Journal of computer science and software applications
https://www.mfacademia.org/
ISSN:2377-0430
Vol. 3, No. 3, 2023

# Research on Data Visualization Design Based on Big Data

**Samuel Scott[1],Lily Adams[2],Jackson Baker[3]**

Texas Christian University[1],Texas Christian University[2],Texas Christian University[3]

Samuelplm@gmail.com[1],lilylilyaa@gmail.com[2],jacksobaker1@gmail.com[3]

## Abstract:

This study addresses the technical domain of big data applications, focusing on data visualization design research based on big data. The methodology includes collecting or inputting raw data, preprocessing and storing the data to obtain precise initial data, visually matching the preprocessed data, completing data modeling, creating and designing visual charts, and presenting the data visually. The aim of this research is to enhance data visualization design by optimizing data processing methods and utilizing the automatic analysis and production capabilities of computers. This approach enables a more intuitive and efficient understanding of the information, knowledge, and insights derived from big data.

## Keywords:

Big Data; Visualization; Network.

## 1. The Former Status Quo

At present, the whole world has stepped into the era of big data. With the rapid development of information technology such as Internet, cloud computing and Internet of Things, information technology is deeply integrated with all aspects of the human world, producing more massive data than ever before. Make statistics and analysis on massive data, and dig out the potential value contained in it. We have been studying it in depth. Once upon a time, data analysts, expertsin statistics and their researchers counted and analyzed the data. However, in the current background of big data, massive data can only show their profundity perfectly after being reasonably collected, interpreted and expressed.Data visualization makes the data more intimate and understandable, and the information conveyed by a picture is worth a thousand words. More than 80% of human information from the outside world comes from human visual system. It is easier and more convenient for analysts to dig out the hidden information behind the data by presenting big data to analysts in an intuitive and visual graphical way.

Big data visualization analysis is an indispensable means and tool for big data analysis. In fact, visualization has always been an important method and means in the field of scientific computing visualization and traditional business intelligence (BI). However, these research fields have not deeply combined the theory and technology of human-computer interaction.Therefore, it is difficult to fully support the human-computer interaction process of visual analysis. At the same time, the new features of big data itself also put forward more urgent needs and severe challenges for visual analysis.

## 2. Design Content

To solve the technical problems.

In view of the shortcomings of the existing technology, this design provides the research of data visualization design based on big data. By optimizing the deep processing method of data, and with the help of the automatic analysis ability and production ability of computer, it helps

people to understand the information, knowledge and wisdom behind big data more intuitively and efficiently.

## 3. Technical Scheme

In order to achieve the above purpose, this design is realized by the following technical scheme: Research on data visualization design based on big data includes the following steps:

S1, collecting or inputting original data, preprocessing and storing the original data, and obtaining accurate initial data after preprocessing;

S2, visually matching the obtained initial data, which mainly includes data filtering, smoothing, normalization, geometric transformation, linear transformation, feature detection and extraction, etc.

S3, mapping: establishing data set for the data processed in S2, converting numerical data into geometric data, and completing data modeling;

S4, drawing and designing charts: according to the needs of business data presentation mode, selecting chart types, matching the display values of the charts to be presented, and drawing the charts by using the drawing engine of the visual class library built in the computer;

S5: Visual presentation: by preparing the page layout of the page to be displayed, customizing local charts, configuring data sources and data sets, and integrating the unified interface for obtaining data from the big data platform, the data sources to be presented are displayed on the front end of the Web page, thus realizing the configuration and presentation of the automatic visual analysis page of the big data platform.

The preprocessing in step S1 mainly includes standardization, analysis, segmentation, classification, compression and decompression of data format.

The raw data collected in step S1 matches the fusion architecture of MPP database and Hadoop, and provides multi-flat SQL driver.

Furthermore, by executing SQL instruction query fields, integrating the fields into data sets and adopting INTEL SSE4.1 and SSE4.2 instruction sets, the instruction execution can be accelerated. The chart types in step S4 include line chart, histogram, scatter chart, bubble chart, data table, flow chart, KPI label, funnel chart, Sankey chart, radar chart and regional map, etc.

The visualization class libraries in step S4 include statistical data visualization class library, relational data visualization class library, geospatial data visualization class library and text data visualization class library, etc.

## 4. Innovation

This design provides data visualization design research based on big data. Compared with the existing known technologies, this design has the following beneficial effects:

After mining and analyzing the massive data in the data platform, this design independently selects different forms of visual class libraries according to different data dimensions and business types, and converts the big data into simple and easy-to-understand gorgeous data visualization graphics or charts, which are displayed in front of users intuitively through the human-computer interaction page.Greatly improve the adaptability of data use, and enrich the expression and aesthetic feeling of data; So that non-professional data analysis experts can easily get the profound meaning of the data stored in the massive data platform.

## 5. Implementation Mode

Example:

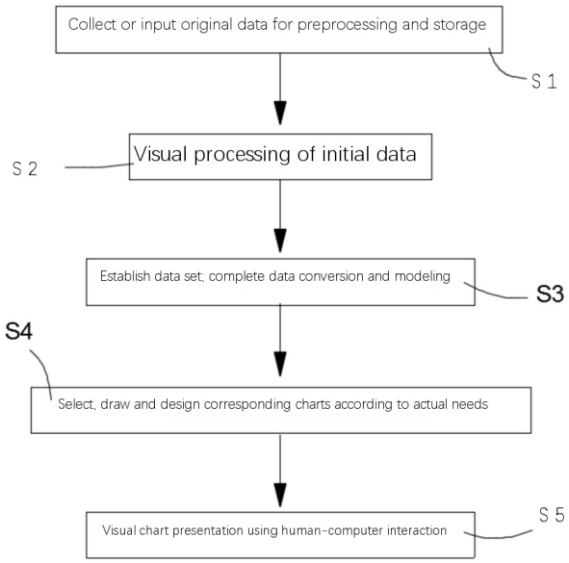


**Figure 1.** A flow chart of the visual presentation method of this design.

The research of data visualization design based on big data in this embodiment, referring to Figure 1, includes the following operation steps:

S1, collecting or inputting original data, preprocessing and storing the original data, and obtaining accurate initial data after preprocessing;

S2, visually matching the obtained initial data, which mainly includes data filtering, smoothing, normalization, geometric transformation, linear transformation, feature detection and extraction, etc.

S3, mapping: establishing data set for the data processed in S2, converting numerical data into geometric data, and completing data modeling;

(Different mapping techniques should be adopted for different data types, and commonly used methods are as follows: verifying the cardinality of dimension by applying recognizable variables in graphics; Use scaling and offset to match the numerical range; Use projection to compress information, obtain metrics, etc.)

S4, drawing and designing charts: according to the needs of business data presentation mode, selecting chart types, matching the display values of the charts to be presented, and drawing the charts by using the drawing engine of the visual class library built in the computer;

S5: Visual presentation: by preparing the page layout of the page to be displayed, customizing local charts, configuring data sources and data sets, and integrating the unified interface for obtaining data from the big data platform, the data sources to be presented are displayed on the front end of the Web page, thus realizing the configuration and presentation of the automatic visual analysis page of the big data platform.

The preprocessing in step S1 mainly includes standardization, analysis, segmentation, classification, compression and decompression of data format.

The preferred implementation of this design: in step S1, the original data collected matches thefusion architecture of MPP database and Hadoop, and provides multi-flat SQL driver. Execute SQL instructions to query fields, integrate fields into data sets, and adopt INTEL SSE4.1 and SSE4.2 instruction sets to speed up the execution of instructions.

The preferred embodiment of this design: The chart types in step S4 include line chart, histogram, scatter chart, bubble chart, data sheet, flow chart, KPI label, funnel chart, Sankey chart, radar chart and regional map, etc.The visualization class libraries in step S4 include statistical data visualization class library, relational data visualization class library, geospatial data visualization class library and text data visualization class library, etc.

## 6. Conclusion

After mining and analyzing the massive data in the data platform, this design independently selects different forms of visual class libraries according to different data dimensions and business types, and converts the big data into simple and easy-to-understand gorgeous data visualization graphics or charts, which are displayed in front of users intuitively through the human-computer interaction page.It greatly improves the adaptability of data usage and enriches the expression and aesthetic feeling of data.

## References

[1] Research on interactive experience system of large screen data using data analysis and visualization technology [J]. Xu Tong. Henan Science and Technology. 2019(07).

[2] Research on data analysis and visualization under the background of big data [J]. Long Hu, Yang Hui. Journal of Kaili University .2016 (03).

[3] The application of visual charts of Power BI, a sharp tool for data analysis [J]. Jia Yunbin. Computer programming skills and maintenance. 2019(03).

[4] Data Science and Big Data Analysis: Visualization and Representation of Data Discovery and Analysis [J]. Automation Expo.2018 (08).

[5] Key points of visualization in the field of big data analysis [J]. Wang Bo, Shen Fangfang. China New Communication. 2017(14).

[6] Computer software collocation of bioinformatics data analysis [J]. Liu Zhen. Science and Technology Wind. 2018(18).

[7] Process design optimization using embedded hybrid visualization and data analysis in genetic algorithm optimization framework [J]. Wang Kefeng, Zhou Yue, Qian Xinhua, Liu Jiang, Han Zhizhong. computers and applied chemistry, 2006(10).

[8] Digital technology makes grass-roots epidemic prevention and control management more efficient [J]. Long Zhizhong, Feng Qiang. China Media Technology.2020 (02).

[9] Big Data and Security Visualization [J]. Hong Xiang, yu zhang, Hu Haibo. Journal of Chongqing University .2016 (02).

[10] Data analysis based on big data [J]. Zhou Mingjun, Liu Hong. Science and Technology Communication .2019 (08).