ISSN: 2277-9655
Impact Factor: 4.116
CODEN: IJESS7

IJESRT

INTERNATIONAL JOURNAL OF ENGINEERING SCIENCES & RESEARCH TECHNOLOGY

REVIEW ON: KEYWORD BASED OPERATIVE SUMMARIZATION USING EVOLUTIONARY TWEETS

Vrushali Muley*, Manisha Kadam, Reshma Magdum, Shruti Jadhav,
C. S. Bhosale
* Department of Compute, MMIT, India

DOI: 10.5281/zenodo.400962

ABSTRACT

Small text messages such as tweets are shared at an unprecedented rate. As the data is growing enormously, analysis of this data is hard to achieve. So in this paper we propose a system which generates summary from the given dataset. In his system we used an algorithm named Tweet Cluster Veector (TCV), which is used in data cluster formation, a TCV- Rank summarization method is used in summary generation, and Topic evolution detection method is used for topic relevant summary generation. The system provides flexibility and effective summarization of the given dataset.

KEYWORDS: 1. Tweet Stream
i) Tweet data
ii) Plugin
iii) Twitter timeline
2. Continuous Summarization
i) Set of documents
ii) Content presentation
iii)Tweet Summarization
3. Timeline Summarization
i)Real-time Timeline
ii)Tweet Cluster Vector Introduction

INTRODUCTION

As the popularity of social networking sites is increasing every new day, large amount of data is shared by people. For example Twitter, which receives over 400 million of tweets every day. So while searching for specific topic on twitter, It may result to millions of tweets and going through all tweets one by one is kind of difficult task. Considering this problem we developed a system which efficiently provides the summary of the given dataset.

SOFTWARE REQUIREMENT

BACK END: MySQL

MySQL is an Open-Source Relational Database Management System, which supports for cross platform. It supports various data manipulating methods. It is written in C and C++. It works on many system platforms like Linux, Microsoft Windows etc.

FRONT END: Eclipse IDE

Eclipse is an Integreted Development Environmet which can be used for customizing the environment. It is used to produce Java applications as it is mostly written in Java language. It also supports other languages such as C, C++ etc.

HARDWARE REQUIREMENT

1. Processor – i3 and above
2. RAM – 512 MB

ISSN: 2277-9655
Impact Factor: 4.116
CODEN: IJESS7

3. Hard Disk – 40 GB

To implement the proposed system we need a 64 bit Operating system having i3 processor. The system should have minimum of 512 MB RAM and a Hard disk of 40 GB.

SYSTEM ARCHITECTURE

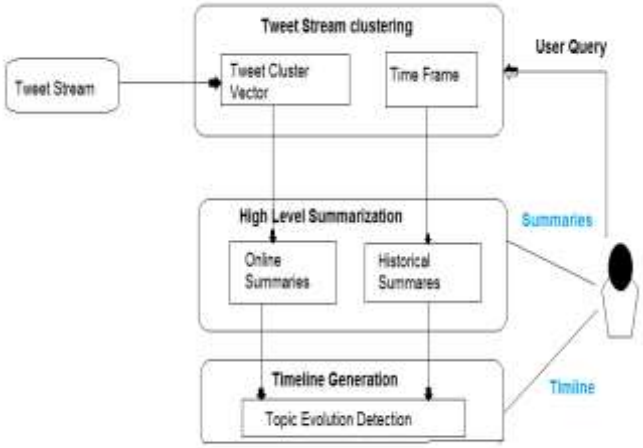

Fig -1: System Architecture

ADVANTEGES

1) It Easily sort the data according to the keyword.
2) Also Useful in E-commerce for determining popularity.
3) Massive data converted into short summary.

LITERATURE SURVEY

Zubiaga, D. Spina, E. Amigo, and J. Gonzalo, Towards real-time summarization of scheduled events from twitter streams, in Proc. 23rd ACM Conf. Hypertext Social Media, 2012,pp. 319320.
This paper explores the real-time sumarization of scheduled events such as soccer games from torrential ow of Twitter streams. It substantially shrinks the stream of tweets in real-time, and consists of two steps: (i) sub-event detection, which determines if something new has occurred, and (ii) tweet selection, which picks a representative tweet to describe each sub-event. We compare the summaries generated in three languages for all the soccer games in Copa America 2011 to reference live reports oered by Yahoo! Sports journalists.

A framework for clustering evolving data streams, Authors: C. C. Aggarwal, J. Han, J. Wang, and P. S. Yu.
In this paper, the clustering problem for data stream applications. The clustering problem is dened as follows: for a given set of data points, we wish to partition them into one or more groups of similar objects. The similarity of the objects with one another is typically dened with the use of some distance measure or objective function. a fundamentally different philosophy for data stream clustering which is guided by application-centered requirements. The idea is divide the clustering process into an online component which periodically stores detailed summary statistics.

[3]. BIRCH: An efficient data clustering method for very large databases.Authors: T. Zhang, R. Ramakrishnan, and M. Livny
This paper presents a data clustering method named BIRCH (Balanced Iterative Reducing and Clustering using Hierarchies), and demonstrates that it is especially suitable for very large databases. BIRCH incrementally and dynamically clusters incoming multi-dimensional metric data points to try to produce the best quality clustering with the available resources (i.e., available memory and time constraints). BIRCH can typically find a good

ISSN: 2277-9655
Impact Factor: 4.116
CODEN: IJESS7

clustering with a single scan of the data, and improve the quality further with a few additional scans. BIRCH is also the first clustering algorithm proposed in the database area to handle ”noise”

CONCLUSION

As the twitter data is growing on faster rate so to analyze the data we develop a system which helps in analyzing the large set of data , which can be used to predict the views of people on a particular topic.

ACKNOWLEDGEMENTS

We express our sincere thanks to Head of Department of Computer Engineering for her kind co-operation. We also express our sincere thanks to Prof. Chaitanya Bhosale.

REFERENCES

[1] SIAM Int. Conf. Data Mining, 2007, pp. 491496. [1] C. C. Aggarwal, J. Han, J. Wang, and P. S. Yu, “A framework for clustering evolving data streams,” in Proc. 29th Int. Conf. Very Large Data Bases, 2003, pp. 81–92.
[2] T. Zhang, R. Ramakrishnan, and M. Livny, “BIRCH: An efficient data clustering method for very large databases,” in Proc. ACM SIGMOD Int. Conf. Manage. Data, 1996, pp. 103–114.
[3] P. S. Bradley, U. M. Fayyad, and C. Reina, “Scaling clustering algorithms to large databases,” in Proc. Knowl. Discovery Data Mining, 1998, pp. 9–15.