DYNAMICS OF LIBRARY FOR EXCELLENCE IN ELECTRONIC REVOLUTION

Editors :

P Visakhi

P K Jain

Debal C Kar

Parveen Babbar

Tools and Technologies for Media Libraries in the era of Big Data

Muruli N
Assistant Librarian
Central University of Himachal Pradesh, Dharamshala, India
E-mail: murulitarikere@gmail.com

Abstract

The term "Big Data" usually refers to describe to data sets that exceed the ability of traditional tools to store, aggregate and process. Big Data analytics is one of the latest trends to analyse, organize and report the massive volume of data or information. Objective of this paper is to highlight the tools and techniques which can be used by Librarians working at Media field and by other professionals. This paper attempts to help the Librarians, reporters, data journalist and others to understand the concept of Big Data, Big Data analytics and the use and characteristics of different technologies that can help them in reporting news or stories easier with the right speed within the right information and with the right framework. This paper is a case study of Google Media Tools and other tools which can be used by the Media journals for better data collection, organization and for reporting.

Keywords: Big Data, Analytics, Librarian, Journalist, Media, Google,

1. Introduction

Big data is a popular term for massive volume of both structured and unstructured data that is large in volume and difficult to process in traditional techniques. It includes the challenges of capture, analysis, data curation, and visualisation and information privacy. The "Big Data" term often refers simply to the use of predictive analytics or other certain advanced methods to extract value from data, and seldom to a particular size of data set. Accuracy in big data may lead to more confident decision making. And better decisions can mean greater operational efficiency, cost reduction and reduced risk.

An example of big data might be petabytes (1,024 terabytes) or Exabyte's (1,024 petabytes) of data consisting of billions to trillions of records of millions of people all from different sources (e.g. Web, sales, customer contact centre, social media, mobile data and so on). The data is unstructured data that is often raw and incomplete.

2. Definition

The term "Big Data" is usually refers to describe to data sets that exceed the ability of traditional tools to store, aggregate and process. Big data is a set of techniques and technologies that require new forms of integration to uncover large hidden values from large datasets that are diverse, complex, and of a massive scale.

According to Oxford English Dictionary Big Data means "Extremely large data sets that may be analysed computationally to reveal patterns, trends, and associations, especially relating to human behaviour and interactions".

Gartner defines "Big Data represents the Information assets characterized by such a High Volume, Velocity and Variety to require specific Technology and Analytical Methods for its transformation into Value".

3. Characteristics

Analysts have identified the following "5V" fundamental characteristics of Big Data.

- **Volume** - The quantity or the size of the Big Data is "Big"
- **Variety** - Big Data contains different types of contents or data.
- **Velocity** - Speed of the data is generated decides Big Data.
- **Variability** - The inconsistency is another character of Big Data.
- **Veracity** - The quality of captured data varies greatly.

4. Big data analytics

Big Data analytics is one of the latest process which collects massive volume of data or large sets of data to understand the hidden trends and other useful information that can be used by the decision makers to make better decisions. Data Scientists, journalist and other professionals can use data analytics to analyze huge sets of data that can't be processed by the traditional methods of analytics. Big data analytics follow Data - Information - Knowledge – Wisdom Cycle.

5. Types of Big Data Analytics

There are four types of big data analytics:

6. Big Data analytics tools

Big Data is not really a new concept, data scientists, information managers and knowledge managers have already worked with massive volume of data, they had computers, databases and data warehouses to manage large sets of data which is raw and incomplete. And now they have understood the importance of Big Data analytics tools that collect, analyse the huge number of raw data in a systematic way to provide real time analysis. Nowadays,

number of Big data tools are available in the market. Interestingly, many of the tools available are open source tools. The top 50 Open source Big data tools are described in more details in an article by Datamation - http://www.datamation.com/data-center/50-top-open-source-tools-for-big-data-1.html

7. Tools and Technologies for Media Librarians

Being a Media Librarian or data journalist most of the cases one has to work in time bound frame work and often need to deal with the huge volume of data. He or she has to collect the data from the different sources, need to organize and finally need to report it in a standard format. Here I have attempted to list few tools which can help the librarians or journalists to help in their day to day activities and improves their productivity and effectiveness of News Hour shows.
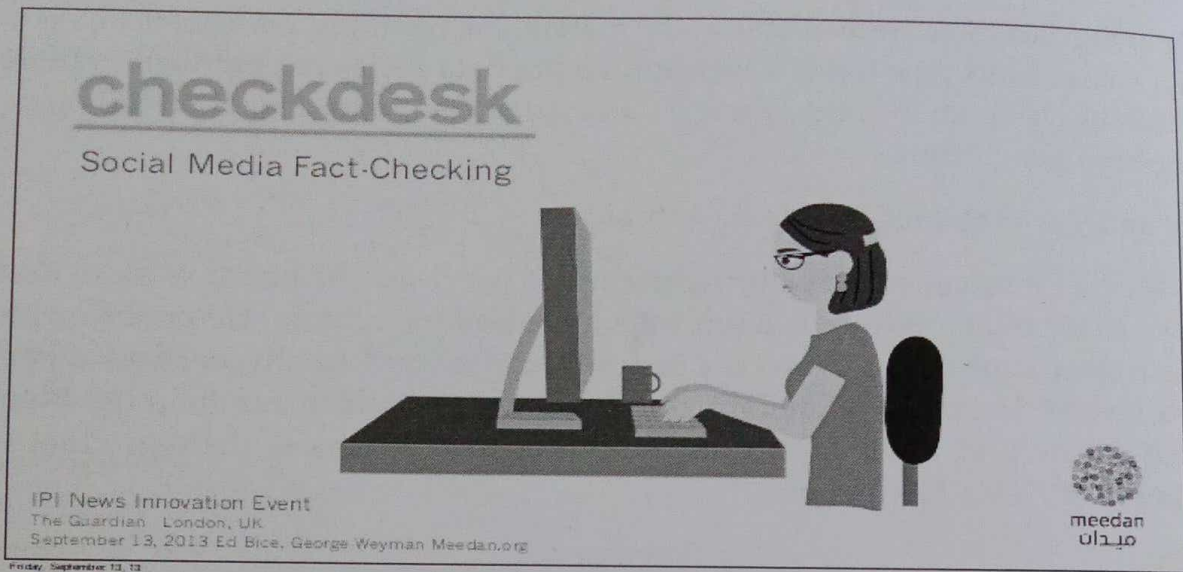
7.1 Tools

a) The spreadsheets

b) Data cleaning tools

c) Visualization tools

d) Mapping software

e) Scripting language

f) Web framework

g) A Flexible editor

h) Revision control

i) Document analysis tools

7.2 Technologies

a) Checkdesk

Checkdesk is an open source application that enables open and participatory investigative journalism for professional newsroom in the Arab region. Checkdesk was founded in 2011 and in 2016 started to build online tools, support journalist that aims to improve investigate the quality of citizen journalist. Checkdesk mainly works as verifying online breaking news spread over the social media. It helps the journalists or the media libraries to verify or cross check the information which they see on social networking sites. There are many chances to people who take the false rumours seriously. With the help of Checkdesk any citizen can upload their pictures or videos or information for verification. Once verified they can share the same with everywhere. [* File contains invalid data | In-line.JPG *]

checkdesk
Social Media Fact-Checking

IPI News Innovation Event
The Guardian London, UK
September 13, 2013 Ed Bice, George Weyman Meedan.org

meedan
ميدان

Friday, September 13, 13

b) Storify Storify is a social networking service that collects or create stories using multiple social media sites such as Instagram, Facebook, Twitter in one place, along with other social networking sites that are important to collect the news or information. Storify was founded in 2010, and since 2011 it open for public. In Sept 2013, it was acquired by Livefyre and currently it is the part of Adobe. Storify helps data journalists to search through multiple social networking sites and to create the content or news in drag and drop interface. Storify can be used by media libraries or by the Media Organizations to coverage ongoing or live events such as Social Movements, Elections, Breaking News, Meetings and other events. Storify has used by CBC to cover London Roits in 2011 and TRT World also used it to report on general elections 2015.
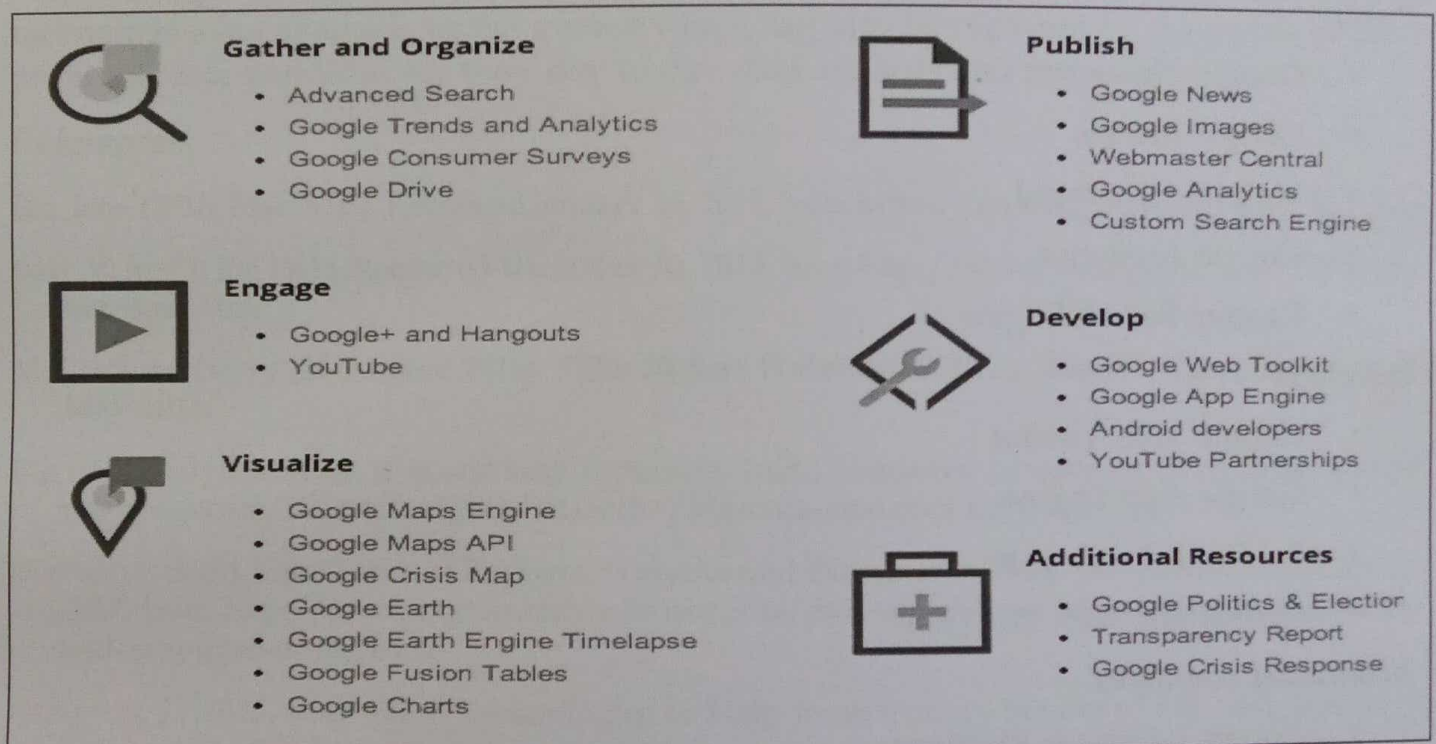

c) Citizen Desk

Citizen Desk is freely available open source software built by the journalist for journalist. It is developed or designed to help citizen journalist to build and communicate with network teams. Citizen Desk helps citizen journalist as well as the professional journalists to verify the information before they post or share. Citizen Desk is user friendly application or a digital tool which allows media libraries or media organizations to create mobile optimised platform for generating, verifying and publishing news. @Verdade a Portuguese Online journal with the help of Citizen Desk gathered and published 1,100 reports in 10 days on their election live blog which had helped Citizen Desk to won the Prestigious "African News Innovation Challenge" award in 2013.

Citizen Desk
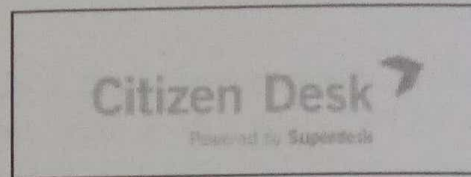
d) Google Media Tools

Google Media tools are designed to help journalist to make newsgathering and reporting as easy as possible. It is empowering the media organization or the journalist to connect with their audience and networks. Google media tools developed as centralized hub for various Google media products like Advanced Search, Google Trends, Google News, Google Plus etc.

Gather and Organize
- Advanced Search
- Google Trends and Analytics
- Google Consumer Surveys
- Google Drive

Engage
- Google+ and Hangouts
- YouTube

Visualize
- Google Maps Engine
- Google Maps API
- Google Crisis Map
- Google Earth
- Google Earth Engine Timelapse
- Google Fusion Tables
- Google Charts

Publish
- Google News
- Google Images
- Webmaster Central
- Google Analytics
- Custom Search Engine

Develop
- Google Web Toolkit
- Google App Engine
- Android developers
- YouTube Partnerships

Additional Resources
- Google Politics & Election
- Transparency Report
- Google Crisis Response

Google Media Tools are divided in to 6 groups:

Collect and Organize

- Advanced Search
- Google Trends and Analytics
- Google Consumer Surveys
- Google Drive

Engage

- Google+ and Hangouts
- YouTube

View

- Google Maps Engine
- Google Maps API
- Google Crisis Map
- Google Earth
- Time lapse of Google Earth Engine
- Google Fusion Tables
- Google Charts

Public

- Google News
- Google Images
- Webmaster central
- Google Analytics
- Custom Search Engine

Develops

- Google Web Toolkit
- Google App Engine
- Android developers
- Partnership with YouTube

Additional resources

- Google Politics & Elections
- Transparency Report
- Google Crisis Response

Google Plus and Hangouts have been used by The New York Times to interview US Secretary of State John Kerry about Syria's Chemical Weapons. The Weather Channel used Google Earth to illustrate the damage of Super storm Sandy through dramatic before and after satellite images. Google Trend tool is used by NBC's one of show to give audience an update on trended news over the weekend. Using Google Media Tools, Media organization or the LIS professionals or journalists can be the first to learn what users, readers or the viewers are talking about now, rather than waiting for other secondary sources.

8. Conclusions

Data analytics is not really a new concept we have been working, organizing, analysing data in traditional methods from many years. When we need to handle or analyse huge volume of data sets which is mostly raw and incomplete requires systematic tools to analyse and manage. Now, the professionals working with massive volume of data have understood the importance of Big Data analytics tools and other tools that collect, analyse the data in a systematic way to provide real time analysis. Nowadays, number of Big data tools are available in the market. Interestingly, many of the tools available are open source tools.

Checkdesk, Storify, Citizen Desk, Google Media tools and other tools are designed or developed to help the Media Libraries or the media organizations in the era of Big Data to make newsgathering and reporting as easy as possible. There are so many other tools and technologies are available in the market which can also be explored by the media libraries or by the data journalist for their day to day data analysis and reporting.

References

Big data. (2016, March 05). Retrieved January 16, 2017, from https://en.wikipedia.org/wiki/Big_data

Beal, V. (n.d.). Big Data. Retrieved December 10, 2016, from http://www.webopedia.com/TERM/B/big_data.html

McCracken, Harry (16 August 2011). "The 50 Best Websites of 2011 - Storify". Time. Retrieved 27 May 2016.

Big data analytics: What it is and why it matters. (n.d.). Retrieved January 06, 2017, from https://www.sas.com/en_us/insights/analytics/big-data-analytics.html

Declues, J. (n.d.). Four Types of Big Data Analytics and Examples of Their Use. Retrieved January 10, 2017, from http://www.ingrammicroadvisor.com/data-center/four-types-of-big-data-analytics-and-examples-of-their-use

Rampton, J. (2015, May 12). 6 Technologies to Help Journalists in the Era of Big Data. Retrieved January 3, 2017, from https://www.forbes.com/sites/johnrampton/2015/05/12/6-technologies-to-help-journalists-in-the-era-of-big-data/#1dffca255b72

Thibodeaux, T. (2016, May 08). 10 tools that can help data journalists do better work, be more efficient. Retrieved December 06, 2016, from http://www.poynter.org/2011/10-tools-for-the-data-journalists-tool-belt/147736/

S. (n.d.). Sourcefabric. Retrieved December 06, 2016, from https://www.sourcefabric.org/en/home/software/

Sawers, P. (2013, October 16). Google Media Tools: A Hub For Journalists. Retrieved November 06, 2016, from https://thenextweb.com/google/2013/10/16/google-guides-journalists-towards-digital-resources-google-media-tools-hub/#.tnw_WKyRvccR#.tnw_A88gXsMx

Sieberg, D. (2013, October 21). Google Media Tools: a new intersection for newsgathering. Retrieved November 12, 2016, from https://www.blog.google/topics/journalism-news/google-media-tools-new-intersection-for/

About the book

The book Dynamics of Library for Excellence in Electronic Revolution is the collection of 56 selected contributions presented at International Conference on Library and Professional Submit (ILIPS 2017) held at IISER, Mohali, India. It gives the perspective through national and international contributions on how to re-envision the spaces in changing perspectives where physical and virtual are blending. The book also discusses some of the emerging and innovative technology applications in Libraries including the issues of collection development and management in electronic age. The user base studies forms the crux of the volume and will be of immense value to practicing librarians, students and teachers of library & information science, publishing industry, and IT professionals working in library and information science sector.

Dr. P Visakhi is presently working as Deputy Librarian (Incharge-Librarian), IISER Mohali and instrumental in Implementing "Learning Commons" in IISER Mohali library. handled two World Bank projects as having 23 years working experience in Library. Previously she was the Chief Librarian of ICAR H.Q, Krishi Bhavan as well as Librarian, IASRI, PUSA, New Delhi. She presented and published 50 articles in different Journals and National & International Conferences edited and compiled 10 books and two Journals in the field. She got trained by OCLC -Dublin for World Cat, Library of Congress- Washington DC. Also visited Libraries at Philippines, Hong Kong, Singapore, France etc. She has been awarded " *Young Information Scientist -2006"* by SIS and *"The Best Citizens of India-2011"*.

Dr. P. K. Jain is the Librarian at Institute of Economic Growth (IEG), University of Delhi Enclave, India. He has more than 30 years of experience in the field of Library and Information Science. He has published more than 30 research papers and presented at national and international conferences in India, USA, Canada, UK, Germany, Japan, France, South Korea, Singapore, Malaysia, Oman, Sri Lanka and Philippines. He is the editor of seven books. He has received several awards from Special Libraries Association, USA and various other nationally including-SLA Fellow 2015 and SIS Fellow Award from Society for Information Science. He has been very active in organizing and initiating various many seminars, national and international conferences under the banner of Special Libraries Association and Society for Library Professionals.

Dr. Debal C. Kar has more than 30 years of experience and holds Ph.D & MLIS Degree in LIS and Master Degree in Philosophy. He is presently working as University Librarian at Ambedkar University Delhi and previously was with TERI. He wrote about 25 articles and edited eight books. He is a recipient of the Endeavour Executive Fellowship 2014 from Government of Australia and worked as Visiting Fellow at University of Technology, Sydney, Australia. He is also recipient of scholarship by Belgium Govt. (ABOS). He was founder editor of World Digital Libraries and Editorial Board member of several journals.

Dr. Parveen Babbar is Deputy Librarian at Jawaharlal Nehru University. He has been the author of more than 50 papers published in various national and international journals and conferences. He is also editor of 4 books published as proceedings of International Conferences. He has received many awards and scholarships. He is associated with several professional associations in various capacities in India and abroad including SLA, USA; SLP, India; SIS, India.