



Data Source Exploration and Visualization in Electronic Social Media for Sports Industry

Shwetha G S¹, Chandrasekhar B H²

Student, Department of MCA, Rashtreeya Vidyalaya College of Engineering, Bengaluru, India¹

Associate Professor, Department of MCA, Rashtreeya Vidyalaya College of Engineering, Bengaluru, India²

ABSTRACT: Data source exploration which analyses different data sources, extract data (via SQL scripts), clean them up and streamlines data to be fit for Tableau visualizations. The data sources can be data from network, data from third party websites, from social media platforms like twitter/Facebook, and from YouTube. This involved usage of querying tool like DB visualizer, Spyder for cleaning data using Python and Big Query to visualize data in Google big tables. Here we mainly focus on extracting data from social media network for analysing the “Major Sports Entertainment Industry” data, by considering the different metrics involved. Main aim is to analyse data according to user requirement. Made use of spyder for cleaning data using python, to accommodate data from more data points and process them faster. Hence there is a need for data to be consistent across different layer of data flow.

Social network analysis enables segment data based on the user behaviour, understand the natural group that have formed around interconnections: topics, personal characters (locations, interests). The fundamental building block of social network analysis is relationship. We analyse data derived from social networking sites, third-party data and network data. The result will be represented using tableau to ensure the veracity of data being visualized in tableau. Representation of data using graphs, charts and maps which helps in easy understanding the represented data. The proposed framework will help in analysing the social media data and helps in creating the interactive Tableau Dashboards.

KEYWORDS: Social Media data: Metrics- Engagements, Opportunities, Tableau-Visualization, and analysis of data

I. INTRODUCTION

Data Source Exploration is an initial step of data analysis which includes visualization and calculation to better understand the characteristics of data. These characteristics would include data size or quantity, data completeness, data quality, potential data element relationships or files / tables in the data. Getting some data, can export directly from the platform like from Facebook, Twitter, Instagram, and YouTube, exporting from a third party where it includes extracting data from an integrated social media platform and getting live data from the third party that is database where in getting robust data in a database. There is a possibility of coding their own solutions using an API and tableaus web data connector.

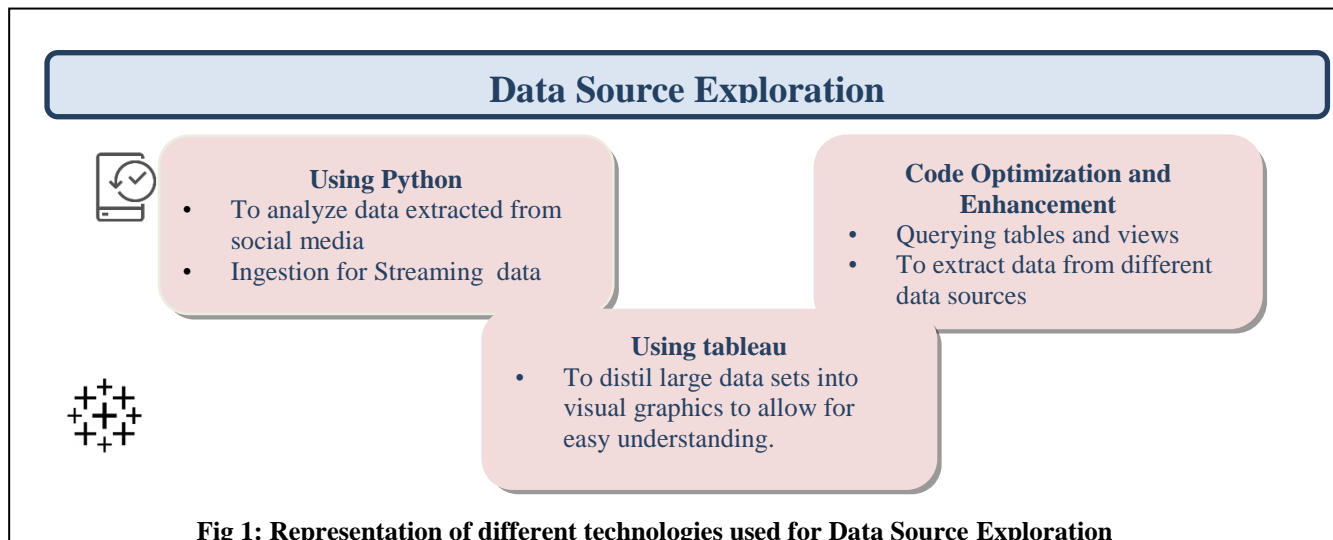
In previous years, across all sources of data, social networking sites included in purpose of the study. In Social Media Data extraction there will be a main focus on some of the metrics. There are different Categories of metrics in which it would be relevant for goals: i) Benchmarks ii) Audience iii) Engagements iv) Conversions v) Opportunities. The first category of metric is Benchmarks it includes-Followers, Mentions, then any category that help in understanding Audience, includes- Impressions, reach, demographics, location, timing. Engagements is all about understanding once they reach to Audience how they engage with them, Engagements includes-likes, comments, views. The actual business impact when posting on social media: conversions which helps to track any metrics such as clicks/leads. There is a little bit difference, can term it as opportunities helps in reaching more people that contains user-generated links, related-hash tags (those are mentioned in tweets).

Extracting data from third party websites and from network includes pdf, database, and text file. Collaborative environments: a forum for various users (e.g. Wikipedia, Wikihow) to generate content simultaneously, Internet news: Website containing user-posted news that others can vote on (e.g., Slashdot, Digg), Online communities and player world: Application that replicates a Virtual object in which users appear as personalized animations and interact with each other (e.g. virtual life, Universe of Battle crafts).

In the prevailing world every organization relays on taking data driven decision. In order to make data driven decision data need to be studied and analysed to identify the factors as mentioned in Fig 1: Tableau is used as best visualization tool as it provides various interactive graphs, diagrams to represent it in dashboards. We can create an interactive worksheets and present two are more worksheets in one dashboard by creating the timeline, can also include different parameters and KPI's in it. With the simple drag and drop interface tableau writes optimized SQL queries at back end to fetch data from data sources in the desired way. And also with its high end graphics and colour schemes



tableau makes the charts to look as apt as possible. Using tableau one can plot Bar chart, pie chart, bubble chart, heap/tree map and even scattered plot on a real geographical map, and also can drill down the data using filters and visualize data at Granular level.



II. RELATED WORK

In [2] authors various data extraction techniques have been used in [2] writers: I HTML parsing utilizing java code: HTML parsers typically are using the computer language Java. In java, HTML parser performs the parsing process using `htmllexer.jar` or `htmlparser.jar`. JSoup could also be used to read and extract HTML documents from certain documents. JSoup is indeed a Java library that does real-time work with an HTML. ii) HTML parse using python: BeautifulSoup is a python library for extracting unique content from web sites, extracting HTML tags and storing information. It is among the information extraction methods commonly which provide resources such as cleaning and decoding of collected documents. In [3] authors When evaluating the tools involved in management of existing systems, they rely on examining the particular characteristics (similarities) demonstrated by the resources concerned. They took full advantage of the system data generated during the running phase in form of a process log. They define similarity of tools, based on the circumstances. The background is considered as a system of method actors relevant to the studied tool.

In [4] authors Had carried out several studies in the observation: I Studied business case: the procedure examined is as follows: the accountant first produces the document, validates it, then sends it to the backers and finally, after returning it, he / she prefers to post or change the approver responsible (ii) data preprocessing of analyzed data and results. The proposed method consisted of the following steps: (a) Identifying and gathering the original logs- the purpose of this step was to locate source logs and convert them to an unified log with the framework. (B) Cleaners and sorting of data-Washing and sorting of data-This system ensured that the report is cleared from lost records and contains only relevant data. (C) Anonymization of data-All log data is anonymized so that no private details can be accessed during study.

In [5] results found a leading internet network has become the cornerstone of emerging technology in the last couple of years. It provides detailed information about individuals by the use of potentially minable profiles. This leads to the related security and ethics difficulties. The data collection industry is a big one-phase spreading mechanism that exists in response to differences. That's not the only way to evaluate collected data. Using edges, scales, data collected can be converted into a graph representing the structural sense of the tests. Digital social media platforms offer excellent map definition and the possibilities of data mining for a range of groups from different fields.

Generally popular process for extracting useful social networks from of the internet is the casual method [6],[7] which utilizes browsers to obtain website content relevant to the participant actors, and then calculates the metric of co-occurrence [1]. Throughout the dynamic sense, this approach collects information about links among all pairs (dyads) where a connection links a pair of participants across one or more relationships, thus giving full system data a full view of community relationships. In addition, a networking site could be very simply modelled by a given Graph, where SNA uses a method to imagine a seems at of vertices throughout V as a set of actors and edge e as just a set of relationships for all [8]. In statistical literature, the researchers model a social network as a Cartesian product of n actors for generating their relation.



III. METHOD OF EXTRACTING DATA

The structure of implementation flow is illustrated in Fig.2 there are four major components to the process:

- 1) Apply verification
- 2) Data Retrieval
- 3) Data Transformation
- 4) Review of the information

The various series of stages:

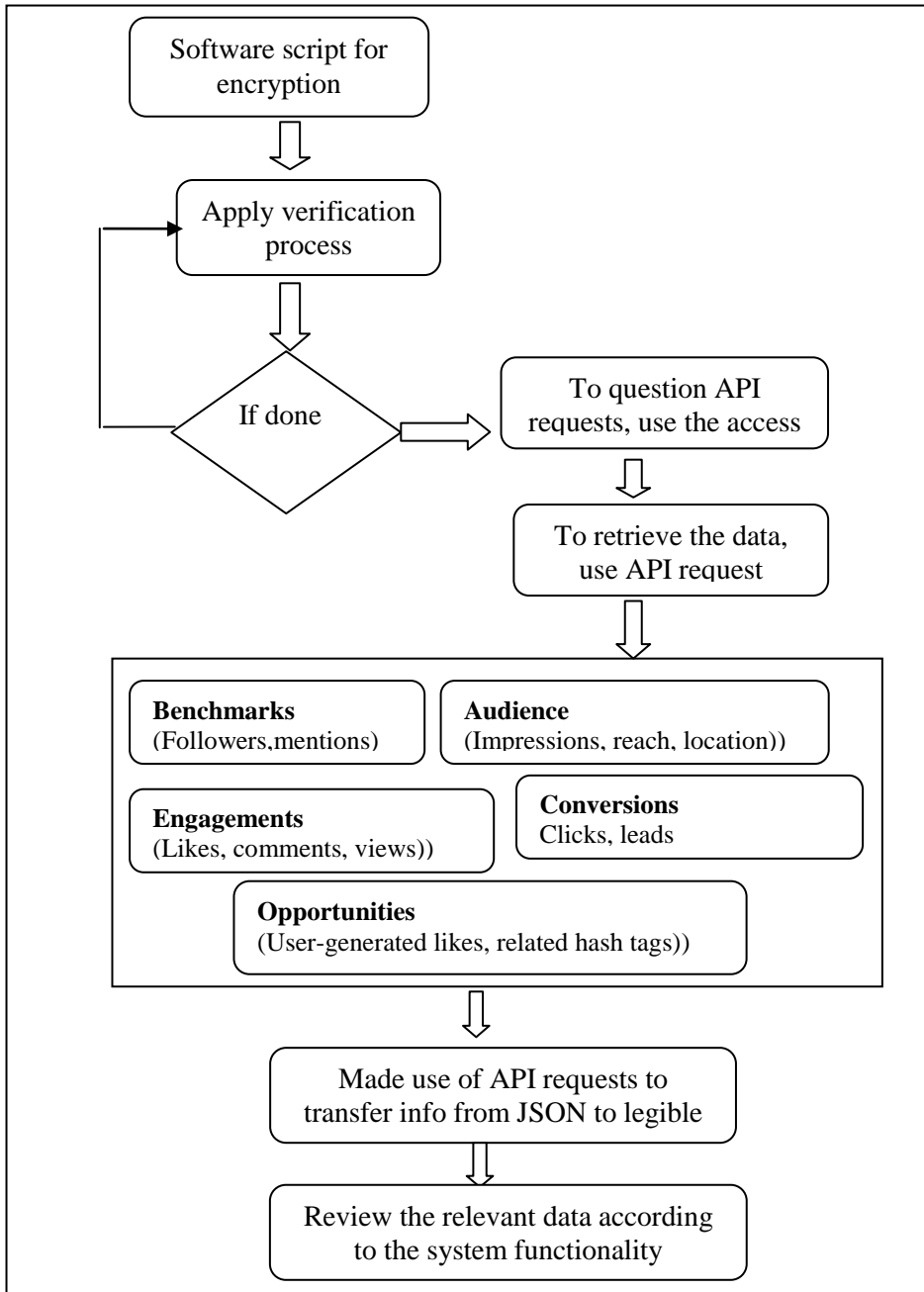


Fig.2. Structure of implementation flow

Step 1: Apply verification

Verified queries are essential to manage API request of SNS. You should submit each and every query using the correct authentication tokens.

The core idea for the structure of implementation is shown in Fig. 3:



Fig.3. structure of the verification process

A. OAuth

It is indeed a configuration of the accessible encryption mechanism described by the Internet standard Oath TF (Task Force) that allows authorized users to access the data through one another. It is an universal method for encryption that Twitter/Facebook supports in addition to be linked to secure information as well as the procedure is performed with a three different way surround.

A user 'requires a key' from either the web application component i.e. Password client and the use the key to verify to some other application component i.e. information processor, these are the details which the user needs to accessor exploit. [10]OAuth provides 3rd parties verification methods which allow web applications to transfer information along all their APIs.

The sections describes if verification method are listed as:

ENCAPSULATES THE METRICS IMPLICATED IN THE APPLYING OAUTH FOR AUTHENTICATION PROCESS FOR CLIENT AND VERSION OF FILES

1. Create the code to use the data from twitter.
2. Private information and client Button problem web application.
3. Application system utilizes to test its passwords for digital signature.
4. Browser server monitors the client.
 - Unless the permission (customer signs in and is accurate) is provided link (voucher via OAuth)
 - Wrong.
 Take Stage 3
5. Confirmation link used it to request the OAuth identifier i.e. Flip Straps.
6. System codes also use PIN and CSK.
 - In which the strap is accurate, SK concerns such as direct connections to credential and hidden access keys.
 - Wrong.
 Stage 5
7. User app uses SK to retrieve necessary data.

Stage 2: Retrieval of the SNS details

When the verification step is finished we will retrieve the data depending on our design specification.

The API seems to be the main way to obtain and entering public data exchange pages. The HTTP-based API is often used to access info, upload new stories, build check-ins, and any of the other tasks and apps will be doing so. Twitter needs two REST-APIs and Filter APIs, whereas Facebook number of buses graphs [11] API which aims to assess new objects in a user's social network context, and helps to create class labels of such behaviour and actions are just like Twitter APIs via graph

Social media group like Facebook/Twitter is a list of coordinates. From of processing shall be included, depending mostly on form of data posted by users. The information on the social media/Twitter website are structured since the same data provided by the user are connected in unreadable text. We submit the request to have the Facebook/Twitter web page material available in JSON format. Stage 3: API conversion tactic, as mentioned earlier in this thread, agreed credit as Xml format which makes it very difficult for clients/ developers to interact with/read the data. Below demonstrates simple JSON object creation code for carbon material.

EXCERPT SCRIPT TO TRANSLATE JSON TO XML

```
JSONObject jsonObject = new JSONObject (json.toString ()); System.out.println (XML.toString (jsonObject));
```



Step 4: Review of the information

Just after data is collected, the descriptive process can be done as needed. Work is program oriented towards to the user.

IV. EXPERIMENTAL RESULTS

The Experiment conducted on “Major Sports Entertainment industry” where in we extracted and analyzed data from twitter, YouTube and Facebook based on the relevant metrics mentioned above. The analysis is done according to the user requirement. Here we made use of Tableau for visualization of data which made easy understanding for the user to Understand represented data. Data will be extracted from social media platform using python and converted to pandas data frame; those data frame will be represented using Tableau. As per the user requirements data will be analyzed.

| Talents Ranked by twitter Mentions | | | | | | | |
|------------------------------------|------|------|--------|----------|-------------|-------------|-----------------------|
| June Followers | Rank | | Talent | Mentions | %Change MoM | %Change YOY | Share of top 20 Views |
| | May | June | | | | | |
| 9 | 2 | 1 | A | 232357 | 20% | 67% | 13% |
| 22 | 1 | 2 | B | 222027 | -24% | 106% | 11% |
| 6 | 2 | 3 | C | 242970 | -17% | 11% | 7% |
| 26 | 8 | 4 | D | 114412 | 44% | -42% | 7% |
| 22 | 4 | 5 | E | 105488 | -17% | 8% | 5% |
| 1 | 10 | 6 | F | 102795 | 14% | -63% | 5% |
| 25 | 12 | 7 | G | 25502 | 22% | 243% | 5% |
| 26 | 11 | 8 | H | 90056 | 6% | 22% | 4% |
| 2 | 17 | 9 | I | 75088 | 24% | 12% | 4% |
| 18 | 8 | 10 | J | 74175 | -25% | 140% | 4% |
| 22 | 50 | 11 | K | 71234 | 174% | 111% | 4% |
| 46 | 18 | 12 | L | 63908 | 15% | -25% | 4% |
| 55 | 27 | 13 | M | 62333 | 5% | 123% | 2% |
| 127 | 12 | 14 | N | 62123 | -22% | 143% | 2% |
| 21 | 18 | 15 | O | 60234 | -23% | -51% | 2% |
| 55 | 24 | 16 | P | 55678 | 24% | -20% | 2% |
| 53 | 26 | 17 | Q | 54234 | 14% | 53% | 2% |
| 25 | 6 | 18 | R | 44987 | -15% | 120% | 2% |
| 125 | 52 | 19 | S | 41888 | 142% | 123% | 2% |
| 62 | 58 | 20 | T | 40124 | 145% | 14% | N/A |

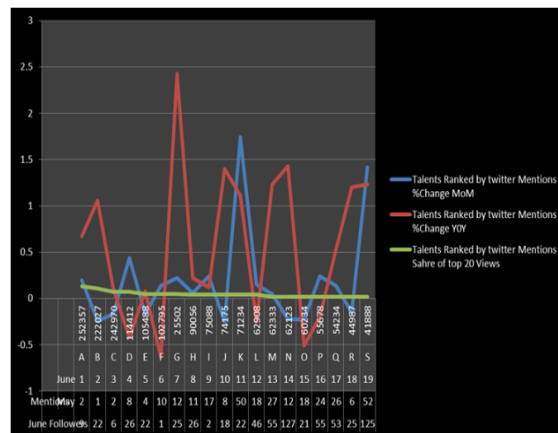


Fig.3. Twitter data analysis based on mentions

In the above Fig 3: it represents the data extracted from twitter based on mention, as mentioned experiment is conducted on Sports Entertainment Industry, here players(Represented as Talents in Fig) performance will be analysed and based on the twitter mentions they will be given higher ranks. As displayed in Fig. 1.

Example:

- Talent I rises in rank from #17 in May to #8 in June, highlighting a 24% MoM(Month on Month) increase in twitter mentions

| Ranked by Facebook Video vies | | | | | | | |
|-------------------------------|------|---|--------|----------|-------------|-------------|-----------------------|
| May | Rank | | Talent | Views | %Change MoM | %Change YOY | Share of top 20 Views |
| | June | | | | | | |
| 1 | 1 | 1 | A | 11024378 | -57% | 10% | 23% |
| 4 | 2 | | B | 8255068 | -12% | 1384% | 13% |
| 38 | 3 | | C | 3337738 | 550% | 175% | 8% |
| 9 | 4 | | D | 3205842 | 100% | 1043% | 7% |
| 6 | 5 | | E | 3033375 | -22% | 1150% | 6% |
| 12 | 6 | | F | 288284 | 24% | 1003% | 6% |
| 50 | 7 | | G | 2709300 | 1283% | N/A | 6% |
| 10 | 8 | | H | 1199434 | 58% | 23% | 3% |
| 8 | 9 | | I | 1700042 | -2% | 74% | 4% |
| 3 | 10 | | J | 1883303 | -89% | 324% | 4% |
| 11 | 11 | | K | 1284303 | 8% | 338% | 3% |
| N/A | 12 | | L | 1237388 | N/A | 45092% | 3% |
| 15 | 13 | | M | 1288888 | 115% | 790% | 2% |
| 2 | 14 | | N | 1230662 | -92% | N/A | 2% |
| 5 | 15 | | O | 1088452 | -90% | 342% | 2% |
| 24 | 16 | | P | 848387 | -9% | 428% | 2% |
| 25 | 17 | | Q | 756487 | 100% | 83% | 2% |
| 23 | 18 | | R | 722708 | 299% | 432% | 2% |
| 30 | 19 | | S | 377248 | 32% | N/A | 1% |
| 24 | 20 | | T | 347375 | 52% | 37% | 1% |

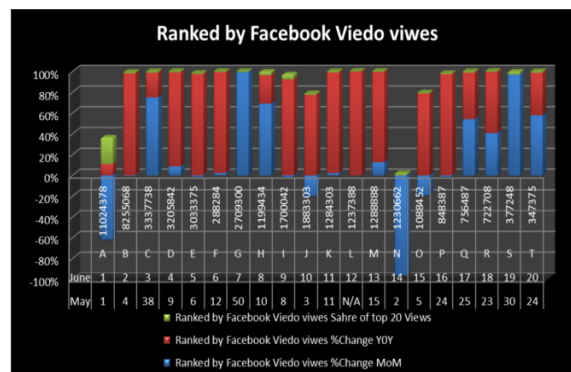


Fig.4. YouTube data analysis based on video views

In the above Fig 4: it represents the data extracted from YouTube based on video views, as mentioned experiment is conducted on Sports Entertainment Industry, here players(Represented as Talents in Fig) performance will be analysed and based on YouTube video views they will be given higher ranks. As displayed in Fig. 2.



Example:

- Talent C rises in rank from #38 in May to #3 in June, highlighting a 550% MoM(Month on Month) increase in YouTube video views.

| TOP 5 JUNE UPLOADED VIDEOS BY YOUTUBE VIDEO VIEWS | | | | | |
|---|------------------------|---------------|-----------------------|----------|---------------------------------|
| Rank | Video Title | Uploaded date | Video Length(minutes) | views | Average views duration(minutes) |
| 1 | video XXXXX June3,2019 | June 3,2019 | 3:33 | 18072066 | 2:18 |
| 2 | video YYYYY June 20,19 | June 24,2019 | 2:28 | 11598487 | 1:34 |
| 3 | video zzzz | June 12,2019 | 22:03 | 8811204 | 6:18 |
| 4 | video aaaa | June 6,2019 | 3:03 | 8808004 | 2:00 |
| 5 | video bbbb | June 17,2019 | 3:33 | 8419031 | 1:48 |

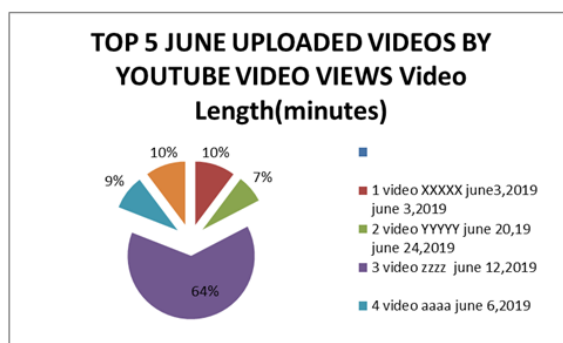


Fig.5. YouTube video analysis based on video views

In the above Fig 5: it represents the data extracted from YouTube based on video views, as mentioned experiment is conducted on Sports Entertainment Industry, here we analysed the top 5 June uploaded videos by YouTube video views.

Example:

- Top 5 videos of June are ranked based on the views, the Fig.3. Includes video title, the uploaded date, and video length in minutes, views and average views duration in minutes.

V. CONCLUSION AND FUTURE WORK

These days social networking sites is seen to be a phenomenon that has combined mechanism, social contacts and terms,images, audios and tapes. It helps to access Family, app accounts and messages. The extraction process and analysis is a difficult function of social networks, as the information is ambiguous and the content is not organized at some places.This study is primarily to support the users knowing the measures by which material is derived from specific sites and analyzed data using Tableau as per the user requirements. Extracted data from different social media for “Major Sports Entertainment Industry” by considering different metrics and give them a higher ranking.

In coming days data will be analyzed by considering data from other data sources such as from data from third party and data from networks. Compare the data extracted from different data sources and analyze the best data source to be considered for analyzing data.

REFERENCES

1. MominSaniyaParvez, Shivankar SnehaRajendra, Khan ShaistaAgahTasneem, Kalpana R. Bodke, ” Analysis Of Different Web Data Extraction Techniques” in 2018 International Conference on Smart City and Emerging Technology (ICSCET)
2. Martin Kopka, Milos Kudelka, JakubStolfa, OndfejKoberski) and VACLAV Smisel, “Extraction and analysis social networks from process data” in 2013 Fifth International Conference on Computational Aspects of Social Networks
3. RuqayyaAbdulrahman, Sophia Alim, Daniel Neagu and Mick Ridley, “Algorithms for Data Retrieval from Online Social Network Graphs” in 2010 10th IEEE International Conference on Computer and Information Technology (CIT 2010)
4. D. Hemavathi, M. Kavitha, Narjiya Begum Ahmed, “Information extraction from social media: Clustering and labellingmicroblogs” in 2017 International Conference on IoT and Application (ICIOT)



5. Narashima S. Purohit, Akshata B. Angadi, MeghanaBhat, Karuna C. Gull, “Crawling through web to extract the data from Social networking site – Twitter”, 2015 National Conference on Parallel Computing Technologies (PARCOMPTECH)
6. Danah M. Boyd and Nicole B. Ellison, “Social Network Sites:Definition, History, and Scholarship”, Journal of ComputerMediated Communication, Vol 13, Issue 1, pp 210–230, doi:10.1111/j.1083-6101.2007.00393.x, Oct 2007.
7. Brad Dinerman, “Social networking and security risks”, in GFI White Paper ,2011, pp.1-8.
8. Yan Guo ,Hui Feng Tang, Linhai Song , Yu Wang and Guodong Ding, “ECON: An Approach to Extract Content from Web News Page”, 12th International Asia-Pacific Web Conference,2010.
William W. Cohen, “Learning and Discovering Structure in Web Pages”, Bulletin of the IEEE Computer Society Technical Committee on Data Engineering.
9. Deng Cai, Shipeng Yu, Ji-Rong Wen and Wei-Ying Ma,“Extracting Content Structure for Web Pages based on Visual Representation”