CHALLENGES AND ISSUES DURING VISUALIZATION OF BIG DATA

Shilpa[1], Manjit Kaur[2]
[1]Student (M.tech), Department of Computer Science and Engineering
[2]Faculty, Department of Computer Science and Engineering
LPU, Jalandhar, India

Abstract: Big data is a collection of massive and complex data sets. There are many computational methods which are used to get the meaningful information from the large data sets. To discover the hidden patterns from the data sets from the massive data collection many technologies are used. Visual analysis is used for the purpose of visualization of data sets in graphics or charts etc. by the visualization of the data; it is very easy to make the decisions and attractive form of the data.

Keywords: Visualization, Visual Analytics, Social Media, Big Data, Hidden Patterns.

I. INTRODUCTION

In today era large amount of information in various forms from variety of sources is generated. To make it meaningful visual analysis is applied and various computational methods are used. To makes it understandable and meaningful big data analytics is performed for the processing of complex and massive data sets.

Big data analytics analyze the large amount of information used to uncover the hidden patterns and the other information which is useful and important information for the use.

II. DIMENSIONS OF BIG DATA

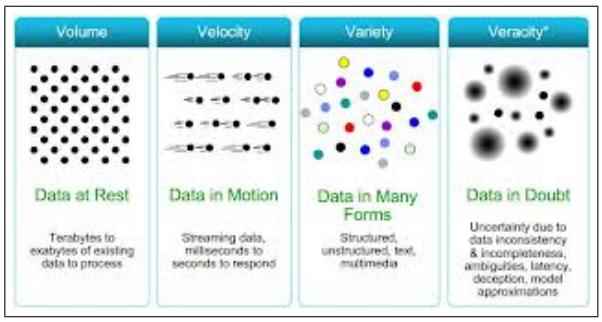

Figure 1: Dimensions of Big data

The Big data is defined by the 3V which are as follows :

Volume

Large amount of information generation in Zetabytes and petabytes. Per Day facebook is collecting 500 terabytes data and 400 millions tweets per day on twitter and 3 billion facebook likes and comments etc.

Velocity

How many data is generated. 150k per second to google, twitter and facebook. Velocity defines the motion of the data. During 2012, 2.5 quintillion bytes of data were created every day.

Variety

Data comes from the various sources in the form of structured and unstructured data such as text, images, videos, logs, social media and so on.

Veracity

The data generated from the various sources can be of various forms of combination of raw data, missing values, dirty data etc.

III. CHALLENGES AND OPPORTUNITIES

As there exist large amount of information, we face the various challenges and problems in the processing of massive and complex data sets.

The challenges include the unstructured data, real time analytics, fault tolerance, processing and storage of the data and many more. The main challenges and opportunities are as follows :

a) Storage and processing issues.

b) Data Accessing and sharing of information.

c) Complexity of data.

The main causes of complexities with big data are as follows :

- Human Perception
- Limited Screen space

Human perception means difficulty in extracting the useful information when the visualized objects become large. In limited screen space, the visibilities of objects are not proper. In short, the challenges with big data are as follows :

a) Acquiring and storing large amount of data.

b) Extracting useful information.

c) Aggregation and integration by representation.

d) Querying , data modeling and analysis.

e) Interpretation of the date to acquire the meaningful information.

IV. VISUALIZATION OF THE DATA

Visualization is the visual representation of the data to make it attractive and understandable easily.

Main objectives of visualization are as follows :

a) Understanding of data properly which is recorded.

b) Graphical representation of data.

c) To run the search query to locate the location of the text .

d) For discovering the hidden patterns.

e) Visibility of all the data items.

V. TOOLS FOR THE VISUALIZATION

ManyEyes for the visualization is a visualization tool launched by the IBM research and IBM Cognos software group. It is web Based tool used for the purpose of visualization and discovery. It is free available tool as a service on IBM's alpha Work Service website. By using this tool we can make the VA [B3] and easily make the decisions.

This tool can be used for structure as well as unstructured data for the visual analysis.

User can use the online available data as well as can upload their own data for the visual analysis. ManyEyes is a community-powered tool. There is more than 150,000 data sets availability online and pre-visualized.

In this tool Visual analytics of big data is done using ManyEyes Tools and data set is uploaded from the spreadsheets or can be text file. But there is a problem related to the proper visualization of data.

Many Eyes is a bet on the power of human visual intelligence to find patterns. Our goal is to "democratize" visualization and to enable a new social kind of data analysis.

With data generation of the large amount of data, there are many challenges and opportunities which are urgent to improve these challenges.

As I have study about the ManyEyes tool there is a problem of searching and proper visualization of the data.

VI. MANYEYES

With the help of ManyEyes the complex data is render that require the high level data input for the Visual analytics. IBM's ManyEyes is used to visualize the data when to make the decisions at a time we can make the interactive picture for the proper understanding.

A. Components of ManyEyes

1. **Word Tree**
   A word tree is a visualization tool for unstructured data such as a book data, article, speech or poem. By the use of this tool we can pick word or phrase and we can analyze the different context where these words appear.

   The given below visualization is a word tree of the NIST Guide IDPS, using search item "Wireless" the visualization is created.

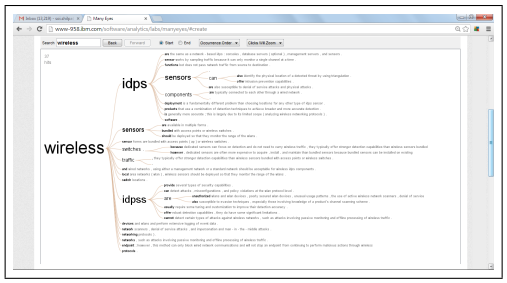

Figure 2: Word Tree Visualization

From this visualization we can easily see the text but difficult to analyze the occurrence of the text location in the given data. This is a searching issue in Word Tree of the Manyeyes Visualization tool.

We can see the next word also easily.

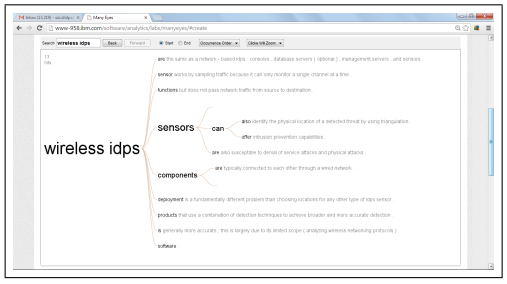

Figure 3: Wordtree

2. **TreeMap**
   A tree map is a visualization tool for the hierarchical structures. It shows the attributes of leaf nodes in effective ways. Tree map enable the user to discover the hidden patterns and exceptions.

   a) **Data Set**
   The data to visualize is taken from the spreadsheet after the uploading of the data in Many Eyes TreeMap it visualized and then display in TreeMap format.

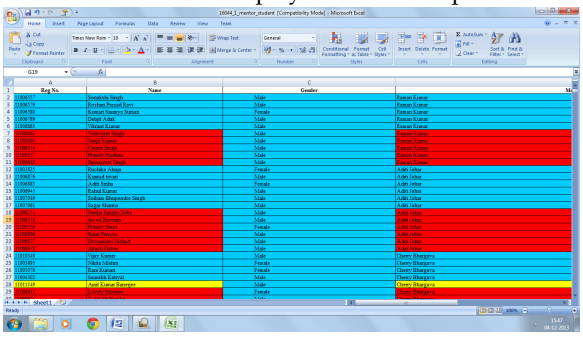

Figure 4: Spreadsheet Data 1

In the TreeMap, we can see the information about the person by just clicking on that column or we

can search.The disadvantage in this visualization is limited screen space.

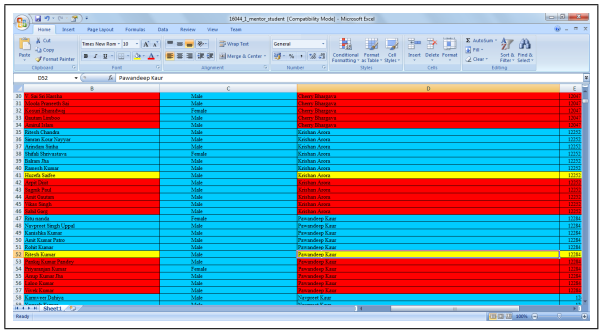

Figure 5: Spreadsheet Data 2

The whole data will upload after upload the visualization will display which is shown as below :

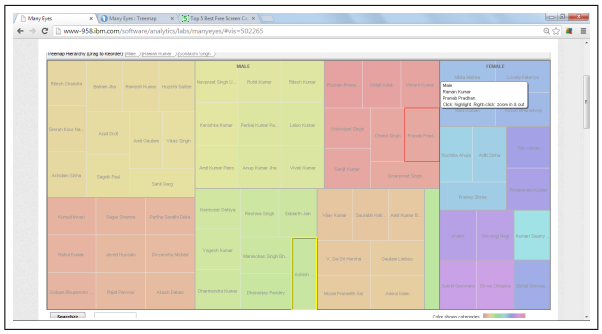

Figure 6: TreeMap Visualization

3. **Network Diagram**
   Network diagram display the relationship between the various entities used in the data sets. Network diagram only telling about the name not detail information about that person it need to be recover.

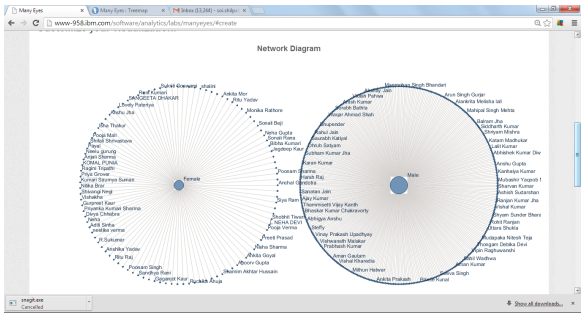

Figure 7: Network Diagra Visualization

VII. FUTURE SCOPE

As the visual analytics of the data helps in better understanding. There are various issues regarding visual representations that include the searching, discovering hidden patterns.

VIII. CONSLUSION

Visualization is a process of graphical representation of the large data sets. The various tools are used for the visual analytics. The graphical representation helps in understanding of the data and help in making decisions.

REFERENCES

[1] Mukherjee A., Datta J., Jorapur R., Singhvi R., Haloi S., Akram W. Shared disk big data analytics with apache hadoop. *High Performance Computing (HiPC) 19th International Conference*, 2012.

[2] Garlasu D., Sandulescu V., Halcu I., Neculoiu G. A big data implementation based on grid computing. *11th Roedunet International Conference (RoEduNet)*, 2013.

[3] Sagiroglu S., Sinanc D. Big data: A review. *Collaboration Technologies and Systems (CTS) International Conference*, 2013.

[4] Zhang Du. Inconsistencies in big data. *Cognitive Informatics andCognitive Computing (ICCI*CC) 12th IEEE International Conference*, 2013.

[5] `http://www-01.ibm.com/software/in/data/bigdata/`.

[6] `http://www.cloudcomputingpath.com/challenges-and-opportunities-with-big-data/`.

[7] Grosso P., de Laat C., Membrey P. Addressing big data issues in scientific data infrastructure. *Collaboration Technologies and Systems (CTS) International Conference*, 2013.

[8] Aditya B. Patel, Manashvi Birla, Ushma Nair. Addressing big data problem using hadoop and map reduce. *Engineering (NUiCONE) Nirma University International Conference*, 2012.

[9] Szczuka Marcin. Ifsa world congress and nafips annual meeting (ifsa/nafips). *Data and Knowledge Engineering 53*, 2013.

[10] Tien J.M. Big data: Unleashing information. *Service Systems and Service Management (ICSSSM) 10th International Conference*, 2013.

[11] `http://www.intel.in/content/dam/www/public/us/en/documents/white-papers/big-data-visualization-turning-big-data-into-big-insights.pdf`.

[12] `http://blogs.computerworld.com/business-intelligenceanalytics/23159/data-visualization-picture-worth-billion-bytes`.

[13] `http://smallbusiness.yahoo.com/advisor/applying-big-data-visualization-data-mining-054555947.html`.

