International Journal of Research Publication and Reviews

A Literature Survey on Data Visualization, Data Extraction, and Data Analysis

Nikhil Rathod[1], Saurabh Shinde[2], Prof. S. R. Kokane[3]

[1, 2, 3] *Dept. Information Technology Engineering, APCOER, Pune, India.*

A B S T R A C T

Data Visualization is a term that refers to a graphic representation of information. It's also the study of visual representations of abstract data to support mortal cognition where abstract data include both numerical and non-numerical data. This field is connected to infographics and scientific visualization, both of which use visual representations to convey information. Data Extraction is a process of retrieving data from complicated data, such data can also be treated as abstract data for visualization. Data Analysis allows us to make sense of visualized data, it helps us to take smart decisions.

Keywords: Data Visualization, Data Extraction, Data Analysis, retrieving data, abstract data

1. INTRODUCTION

Currently, Data is essential in almost every sector in companies, businesses, games, sports, schools, universities, etc. Sometimes it becomes complicated to extract, analyze, and sort the data, due to their vast entries. Data Visualization is a technique representing data in a visual form for a better understanding of underpinned data. The part of visualization is essential to easily convey what the data means and how it can significantly impact decision-making through similar representation. Its basic concept is Data Sorting which helps us to sort the data. In other words, sorting is a process of arranging data in a more meaningful manner (i.e. in sense of order) so that it becomes easier to visualize. Data Sorting has several types or ways, each having different algorithms, methods, and time & space complexity. Sorting Visualizer helps us to better understand sorting algorithms. Pathfinding is another basic concept in visualization, it refers to the process of finding a more efficient or effective path between two points (nodes) in graphs. It is a more practical variant of solving the maze. It has many algorithms like Dijkstra's algorithm, A* algorithm, etc. each having different time & space complexity. Pathfinding Visualizer helps us to understand pathfinding algorithms more effectively. Data Extraction is the process to collect or retrieve data from vast quantities, it also refers to retrieving disparate types of data from a variety of sources. It consists of making abstract data from poorly organized or completely unstructured data. Due to abstract data, it becomes easier to visualize data and also easier to understand. Data Analysis is a process of applying statistical and logical techniques to describe the data. It also refers to making a report or summary of visualized data so that it helps organizations or companies make smart decisions. It is also used in machine learning (ML) to train datasets. When Machine Learning (ML) predicts advanced results, it becomes important to visualize the output of the results to ensure that models are performing well.

2. LITERATURE SURVEY

Sr. No.	Paper Title	Authors & Year	Observation/Conclusion
1.	DATA VISUALIZATION	Matthew N. O. Sadiku, Adebowale E. Shadare, Sarhan M. Musa and Cajetan M. Akujuobi; (2016)	Data Visualization involves presenting data in graphical form which make information easy to understand. It explains the summary of any vast data. It will benefit any field of study that requires innovative ways of presenting large and complex information to understand the underlying data. Advanced computer graphics has reshaped data visualization.
2.	Super Sort Sorting Algorithm	Yash Gugale; (2018)	Sorting Algorithms is one of the important areas of research in computer science and engineering where a simple task of sorting the number is the backbone of complex algorithms. In recent times, many types of research are done to enhance time & space complexity of the algorithms.
3.	Data Visualization Techniques for real-time information -A Custom and	Renato Toasa, Marisa Maximiano, Catarina Reis, David Guevara; (2020)	Implement a generic and dynamic dashboard based on real-time information to assess the impact of the Data Visualization Techniques. The dashboard helps to interact with users based on initial set and existing set of Data Visualization Techniques.

	Dynamic Dashboard for Analyzing Surveys' Results		
4.	Visualization of Big Data Text Analytics in the Financial Industry	Živko Krstić, Sanja Seljan, Jovana Zoroja; (2019)	Textual data can decide new insights and bring precious business insights. Sources that are used for text analysis in financial industry contains internal document like emails and external documents like social media, websites, etc. Data Extraction uses to gather valuable information.
5.	Surface Optimal Path Planning Using an Extended Dijkstra Algorithm	MIN LUO, XIAORONG HOU, AND JING YANG; (2020)	Dijkstra algorithm is a classical well-known shortest path routing algorithm. It's a simple algorithm for the single-source shortest path problem, which can effectively calculate the shortest path to all destinations.
6.	Seaborn: statistical data visualization	Michael L. Waskom; (2021)	Seaborn is a high statistical graphical library in Python used for data visualization. It provides a high interface to the matplotlib library. It integrated with the pandas data structure. When plotting the dataset has to make, seaborn astronomically maps the data values to visualize the color, size, and style.
7.	Narrative Scientific Data Visualization in an Immersive Environment	Richen Liu, Hailong Wang, Chuyu Zhang, Xiaojian Chen, Lijun Wang, Genlin Ji, Bin Zhao, Zhiwei Mao, Dan Yang; (2021)	Narrative visualization for scientific data studies can help users better understand the domain knowledge because narrative visualizations frequently present a sequence of data and observations linked together by a unifying theme or argument. Narrative visualization in immersive surroundings can provide users with an intuitive experience to interactively explore scientific data because immersive surroundings give a brand new strategy for interactive scientific data visualization and exploration.
8.	What is Interaction for Data Visualization?	Dimara, Evanthia, Perin, Charles; (2019)	In this paper, they define that interaction is fundamental to Data Visualization. To tackle the problems, they synthesize an inclusive view of interaction in the visualization which includes insights from information visualization, visual analytics, and scientific visualization.
9.	DeepEye: Towards Automatic Data Visualization	Luo, Yuyu, Qin, Xuedi, Tang, Nan, Li, Guoliang; (2018)	Data Visualization is extremely useful to explain the significance of data to people who are visually oriented. The main task of automatic data visualization is, given a dataset to visualize its storytelling by transforming the data (e.g. selecting attributes, grouping and binning values) and deciding which visualization (e.g. bar, pie, or line charts) is the right type to a particular data. In this paper, they present DeepEye: A novel system for automatic data visualization that tackles three problems- Visualization Recognition Visualization Ranking Visualization Selection.
10.	The Case for a Learned Sorting Algorithm	Ani Kristo, Kapil Vaidya, Ugur Çetintemel, Sanchit Misra, Tim Kraska; (2020)	Sorting is one of the most basic algorithms in the field of Computer Science. Its operation is not just for sorting the data, but also a part of joins. In this paper, they introduced a new type of distribution sort that leverages the learning model of the eCDF (i.e. empirical Cumulative Distribution Function). Algorithms which used in a model to effectively get an approximate value of the scaled empirical CDF for each record key and map.

PROPOSED SYSTEM

We mainly focused on data preparation of finance related to banking and the stock market. Further, Data Extraction will do where users extract according to their wants. Later they use the abstract data for visualization or analytics.
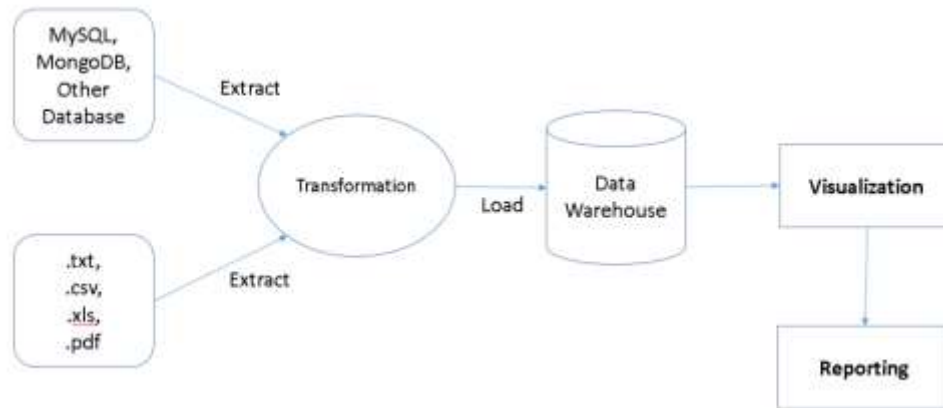
The above fig. gives an idea of Data Extraction to be done from different databases and different extensions. Collectively transform it and load it in data warehouse. Further, that data is treated as abstract data for visualization and analytics. It helps to make summaries or reports to boost decision-making. In visualization, we have the concepts of Sorting and Pathfinding. Sorting helps to sort the elements of data or abstract data. It has different algorithms and methods, each having different space & time complexities. Sorting visualization helps to understand different algorithms and techniques easily. Pathfinding visualization helps to find to shortest or most effective distance between two nodes. This concept helps us to develop tracking systems.

4. TECHNOLOGY

For Data Visualization, we use Power BI, python and its libraries. *Matpoltlib* for graphs and *Seaborn* for highly interactive graphs. For Data Extraction we use *pandas* and *numpy*. For Data Analysis we use Jupyter notebook. *Django/Flask* to make it easier to build a web page using python.

5. CONCLUSION

We conclude that solving data inefficiencies & absorb vast amounts of data presented in a visual format, increases the speed of decision-making. It helps to identify errors and inaccurate data quickly. It promotes storytelling and conveys the right information to the audience in an understandable form. The main goal of the project is to design and develop a strong application for building data visualization, extraction, and analysis tools to perform practical skills.

References

[1] Matthew N. O. Sadiku, Adebowale E. Shadare, Sarhan M. Musa and Cajetan M. Akujuobi (2016), "DATA VISUALIZATION", International Journal of Engineering Research And Advanced Technology(IJERAT).

[2] Yash Gugale (2018), "Super Sort Sorting Algorithm", International Conference for Convergence of Technology (I2CT).

[3] Renato Toasa, Marisa Maximiano, Catarina Reis, David Guevara (2020). Data Visualization Techniques for real-time information -A Custom and Dynamic Dashboard for Analyzing Surveys' Results.

[4] Živko Krstić, Sanja Seljan, Jovana Zoroja (2019), "Visualization of Big Data Text Analytics in Financial Industry", ENTRENOVA.

[5] MIN LUO, XIAORONG HOU, AND JING YANG (2020), "Surface Optimal Path Planning Using an Extended Dijkstra Algorithm", IEEE Access.

[6] Michael L. Waskom (2021), "Seaborn: statistical data visualization", Journal of Open Source Software.

[7] Richen Liu, Hailong Wang, Chuyu Zhang, Xiaojian Chen, Lijun Wang, Genlin Ji, Bin Zhao, Zhiwei Mao, Dan Yang (2021), "Narrative Scientific Data Visualization in an Immersive Environment", Bioinformatics.

[8] Dimara, Evanthia, Perin, Charles (2019), "What is Interaction for Data Visualization?" IEEE Transactions on Visualization and Computer Graphics.

[9] Luo, Yuyu, Qin, Xuedi, Tang, Nan, Li, Guoliang (2018), "DeepEye: Towards Automatic Data Visualization", International Conference on Data Engineering (ICDE).

[10] Ani Kristo, Kapil Vaidya, Ugur Çetintemel, Sanchit Misra, Tim Kraska (2020), "The Case for a Learned Sorting Algorithm", International Conference on Management of Data.

[11] Charis Ntakolia, Dimitris K. Iakovidis (2021), "A swarm intelligence graph-based pathfinding algorithm (SIGPA) for multi-objective route planning", Computers & Operation Research.