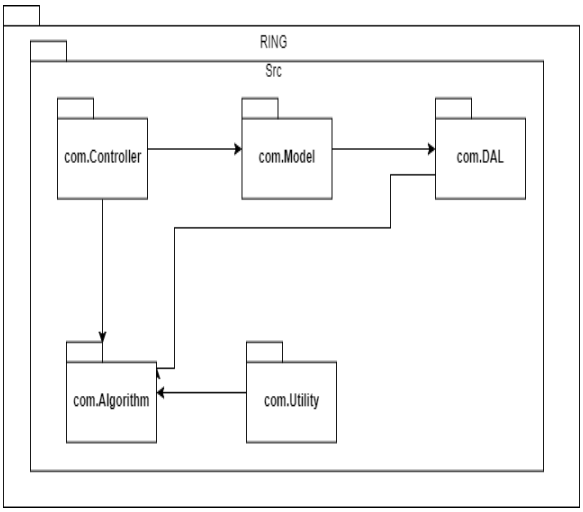
Class Diagram:



Sequence Diagram:

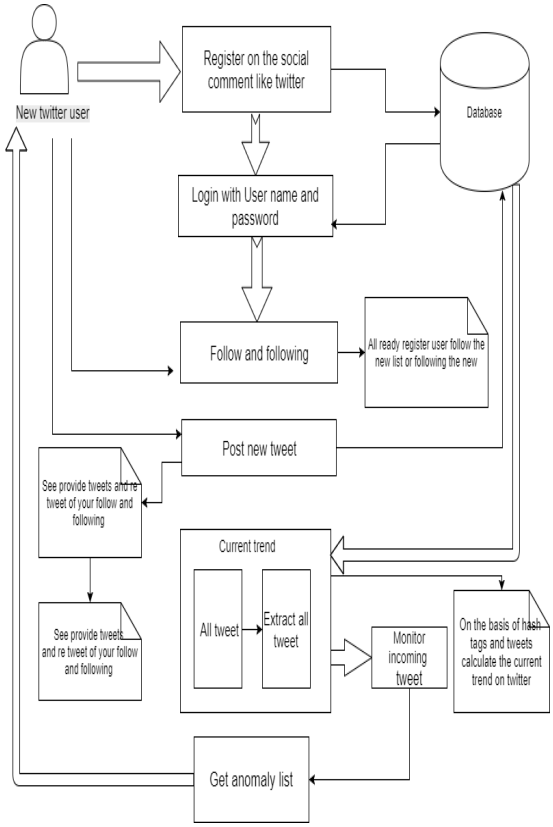


Package diagram:



V. SYSTEM ARCHITECTURE

The prime focus of this system is to analyze the post to their tweets on social network. We are applying our system on social media like Twitter, Twitter have some current trends with his hash tags and current topics. And we will use those current topics and hashtags to detects trending topics which will have anomalies. Using this real-time diffusion of information is going on. So, in that so many abnormal discussions on microblog, this thing are trending on the social network and be able to monitor their evolution and find related anomalies. So, in this paper we proposed a system RING (real-time emerging anomaly monitoring system over microblog text streams).



Related Works

VI. HYPOTHESES

- In this, we are developing a system that helps a user to gives their opinion related to their topics. User can define the problem of their topics.
- Naive Bayes is applied to differentiate news, ads and wisdom words among detected trending events, where the latter two classes are major types of spams on Weibo.
- The classifier is trained with manually labeled data depends on features of content, users and temporal information.

VII. DELIMITATION OF THE STUDY

- In this paper we used the naïve bayes Algorithm for classification of topics.
- Generate a graph of emerging disappearing and reoccurring topics.

- And then we find the list of anomalies.

DESIGN OF THE STUDY

We designed a graph stream model for anomalous event detection over the short and noisy text in microblog services. To represent texts with a graph stream model, we consider keywords as nodes and their co-occurrence relationships in each tweet as edges. For each incoming tweet, we generate a binary clique graph over its keywords and retrieve the edge set.

SAMPLE OF THE STUDY

In this, we are developing a system that helps a user to find the anomalies.

TOOLS UDED

Software Requirement: A social networking service is an online platform which people use to build social relation with other people who share similar personal or career interests, activities, backgrounds or real-life connections. Social networking services are Internet-based applications.

- Operating System: Windows 95/98/2000/XP
- Front end : HTML, JSP, CSS
- Backend: MySQL
- JDK 1.8

Hardware Requirement: The hardware design of the system includes designing the hardware units and the interface between those units.

- Processor – Pentium –III
- Speed – 2.4 GHz
- RAM - 256 MB (min)
- Hard Disk - 20 GB

VIII. STATISTICAL TECHNIQUE USED

We have developed Login and Registration which manages the user profiles, so that the users can post their tweets with their profile pictures. As soon as Users post the tweets related to any topics, the database stores these all details related to the post. Likewise, Database stores the information related to post of all users when they post with tweets.

Input: Tweets and find the emerging topics and using hashtag keyword create a node of graph. If any node not in relationship with topic node then find relation of incoming trends based on hashtag keywords and then consider it as anomaly.

ALGORITHM

This paper use three algorithms.

1] **Naive Bayes:**

- Naive Bayes is applied to differentiate news, ads and wisdom words among detected trending events, where the latter two classes are major types of spams on Weibo.
- The classifier is trained with manually labeled data depends on features of content, users and temporal information.

2] **Sub event Hierarchy construction:**

Input: Tweets
Output: Get anomaly list.

1. User posts any tweets related to their trending topic.
2. Find the anomaly things in current trends
3. And then generate the graph of that anomaly things
4. Get the list of that anomaly things.

3] Graph model:

- It take input as a tweets and then find emerging topic using keyword create node of graph find relations of incoming trends based on keywords and then consider it as anomaly.

System Workflow

The new user register on the social network like twitter and then login. The user here has to post their comment on the social network and be able to monitor their evolution and find related anomalies and the system then retrieves the anomaly thing of that particular post from database which the user has posted and these posts are displayed in the list. And detect anomaly event in trends which store into database. So, for every post there is a different list and there is also facility for user to view all the post. Here the user can see the post in their system.

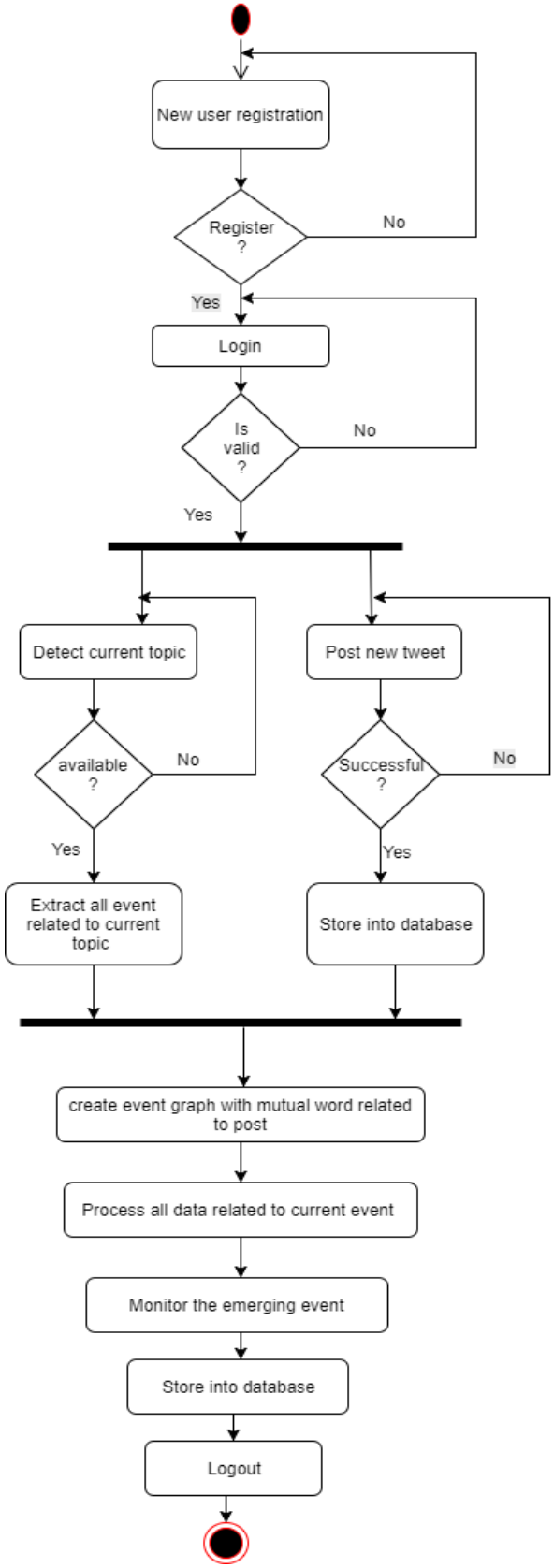


Fig 2: Activity Diagram

OUR APPROACH

Symbol	Description
$G(t)$	Undirected temporal graph of keywords.
λ	Decay rate of the weight of an edge arrival
$F(i,j,t)$	The weight of an edge accumulated till current time t.
$S(i,t)$	The set of neighboring nodes in node i's locality
$\alpha(i,t)$	The sum of the edge frequencies in locality set $S(i,t)$.
$HA(i,t,\lambda)$	Half-life activity change of a node i.
$G_E(t)$	Denoised keyword graph with rich information
$G_T(t)$	Trending keyword graph which is binary

Table 1: Notation Table

Step 1: To form undirectional graph of G(t),
G(t) = (N(t),A(t))….1
Where, N(t)- Set of all distinct nodes in G(t) at time t,
A(t)- As the sequence of edges received so far.

Step 2: Decay weight: To calculated arrival time of incoming data

Time t=1/ λ…2

Step 3: Weighted Frequency: Accumulated decay weights over all instances of its arrivals till time t.

$$F(i,j,t) = \sum_{k=1}^{n_{ij}^t} N(i,j,T(i,j,k)) \cdot 2^{-\lambda \cdot (t - T(i,j,k))}$$

….3

Step 4: Node Activity Frequency: Sum of the edge frequencies.

$$\alpha(i,t) = \sum_{k=j_1^i(t)}^{j_{|S(i,t)|}^i(t)} F(i,k,t)$$

….4

Step 5: Denoised Keyword Graph): Is a directed weighted graph

GE(t) = (VE(t),EE(t)), VE(t)….5

Where, EE(t)- Contain edges among all keywords with their weighted frequency.

Step 6: Trending Keyword Graph: A binary graph

GT (t) = (VT (t), ET (t)), ….6

where VT (t) - contains only trending keywords.
ET (t) - Edges between trending keywords.

Step 7: Trending Event: To calculate the current trending event, we used scalability and event denoising.

IX. EXPERIMENT RESULT

Extensive experiments have been conducted on a large-scale real-life tweet dataset. The results demonstrate the promising performance of our event indexing and monitoring methods on both efficiency.

X. FUTURE SCOPE

In the future, we would conduct research on automatic story telling and developing knowledge base into our results to provide better intelligence.

XI. ACKNOWLEDGMENT: (OPTIONAL)

It gives us great pleasure in presenting the preliminary project report on ‘ Social Monitoring System for dynamically evolving anomaly over text streams’.

I would like to take this opportunity to thank my internal guide Prof. Guide
Name…….. for giving me all the help and guidance I needed. I am really grateful to them for their kind support. Their valuable suggestions were very helpful.

I am also grateful to Dr. Prasad S. Halgaonkar, Head of Computer
Engineering Department, ZES' Zeal College of Engineering and Research for
his indispensable support, suggestions.

In the end our special thanks to Other Person Name ……for providing various resources such as laboratory with all needed software platforms, continuous Internet connection, for Our Project.

Yogesh Khedkar
Sayli Abhang

XII. CONCLUSION

In this paper, the advantages RING system is using the existing and proposed system using the monitoring the abnormal emerging event task. Further, RING’s infrastructure is equipped with customized optimization on its full-text search engine and distributed graph processing engine to perform event monitoring more efficiently

REFERENCE

[1] Earthquake shakes twitter users: real-time event detection by social sensors http://ieeexplore.ieee.org/document/7402054/
[2] Scalable distributed event detection for twitter http://ieeexplore.ieee.org/document/6691620/
[3] Topicsketch: Realtime bursty topic detection from twitter http://ieeexplore.ieee.org/document/7457245/ HYPERLINK "http://ieeexplore.ieee.org/document/7457245/S"S
[4] Sub-Event Detection from Tweets http://ieeexplore.ieee.org/document/7966112/
[5] Indexing Evolving Events from Tweet Streams http://ieeexplore.ieee.org/document/7498413/
[6] PAGE: A Partition Aware Engine for Parallel Graph Computation http://net.pku.edu.cn/~cuibin/Papers/2013cikm-page.pdf
[7] Evaluating Real-Time Search Over Tweets https://pdfs.semanticscholar.org/b488/24a9a2197048fea578d77b1db7a8cc3d0b01.pdf
[8] Dense subgraph maintenance under streaming edge weight updates for real-time story identification http://vldb.org/pvldb/vol5/p574_albertangel_vldb2012.pdf
[9] Learning evolving and emerging topics in social media: a dynamic nmf approach with temporal regularization http://vikas.sindhwani.org/wsdm227-saha.pdf
[10] Quantifying information flow during emergencies http://barabasi.com/f/451.pdf