                    Authors:

          Anjali Barskar, Ajay Phulre

Abstract

Twitter, one of the largest and famous social media site receives millions of tweets every day on variety of important topic. This large amount of raw data can be used for industrial , Social, Economic, Government policies or business purpose by organizing according to our need and processing. Hadoop is one of the best tool options for twitter data analysis and hadoop works for distributed Big data , Streaming data , Time Stamped data , text data etc. This paper discuss how to use FLUME for extracting twitter data and store it into HDFS for opinion mining because twitter contains variety of opinions on various topics so we have to analyse these opinions using hadoop and its ecosystems to check every tweets polarity either tweets contains positive ,negative or neutral opinions on particular topic. This paper provides an efficient mechanism to perform opinion mining by coming up with a finish to finish pipeline with the assistance of Apache Flume ,Apache HDFS, and Apache Pig.

Here we have used dictionary based approach for analysis for which we have implemented pig statements through which we can analysis these complex twitter data to check polarity of the

tweets based on the polarity dictionary through which we can say that which tweets have negative opinion or positive opinion.

References

1. Marco Furini, Manuela Montangero, "TSentiment: On Gamifying Twitter Sentiment Analysis", IEEE ISCC 2016 Workshop: DENVECT, IEEE 2016, ISSN: 978-1-5090-0679-3/16.

2. Rahul Kumar Chawda, Dr. Ghanshyam Thakur, "Big Data and Advanced Analytics Tools", 2016 Symposium on Colossal Data Analysis and Networking (CDAN), IEEE 2016, ISSN: 978-1-5090-0669-4/16.

3. Mahalakshmi R, Suseela S , "Big-SoSA:Social Sentiment Analysis and Data Visualization on Big Data", International Journal of Advanced Research in Computer and Communication Engineering, Vol. 4, Issue 4, April 2015 , pp 304-306, ISSN : 2278-1021.

4. Manoj Kumar Danthala, "Tweet Analysis: Twitter Data processing Using Apache Hadoop", International Journal Of Core Engineering & Management (IJCEM) Volume 1, Issue 11, February 2015, pp 94-102.

5. Manoj Kumar Danthala, "Bigdata Analysis: Streaming Twitter Data with Apache Hadoop and Visualizing using BigInsights", International Journal of Engineering Research & Technology, Volume. 4 - Issue. 05 , May – 2015.

6. Judith Sherin Tilsha S , Shobha M S, "A Survey on Twitter Data Analysis Techniques to Extract Public Opinion", International Journal of Advanced Research in Computer Science and Software Engineering, Volume 5, Issue 11, November 2015, pp 536-540.

7. Mr. Sagar Nadagoud, Mr. Kotresh Naik.D, "Market Sentiment Analysis for Popularity of Flipkart ", International Journal of Advanced Research in Computer Engineering & Technology (IJARCET), Volume4Issue5,May2015,pp 2117-2123.

8. Ramesh R, Divya G, Divya D, Merin K Kurian , "Big Data Sentiment Analysis using Hadoop ", (IJIRST)International Journal for Innovative Research in Science & Technology,Volume 1 , Issue 11 , April 2015 ISSN : 2349-6010

9. Sunil B. Mane , Sunil B. Mane, Yashwant Sawant, Saif Kazi, Vaibhav Shinde , "Real Time Sentiment Analysis of Twitter Data Using Hadoop", (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 5 (3) , 2014, 3098 – 3100 , ISSN:0975-9646.

10. Praveen Kumar, Dr Vijay Singh Rathore," Efficient Capabilities of Processing of Big Data using Hadoop Map Reduce", International Journal of Advanced Research in Computer and Communication Engineering Vol. 3, Issue 6, June 2014, pp 7123-7126.

11. G.Vinodhini , RM.Chandrasekaran, "Sentiment Analysis and Opinion Mining: A Survey" , International Journal of Advanced Research in Computer Science and Software Engineering, Volume 2, Issue 6, June 2012 ISSN: 2277 128X.

12. Aditya B. Patel, Manashvi Birla, Ushma Nair, "Addressing Big Data Problem Using Hadoop and Map Reduce",6-8 Dec. 2012.

13. Michael G. Noll, Applied Research, Big Data, Distributed Systems, Open Source, "Running Hadoop on Ubuntu Linux (Single-Node Cluster)", [online], available at http://www.michael-noll.com/tutorials/running-hadoop-on-ubuntu-linux-single-node-cluster/

Index Terms

Computer Science   Information Sciences

Keywords

Hadoop, twitter, Flume, opinion mining, social analysis, apache pig.