


# Effective Information Retrieval Framework for Twitter Data Analytics

Ravindra Kumar Singh, National Institute of Technology, Jalandhar, India\*

 <https://orcid.org/0000-0003-1142-1954>

## ABSTRACT

The widespread adoption of opinion mining and sentiment analysis in higher cognitive processes encourages the need for real time processing of social media data to capture the insights about user's sentiment polarity, user's opinions, and current trends. In recent years, lots of studies were conducted around the processing of data to achieve higher accuracy. But reducing the time of processing still remained challenging. Later, big data technologies came into existence to solve these challenges but those have its own set of complexities along with having hardware deadweight on the system. The contribution of this article is to touch upon mentioned challenges by presenting a climbable, quick and fault tolerant framework to process real-time data to extract hidden insights. This framework is versatile enough to support batch processing along with real time data streams in parallel and distributed environment. Experimental analysis of proposed framework on twitter posts concludes it as quicker, robust, fault tolerant, and comparatively more accurate with traditional approaches.

## KEYWORDS

Cache Management, Data Processing Framework, Message Broker, MongoDB, Parallel Processing, Python-dash, Real Time Analytics, Redis, Social Media Analytics, Visualization

## 1. INTRODUCTION

Social media has evolved with the phenomena that each individual in this virtual world is a social player and love to discuss their viewpoints on different domains (Zeng et al., 2010), such as politics and journalism (Stieglitz & Dang-Xuan, 2013), business (Beier & Wagner, 2016), sports and entertainment (Shen et al., 2016), science and technologies (Baars & Kemper, 2008), etc. on public forums day in day out without bearing any cost of membership. The most lucrative virtue of social media is its real time wide range of acceptableness of user's posts in multiple languages across the domains about various events, situations, feelings, opinion, day to day thoughts, feedback, shopping experiences, best wishes, knowledge sharing, advice, wish lists, future plans or anything out of the imaginations. These posts may be in form of text messages, images, videos, emoji, maps etc and accepts the reactions of other users on it as well as extending the features to be shared on multiple

DOI: 10.4018/IJIRR.325798

\*Corresponding Author

This article published as an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>) which permits unrestricted use, distribution, and production in any medium, provided the author of the original work and original publication source are properly credited.

such portals to attain more and more engagements. It leads the wide acceptability and exponentially grown popularity among the users, therefore produces immense quantity of data (Azab & ElSherif, 2018). Social media platforms are capturing the associated metadata along with users post to better define the matter of the post.

According to a survey most of the online users invest their couple of hours in online activities like social media networking, shopping, banking, reading, coding, playing, watching videos, etc. This increasing interest of users in online activities accelerated the growth of online generated data up-to 2.5 quintillion bytes per day (BSA, 2015). According to a study of IBM, approximately 35 zettabytes of data would be generated annually by fall of 2020 (2012). Besides, in every second data are produced continuously; 34,722 Likes on Facebook, about 571 new websites, and almost 175 million tweets, billions of transaction logs, etc, accumulated all these data is known as Social Media Big Data (Lynn et al., 2015). This growing set of user generated data brings the opportunities for understanding and analyzing several aspects, patterns, trends and sentiments (Kim et al., 2016) and extract useful insights from it to take the edge in this competitive world (Oh et al., 2016).

Social media platforms have been established in recent years as sources of trends, feedback, debate and sentiments on across the domain including politics (Khatua et al., 2015), journalism, business (Kurniawati et al., 2013), entertainment, sports, etc. Business leaders across the domains are curious enough to use the real time insights from social media data in their policy and decision making, moreover to extend their support to gain better customer satisfaction (Nulty et al., 2016). Therefore “Social Media Analytics” coined as a special term and all these activities started from collecting posts to storing, analyzing till insight extraction and visualization fall under its umbrella (Stieglitz et al., 2014). All these factors made social media a separate world in itself among the internet domain, considerably influencing both academic and industrial outlooks (Batrincea, 2015).

## **2. BACKGROUND**

Social media analytics have gained foremost attention (Tsantarliotis & Pitoura, 2017) since last couple of years in researcher’s world for sentiment analysis (Kane et al., 2014) and opinion mining (Maynard et al., 2012). Researchers are more focused on real-time data (Jose & Chooralil, 2015), by using streaming application programming interface (API) of social media platforms like twitter (Tweepy, ), facebook, reddit, etc. to grasp the current trends and public opinions. Social media networking portals are operating on ideology of web 2.0 that defines that data would be created and updated by the users in collaborative manner rather than just being published by individuals who owns it (Kaplan & Haenlein, 2010). Therefore social media data are vague and not bound by any rule or specified formats (Hogenboom et al., 2011), so the research on this data arises the need of some framework to better define the steps of processing and could minimize the complexities in the research.

### **2.1 Social Media Analytics Framework**

There are few frameworks to demonstrate and build a typical basis for conducting social media analytics. It is a multi step journey that contains data collection, data processing, data analysis and lands up with presentation and visualization of the outcome. A better flow is defined in the CUP framework (Fan & Gordon, 2014) which defines it in 3 phases-

- Capture is the process of data aggregation and applying pre-processing on it to get the better shape to the data.
- Understand is the process of applying various analytics per the use case on the captured data post handling noisy and skewed data to attain the insights of the analytics.
- Present is the process of evaluating and summarizing the findings.

19 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage: [www.igi-global.com/article/effective-information-retrieval-framework-for-twitter-data-analytics/325798](http://www.igi-global.com/article/effective-information-retrieval-framework-for-twitter-data-analytics/325798)

## Related Content

---

### Illustration and Validation of the Interactive IR Framework

Iris Xie (2008). *Interactive Information Retrieval in Digital Environments* (pp. 263-293).

[www.irma-international.org/chapter/illustration-validation-interactive-framework/24530](http://www.irma-international.org/chapter/illustration-validation-interactive-framework/24530)

### Answering Top-k Keyword Queries on Relational Databases

Myint Myint Thein and Mie Mie Su Thwin (2012). *International Journal of Information Retrieval Research* (pp. 36-57).

[www.irma-international.org/article/answering-top-keyword-queries-relational/78313](http://www.irma-international.org/article/answering-top-keyword-queries-relational/78313)

### A Review of Answering Queries over Ontologies Based on Databases

Fu Zhang, Z. M. Ma and Haitao Cheng (2016). *Handbook of Research on Innovative Database Query Processing Techniques* (pp. 479-498).

[www.irma-international.org/chapter/a-review-of-answering-queries-over-ontologies-based-on-databases/138711](http://www.irma-international.org/chapter/a-review-of-answering-queries-over-ontologies-based-on-databases/138711)

### Speech Emotion Analysis of Different Age Groups Using Clustering Techniques

Hemanta Kumar Palo, Mihir Narayan Mohanty and Mahesh Chandra (2018). *International Journal of Information Retrieval Research* (pp. 69-85).

[www.irma-international.org/article/speech-emotion-analysis-of-different-age-groups-using-clustering-techniques/193250](http://www.irma-international.org/article/speech-emotion-analysis-of-different-age-groups-using-clustering-techniques/193250)

### Venkatesh et al.'s Unified Theory of Acceptance and Use of Technology (UTAUT) (2003)

Manishankar Chakraborty and Salim Al Rashdi (2015). *Information Seeking Behavior and Technology Adoption: Theories and Trends* (pp. 220-236).

[www.irma-international.org/chapter/venkatesh-et-als-unified-theory-of-acceptance-and-use-of-technology-utaut-2003/127134](http://www.irma-international.org/chapter/venkatesh-et-als-unified-theory-of-acceptance-and-use-of-technology-utaut-2003/127134)