Twitter Based Event Detection System

Ankit Kumar, Mimansa Singh, Rajan Yadav, Shubham Ghare, Prof. Abhijit D. Jadhav

[#]DEPARTMENT OF COMPUTER ENGINEERING,

SAVITRIBAI PHULE PUNE UNIVERSITY, PUNE

Address Including Country Name

[1]first.author@first-third.edu

[*]Second Company

Address Including Country Name

[2]second.author@second.com

ABSTRACT

Microblogging, today has become a very popular communication tool among Internet users. Millions of messages are appearing daily in popular web-sites that provide services for microblogging such as Twitter, Tumblr, Facebook. Because of a free format of messages and an easy accessibility of microblogging platforms, Internet users tend to shift from traditional communication tools to microblogging services. Traditional media can be in some or form inaccurate. As more and more users post about products and services they use, or express their political and religious views, or express and directly report a current event/incident happening around them, microblogging websites become valuable sources of peoples opinions and sentiments.With its over 240 million users tweeting out more than 500 million messages daily, Twitter is shaping public opinion like never before.

Keywords— tweets,events,analysis, IDC(Importance Degree Calculating), Tweet Preprocessing (TP), Remarkable Words Detecting (RMD).

I. INTRODUCTION

The project aims to demonstrate how Twitter Data Analysis can be used to give a clear picture about a particular event, so that the user can get an overall idea about events happening currently.The contribution of our project towards the technology is that we are demonstrating how the power of Data Analytics can be used for the benefits of the general public.

II. LITERATURE SURVEY

1.Detecting Real Time Events using Tweets Koichi Sato, Junbo Wang, Zixue Cheng

According to author, Big Data has been one of main topics in the field of computer science. Additionally, demand for observations of the real world in real time has increased to provide services or information to people accordingly. For example, when disaster occurs, government can appropriately respond to the disaster if the situations in the disaster-stricken areas are real timely grasped. Although there are many kinds of blog services and they are functioning as one of Big Data source, Twitter is considered as the most active Big Data source. Users can feel free to post a tweet anywhere in real time, since twitter limits a tweet to 140 characters. In this paper, a scheme is developed which can detect what happens in real world in real time only by analyzing tweets as Big Data and let a user know the event. To this end, the following problems have to be solved. They are a) quantifying importance of words accurately and b) evaluating the quanti_ed values dynamically. As the solutions for the problems, two new methods are proposed which are the Extended Hybrid TF-IDF and the Remarkable Word Detecting Method, and they are used in the proposed scheme. Finally an experiment is executed to evaluate the proposed methods and scheme.

2. TEDAS: a Twitter based Event Detection and Analysis System Rui Li, Kin Hou Lei, Ravi Khadiwala, Kevin Chen- Chuan Cheng

Witnessing the emergence of Twitter, we propose a Twitter-based Event Detection and Analysis System (TEDAS), which helps to (1) detect new events, to (2) analyze the spatial and temporal pattern of an event, and to (3) identify importance of events. In this demonstration, they show the overall system architecture, explain in detail the implementation of the components that crawl, classify, and rank tweets and extract locations from tweets, and present some interesting results of system .

3.A Survey of Techniques for Event Detection in Twitter - Farzindar A Tefeh and Wael Khreich.

According to author, Twitter is among the fastest-growing microblog- ging and online social networking services. Messages posted on Twitter have been reporting everything from daily life stories to the latest local and global news and events. Monitoring and analysing this rich and continuous user generated content can yield unprecedentedly valuable information, enabling users and

organizations to acquire actionable knowledge. The author in this paper provides a survey of techniques for event detection from Twitter streams. These techniques aim at _nding real-world occurrences that unfold over space and time. In contrast to conventional media, event detection from Twitter streams poses new challenges. Twitter streams contain large amounts of meaningless messages and polluted content, which negatively affect the detection performance. In addition, traditional text mining techniques are not suitable, because of the short length of tweets, the large number of spelling and grammatical errors, and the frequent use of informal and mixed language. Event detection techniques presented in literature address these issues by adapting techniques from various _elds to the uniqueness of Twitter. [3]

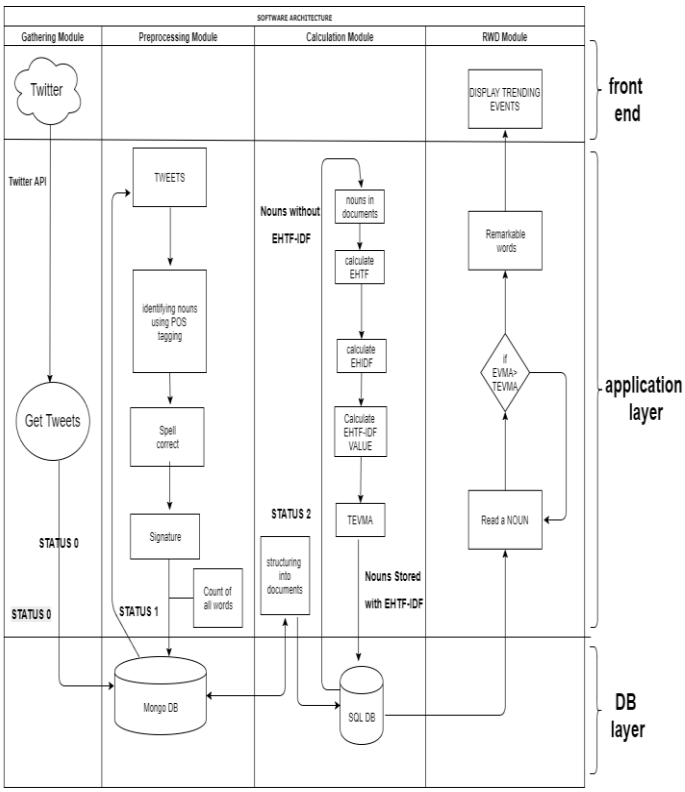
4. Using TF-IDF to Determine Word Relevance in Document Queries Juan Ramos

In this paper, we examine the results of applying Term Frequency Inverse Document Frequency (TF-IDF) to determine what words in a corpus of documents might be more favorable to use in a query. As the term implies, TF-IDF calculates values for each word in a document through an inverse proportion of the frequency of the word in a particular document to the percentage of documents the word appears in. Words with high TF-IDF numbers imply a strong relationship with the document they appear in, suggesting that if that word were to appear in a query, the document could be of interest to the user. The author has provided evidence that this simple algorithm efficiently categorizes relevant words that can enhance query retrieval

III. ARCHITECTURE DIAGRAM

There are 4 Modules- Gathering or extracting, Preprocessing,

There are 3 layers - Front End(visible layer), application layer and Database layer.



IV. CONCLUSION

Thus, we conclude that the Twitter based eventdetection system is a NP Hard class problem because it cannot besolved in deterministic polynomial time. we have just proposed andwe are in designing phase, so our system is NP-hard..

ACKNOWLEDGEMENT

We would like to take this opportunity to thank our internal guide **Prof. Abhijit D. Jadhav** for giving us all the help and guidance we needed.We are really grateful to them for their kind support. Their valuable suggestions were very helpful. We are also grateful to **Dr. Pramod Patil** , Head of Computer Engineering Department, Dr. D.Y.Patil Institute of Technology for his indispensable support, suggestions.

REFRENCES

[1] Koichi Sato, Junbo Wanga and Zixue Cheng, "*Detecting Real-time Events using Tweet*" 2016 IEEE Symposium Series on Computational Intelligence.

[2] Rui LI, Kin Hou Lei, Ravi Khadiwala and Kevin Chen-Chuan Chang, "*TEDAS: a Twitter Based Event Detection and Analysis System*", 2012 IEEE 28th International Conference on Data Engineering.

[3] Farzindar Atefeh, Wael Khreich,"*A Survey of Techniques for Even Detection in Twitter* ", published by - Wiley Periodicals, Inc,September 2013

[4] Kumar Shamanth, Liu Huan, Mehta Sameep, L. Venkata Subramaniyam, " *Exploring a Scalable Solution to Identifying Events in Noisy Twitter Streams*", Proceedings of the 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, 2015.

[5]"*Using TF-IDF to Determine Word Relevance in Document Queries* ",by Juan Ramos, January 2003

[6] T. Sakaki, M. Okazaki and Y. Matsuo, *"Earthquake shakes twitter users: Real-time event detection by social sensors"*, Proc. 19th International Conference on WWW2010, 2010.

[7]1. D. Inouye and J. K. Kalita, "*Comparing twitter summarization algorithms for multiple post summaries*", 2011 IEEE International Conference on Privacy Security Risk and Trust and IEEE International Conference on Social Computing, 2011.