ISSN:2790-1688 Volume-15-(2025)

A Review of the Classification and Application of Big Data Visualization Techniques

Jianhao Yu

Cambridge International School LITAI College, SISU, Shanghai,200083, China

Abstract. This article systematically categorizes the fundamental theories and technological developments in data visualization, providing a detailed introduction to the basic visualization methods for various types of data, including spatiotemporal data, geographic information data, time series data, associated data, and text data. Moreover, in light of the characteristics of these different data types, the article will also propose customized visualization solutions. For instance, using step charts and fitting curves can clearly outline the patterns of data changes over time. By exploring the interrelationships among multiple variables through scatter plot matrices and bubble charts, we can uncover the hidden patterns behind these variables more deeply. In addition to the visualization technology itself, this article also incorporates relevant content on data management tools, including distributed storage systems and columnar database management systems. The addition of these tools can further enhance the scalability and efficiency of data processing, providing a more solid underlying support for visualization work. In addition, based on the principles of visual coding and human cognitive laws, this paper also proposes corresponding design criteria to ensure that the introduced visualization methods conform to human perception habits, thereby making the final visualization results more interpretable and more practical.

Keywords: Data Science; Big Data Visualization; Visualization Methods; Combination of Pictures and Texts; Classification.

1. Introduction

As a derivative field of computer graphics, data visualization has undergone a significant technological leap since its development in the mid-twentieth century [1,2]. At the technical level, the field has changed from the original two-dimensional geometry. For example, linear regression plots and histogram distribution maps have been introduced to the dynamic mapping of multidimensional spatiotemporal data.

The primary purpose of this paper is two-fold. First, this paper establishes a comprehensive technical framework for data visualization, encompassing key stages such as data collection, cleaning, mapping, and interaction. Then, the visualization methods of typical scenarios, such as spatiotemporal, relational, and text data, are classified and analyzed, and their design principles are summarized. The significance of this paper lies in the following aspects. First, the visual design criteria are proposed based on visual coding and cognitive rules. The second point proposes introducing the topic river model for temporal text analysis. The final point is to systematically evaluate the suitability of tools such as columnar databases and graph databases in data management.

In detail, for multiple data types, including spatiotemporal data, time series, relational data, and text information, a corresponding visualization method is proposed. A step graph and a fitting curve illustrate the continuous-time trend. The multivariate correlation is explored using a scatter plot matrix and a bubble map, and a new model is introduced to present the topic intensity evolution of the temporal text dynamically. This paper also utilizes data management tools, including distributed storage and columnar databases. It proposes design criteria based on visual coding principles and cognitive rules to provide methodological support for cross-domain applications.

The rest of this paper is organized as follows. Section 2 introduces the basic concepts, development history, and core data visualization processes, with a focus on the technical details of data preprocessing and visual mapping. Section 3 presents visual design principles and tools, and compares and analyzes data management solutions, including distributed storage and document databases. Subsequently, Section 4 focuses on the visualization of temporal data, proposing

application strategies for ladder diagrams and fitting curves in both continuous and discrete scenarios. Section 5 presents the correlation expression of relational data. The potential association between variables is revealed by combining a scatter plot matrix, bubble chart, and other tools. Then, in Section 6, a dynamic analysis method based on keyword extraction and topic river modeling was proposed for the in-depth analysis of text data visualization. Finally, Section 7 concludes the whole paper.

2. Data Visualization Basics

2.1 The Definition and Evolution of Data Visualization

2.1.1 Definition

As a cutting-edge technology system, data visualization primarily utilizes graphic symbols and dynamic media to skillfully transform complex multidimensional information into intuitive and easily understandable visual forms. Data visualization is also a scientific and technological research field focused on the visual representation of data. Among them, the visual expression of this data is defined as a type of information presented in a concise and summarized form, encompassing various attributes and variables of the corresponding information unit.

2.1.2 Development

In the early 1950s of the 20th century, computer graphics processing technology ushered in a breakthrough[4], enabling the transformation of data into images for the first time. This pioneering achievement laid the groundwork for the basic methodology of visualization technology. With the popularization of basic tools and the expansion of technology in the 1970s, graphics display devices and desktop operating systems gained increasing popularity, and interactive visual programming technology broke free from the confines of the laboratory, gradually extending to a broader range of practical scenarios. The gradual improvement and expansion of the technical system (from the 80s to the 20th century) has led to the emergence of multidimensional data innovation mapping, dynamic interaction technology innovation, and deep integration of analysis tools. The boom in modern multi-application (2020s to present) has partnered with neurocognitive science to develop visual saliency models, significantly improving the cognitive absorption efficiency of information[6]. At the same time, generative AI technology is used to automate the generation of charts, elevating visualization technology to a new level of intelligence.

2.2 The process of visualization

2.2.1 Data Collection

Data collection is the first step of data analysis and visualization. The method and quality of data acquisition significantly determine the final effect of data visualization. There are various classification methods for data collection, which can be categorized into internal data collection and external data collection based on the perspective of data sources.

Firstly, internal data collection refers to the process of gathering data on an enterprise's internal business activities. Typically, this data originates from an enterprise database, such as an order transaction system.

Secondly, external data collection generally refers to the amount of data obtained outside the enterprise through some method. However, the data obtained by the above two collection methods is not first-hand. In scientific research, researchers often need first-hand data. Therefore, we often need to collect data through experiments. This method is more effective.

2.2.2 Data Processing and Transformation

Data processing[1] and data transformation are the prerequisites for data visualization, including data preprocessing and data mining. The data obtained through early data acquisition inevitably contains noise and error, resulting in low data quality. In addition, the characteristics and patterns of

data are often hidden in massive data, which requires us to extract key information through data mining technology.

It is precisely because of the existence of the above problems that the conclusions drawn by directly analyzing or visualizing the collected data often mislead users into making wrong decisions. Therefore, data cleaning and normalizing the collected raw data are essential for effective data visualization. In the era of big data, the data we collect usually has 4V characteristics: Volume, Variety, Velocity, and Value. Mining valuable information from high-dimensional, massive, and diverse data to support decision-making requires data cleaning, noise removal, and secondary data processing for business purposes.

2.2.3 Visual Mapping

It is usually explained that the next step is visual mapping after the data has been cleaned, denoised, and processed according to the business purpose. Visual mapping[2] is the core of the entire data visualization process, which refers to mapping processed data information into visual elements. The visualization element consists of 3 parts: the visualization space, the marker, and the visual channel. The display space of data visualization is usually two-dimensional. Visualizing three-dimensional objects solves the problem of displaying them in a two-dimensional plane through graphic drawing technology, such as 3D donut charts and 3D maps.

Tags are mappings of data attributes to visual geometry elements to represent the categorization of data attributes. According to the difference in spatial degrees of freedom, markers can be categorized into points, lines, surfaces, and volumes, corresponding to zero degrees of freedom, one-dimensional, two-dimensional, and three-dimensional degrees of freedom, respectively. For example, our standard scatter charts, line charts, rectangular tree charts, and three-dimensional column charts use four markers: points, lines, surfaces, and volumes. However, a visual channel refers to mapping the value of a data attribute to the visual presentation parameters of a marker. It is often used to present quantitative information about a data attribute.

2.2.4 Human-Computer Interaction

Standard interaction methods include Scroll and Zoom, Color Mapping Control, Control of data mapping methods, and control of data details hierarchy. Scrolling and Zooming, when data cannot be fully displayed on devices with current resolutions, are very effective ways of interaction, such as viewing information details on maps and line charts. However, the specific effects of scrolling and scaling are related to the page layout and the particular display device. Color Mapping Control refers to visual, open-source tools that provide color palettes, such as D3. Users can configure the colors of visual graphics according to their preferences. There are relatively more platform tools, such as self-service analysis. Still, for some self-developed visualization products, professional designers are generally responsible for this work, so the visual communication of these products has aesthetic value. Control of the data mapping method refers to the user's selection of data visual mapping elements. Generally, a data set has multiple sets of features. It provides users with flexible data mapping methods, allowing them to explore the information behind the data according to their interests. It is available in commonly used visual analysis tools, such as Tableau and Power BI.

3. Visualization methods and principles

3.1 Visual Design Tools and Principles

3.1.1 Organization and Management Tools

First, we will introduce the first visual data organization method, the distributed file system[3]. Generally, it refers to the fact that a file may be physically dispersed and stored on nodes in different locations. Each node communicates and transmits data through a computer network, but is still logically a complete file. We don't need to know which specific node the data is stored on when we use a distributed file system.

For document storage[4], the document storage model supports a nested structure. The MongoDB database achieves similar functionality by allowing you to specify the path of a JSON field in a query.

There are also methods, such as columnar storage[5]. Columnar databases are fast because they require reading fewer data blocks during a query. Because the same type of columns is stored together, the data compression ratio is high, simplifying the complexity of data modeling. However, it is stored in columns, and insert and update operations are slower, making it less suitable for databases with frequent data changes. It is ideal for use in decision support systems, data marts, and data warehouses, but not for online transaction processing.

The last type of storage described is key-value storage. Key-value storage, also known as KV storage, is called KV storage. Its data is organized, indexed, and stored in key-value pairs. Key-value storage, as a NoSQL database, can effectively reduce the number of reads and writes to disk, offering better read and write performance than SQL database storage.

Databases are also visual data management tools, such as graph databases[6]. It is used when there are complex network relationships between entities (these relationships can be represented as graph data). A typical example is the relationship between people in social networks, where the use of relational databases to store this "relational" data is not very effective. Its queries are complex, slow, and often exceed expectations, and the emergence of graph databases has addressed this shortcoming.

Then, there's what is called a relational database. The relational model is the most traditional data storage model, which stores data in rows within schematically defined tables. Each column in the table has a name and a type, and all records in the table must conform to the table's definition.

Finally, an in-memory database[7] is a database that puts data in memory and operates directly. In-memory data can be read and written orders of magnitude faster than disk data. The most prominent feature of MMDB is that its data is in-memory; that is, the active transaction only "deals" with the in-memory data of the real-time in-memory database, and the processed data is usually "short-lived" for a specific period of time. New data is generated when it is outdated. Therefore, in practical applications, in-memory databases enable the processing of real-time business logic.

3.1.2 Visual Design Principles

There are generally three principles of visual design that need to be adhered to: the principles of data filtering, intuitive mapping of data to visualization, and design of view selection and interaction. The principle of intuitive mapping of data to visualization requires designers to anticipate user behavior and expectations when using visualization results, thereby enhancing the usability and functionality of visualization design and facilitating user understanding of the results. Designers can reduce the time it takes for users to perceive and process information by leveraging existing prior knowledge. View selection and interaction design principles enable the use of simple data in basic visual views.

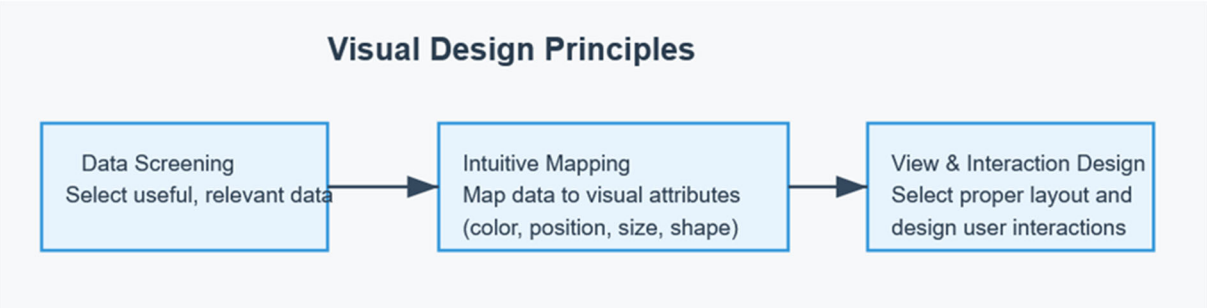

Fig. 1 Visual Design Principles

3.2 Visualization of major types of data

3.2.1 Visualization of space field data

Spatial field data[8] are named according to the dimensions of the space and the characteristics of the attribute values.

The cell structure is closely related to the method of sampling in space and the cell division strategy. When sampling is referred to as raster, the common sampling strategies include uniform raster sampling at the same interval, non-uniformly distributed linear raster sampling, and sampling based on geographic coordinates, among others. Although the spatial field data is obtained through sampling, its value does not correspond to a specific point. Still, it is a measure of a particular range in space, and all samples are continuously and adjacent to each other distributed in the entire spatial domain.

3.2.2 Visualization of geographic data

Map projection is the basis of geographic data visualization[9]; its purpose is to map the spherical surface to a specific surface and establish a correspondence between each point on the spherical surface and a point on the plane, thereby realizing the parameterization of the spherical surface. Simply put, it is a method of projecting geographic information data onto the Earth's surface.

4. Time data visualization

4.1 Application of time data in big data

For data, time is a critical dimension and property. The accumulation of historical data is an essential reason for the "big" of big data. At the same time, time series data are found in various fields, including financial and commercial transaction records, socio-economic indicator records, meteorological observation data, and data on animal and plant populations, among others.

Time data is divided into two types: discrete and continuous. Regardless of the data visualization method used, the primary purpose is to identify the trend of data changing over time. The patterns in these changes transcend a particular moment and contain rich information that can only be discovered through observation and analysis over time.

Moreover, Time-varying data refers to data that changes over time and possesses time attributes. The first type of time series data is generally arranged on a time axis. For example, stock charts, the Olympic schedule, and other similar resources. The second is not a time variable but has an intrinsic arrangement order. For example, text, DNA sequencing, etc. Characteristics of time-varying data: large quantities, dimensions, and many variables in practical applications, as well as rich types and wide distribution ranges.

4.2 Continuous-time data visualization

Continuous-time data[10] refers to an infinite number of numerical values that can be subdivided between any two time points, representing a record of continuous and changing phenomena. For example, temperature is the constant time data we are most often exposed to, and temperature can be measured at any moment of the day. The real-time price of stocks can also be regarded as continuous-time data.

4.2.1 Ladder diagram

The ladder diagram[11] is a type of X-Y diagram, typically used for discrete changes in Y values, where a sudden change occurs at a specific X-value position. At the same time, the step diagram can also express the change of numerical values over time in irregular and intermittent steps. For example, bank interest rates can be represented by ladder charts, as they typically remain unchanged for an extended period.

4.2.2 Line Chart

A line chart[12] is a graph composed of a straight line segment connecting each data point, which displays the changing trend of data in a line format. In a line chart, time is distributed evenly along the horizontal axis, and numerical values are distributed evenly along the vertical axis.

The Line chart is more suitable for displaying performance trends and often shows time data, such as population growth trends, book sales volume, and fan growth progress. However, it should be noted that the length of the horizontal axis will affect the curve trend displayed. If the horizontal axis

in the figure is too long and the separation distance between points is relatively large, it will exaggerate the entire curve; if the horizontal axis is too short, the user may not be able to see the change trend of the data. Therefore, it is essential to set the length of the horizontal axis reasonably.

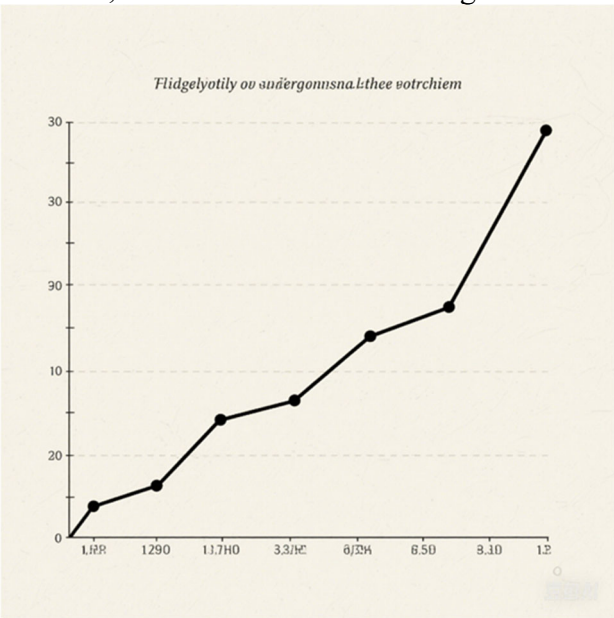

Fig. 2 Example line chart

4.2.3 Fitting Curve

A fitted curve[13] is drawn based on a given discrete data point, also known as an irregular curve. In real life and work, variables often do not follow a linear relationship. Fitting a curve refers to selecting the appropriate curve type to fit the observation data and analyzing the relationship between two variables using the fitted curve equation.

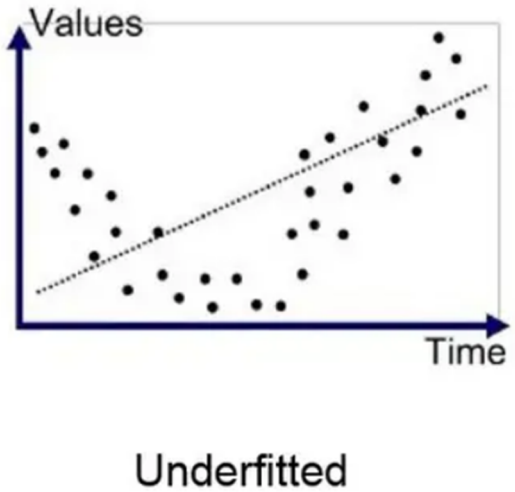

Fig. 3 Example of Fitting Curve

The fitting curve method involves establishing a data relationship (mathematical model) from a given set of discrete data points, finding a series of small straight-line segments, and connecting these interpolated points into curves. As long as the intervals of interpolated points are appropriately selected, a smooth curve can be formed. If a large amount of data is obtained or the data is very messy, it may be difficult or even impossible to identify the development trends and patterns. Therefore, fitting estimation can be used to simulate the trend.

4.3 Discrete-time data visualization

Discrete-time data is also referred to as discontinuous time data, and the number of such data points between any two time points is limited. In discrete time data, the data comes from a specific time

point or time period, and the possible values are also restricted. For example, the total number of medals in each Olympic Games or the number of gold medals in each country is discrete data.

4.3.1 Scatter plot

A scatter plot refers to the distribution plot of data points on the plane of the Cartesian coordinate system in mathematical statistical regression analysis. A scatter plot represents the relationship between the dependent variable and one or more independent variables. Therefore, the trend can be selected to fit the empirical distribution, and then the functional relationship between the variables can be found. For discrete-time data, the horizontal axis represents time, and the vertical axis represents the corresponding numerical value.

4.3.2 Column Chart

A column chart, also known as a bar chart and histogram, is a graph that displays the numerical values of statistical indicators based on the difference in height or length. The column chart is concise and eye-catching, and is a commonly used statistical graph. Column charts are typically used to display changes in data over time or to compare items across different periods. In addition, the numerical value's reflection is the cylinder's height. The shorter the column shape, the smaller the value; the higher the column shape, the larger the value.

It should also be noted that the spacing between the degrees of the column and the adjacent columns determines the aesthetics of the visual effect of the entire column chart. If the width of the column is smaller than the spacing, the reader will focus on the blank space and ignore the data; therefore, it is essential to choose a width that is reasonable.

4.3.3 Stacked column chart

Stacked column charts are variants of ordinary column charts. Stacking column charts will superimpose one or more other columns on top of a single column, typically with different colors. If the data exists in subclasses and the sum of these subclasses is meaningful, then stacked column charts can be used.

5. Relational Data Visualization

5.1 Application of relational data in big data

A significant value of big data is that it can help us identify connections between variables and uncover the cause-and-effect relationships behind various phenomena. An essential step before big data mining is to explore the correlation between variables and then explore the causal relationships that may be hidden behind them. When analyzing data, we can not only observe from the whole but also pay attention to the distribution of the data. The correlation between each distributed data point can also be observed from a broader perspective. In fact, the most crucial point is the meaning of the charts presented to readers after the data is visualized. Relational data are related and distributed. The following provides a detailed explanation of relational data, including examples and guidance on observing correlations between data.

5.2 Data relevance

The correlation between two things is relatively easy to discover, but it does not necessarily imply a causal relationship. For example, when the price of soybeans rises, the cost of pork may also increase, but the price of soybeans may not be the reason for the rise in pork. Despite this, correlation can still bring me great value. For example, soybean prices have risen, so we can seize the opportunity to stock up on some pork, which can often save a significant amount of money. Whether there is a causal relationship behind it is not that important. Big data visualization is used to convey the analysis results, not the reasoning behind them.

The core of data correlation[14] refers to the mathematical relationship between two sets of quantized data. A strong correlation indicates that when one value changes, the other value changes

in a corresponding manner. On the contrary, a weak correlation means that when one value changes, the other value changes by almost nothing. Through data correlation, the change of another numerical value can be predicted based on the change of one known numerical value. Below, we will examine this relationship using scatter plots, scatter plot matrices, bubble plots, and other visualizations.

5.2.1 Scatter chart

There are generally three relationships between variables: positive correlation, negative correlation, and uncorrelation, as shown in the figure. When a positive correlation exists, the change trends of the horizontal axis data and vertical axis data are the same; when a negative correlation exists, the change trends of the horizontal axis data and vertical axis data are opposite; when uncorrelated, the arrangement of scatter points is disorganized.

There are more scientific methods in statistics (such as correlation coefficients[15]) to measure the correlation between two variables. Still, scatter plots are often the most straightforward and most intuitive way to judge correlations. Before calculating correlation coefficients, scatter plots are usually used to make preliminary judgments.

5.2.2 Scatter plot matrix

The previous scatter plot uses two sets of data to form multiple coordinate points, and then observes the distribution of these points to determine whether there is a specific correlation between the two variables or summarizes the distribution pattern of the coordinate points. But in many cases, there are more than two variables, so the relationship between multiple (more than two) variables should be examined at the same time. Still, it is very cumbersome to draw a simple scatter plot between them one by one. At this time, you can use the scatter plot matrix[16] to draw scatter plots of multiple variables at the same time, so that you can quickly discover which variables have higher correlations. This method is beneficial in the data exploration stage.

5.2.3 Bubble diagram

Compared to the bubble chart and the scatter chart, there is an additional dimension of data. A bubble chart[17] turns a "point" with no size in the scatter plot into a "circle" with size, and the size of the circle can be used to represent the size of the extra one-dimensional data. The bubble chart enables us to compare three variables simultaneously. At the same time, the smaller the two indicators, the larger the bubble, indicating a higher price; conversely, the larger the two indicators, the smaller the bubble, indicating a lower price.

5.2.4 Histogram

A histogram[18], also known as a frequency distribution map, is an accurate graphical representation of the distribution of numerical data. The height of the column in the histogram represents the numerical frequency, and the width of the column is the value range. The horizontal and vertical axes differ from those of the general column chart, as they are continuous, whereas the horizontal axes of the general column chart are separated.

5.2.5 Density diagram

In 5.2.4, it was mentioned that the histogram reflects the distribution of a set of data. The horizontal axis of the histogram is continuous. The entire chart shows a column shape, and the user cannot know the internal changes of each column. For the stem and leaf chart, users can see specific numbers, but the requirement to compare the size of the gap between the values is not very clear. To present more details, a density map was proposed, which can be used to visualize the distribution's details.

When the histogram is enlarged in segments, the group distance between segments will be shortened. At this point, the polyline drawn according to the histogram will gradually become a smooth curve, which is referred to as the overall density distribution curve. This curve can reflect the density of the data distribution.

6. Text Data Visualization

6.1 Application and extraction of data

Human society continues to accumulate textual information [19], and in the computer era, a large amount of data can be stored on a tiny hard disk. On the Internet, there is a massive amount of "user-generated content" every day. People receive information at a rate that is slower than the rate at which information is generated, especially text information. Massive amounts of information make it increasingly difficult for people to process and understand. However, the information extracted by traditional text analysis technology still cannot satisfy people's reasonable analysis, understanding, and application of browsing and filtering methods. At this time, the critical role of text visualization is reflected. Text visualization presents text content in the form of visual symbols, allowing people to understand the information quickly. From humanities research to government decision-making, from precision medicine to quantitative finance, from customer management to marketing, these massive texts play an important role everywhere as one of the most essential information carriers. Related occupations, such as intelligence analysts, network content analysts, sentiment analysis specialists, or literary researchers, often require text visualization.

Text visualization relies on natural language processing, so techniques such as bag-of-words modeling, named entity recognition, keyword extraction, theme analysis, and sentiment analysis are more commonly used in text analysis. The text visualization process primarily involves text data preprocessing and filtering out invalid information. Feature extraction involves extracting text vocabulary and content, as well as analyzing the similarity between texts and performing text clustering, among other tasks.

6.2 Web crawler to extract text data

The web crawler[20] refers to a type of program that can automatically access the network and crawl specific information, and is sometimes called a "web robot" or "network robot". People can obtain and use the information they are interested in more efficiently and effectively, thereby completing a significant amount of valuable work conveniently. The most popular one currently is writing crawlers in Python[21]. There are many third-party libraries available for use, including Requests, urllib, and Scrapy, among others. Among them, the Scrapy library provides a relatively comprehensive crawler framework, as shown in the figure, which can save a significant amount of time and effort.

6.3 Keyword visualization

If a word appears frequently in a text, then the word may be the keyword of the text. The TF-IDF method is often used to calculate the importance of words in conveying textual information [22]. Where TF refers to the frequency of occurrence of words in the target text, IDF is the inverse document frequency.

The higher the frequency of a word in the target text and the lower the frequency in other texts, the higher its TF-IDF weight, and the more it can represent the content of the target text.

6.4 Timing text visualization

Timing texts are temporal and sequential. For example, news will evolve, novel storylines will change with time, and comments on a specific news event on the Internet will evolve as the truth is gradually disclosed. This change needs to be reflected in the results when visualizing text with obvious timing information.

Next, let's introduce the theme river method[23]. Theme river is a time series data visualization method proposed by scholars, such as Susan Havre in 2000, which is primarily used to illustrate the evolution of a text theme's strength over time.

The visual example of the theme 'river' is shown in the figure below. The horizontal axis represents time, and the surging currents of different colors in the river represent different themes. The flow of

the surging current represents the theme change. At any point, the vertical width of the inrush indicates the strength of the subject.

6.5 Text relationship visualization

Text relationships include relationships within or between texts, as well as relationships between text sets. The purpose of visualizing text relationships is to present them. The relationships in text include the relationship between words before and after; the relationships between texts encompass the similarity of content between texts, quotations between texts, and other connections. The relationships between text sets refer to the hierarchy of content within the text set.

6.5.1 Graph-based text relationship visualization.

A word tree uses a tree diagram to display the appearance of words in the text, allowing for an intuitive presentation of a word and its context before and after. The user can customize the words of interest as the central node. The central node expands forward, encompassing the words in the text that precede the word; the central node expands backward, encompassing the words in the text that follow the word. Font size represents the frequency of words appearing in text. As shown in the figure, the word tree method is used to present all the words that are connected to the word "Child" in a text, both before and after it.

6.5.2 Graph-based text relationship visualization

The phrase 'network' encompasses the following two attributes. Node, representing a word or phrase. The connection line with arrows indicates the relationship between nodes. This relationship needs to be defined by a user. For example, "A is B", which is represented by a connection line. A and B are the two node words before and after is. A is in front of B, and B is behind A, then the arrow points from A to B. The wider the width of the connection line, the higher the frequency of the phrase appears in the text. The picture uses a phrase network to visualize the "*the*" relationship in a novel.

7. Conclusion

This paper systematically discusses the theoretical framework, technical process, and application practice of data visualization. We firstly introduce the basic concept and development context. This paper also examines the technological evolution of data visualization from early computer graphics to modern intelligent generation, revealing the internal logic of the four core processes: data collection, cleaning, mapping, and interaction. It emphasizes the decisive role of data quality and preprocessing in visualization performance. Secondly, the classification of data visualization is introduced.

In the future, the combination of artificial intelligence and machine learning, along with their continued development, will provide more possibilities for data visualization. With AI technology, data visualization tools can analyze data more intelligently and generate more insightful visualizations. For example, machine learning-based data analytics can help users uncover hidden patterns and associations in data, enabling them to design more effective visualizations. As visualization continues to mature, real-time data visualization becomes very important. With the advent of the significant data era, an increasing number of organizations need to monitor and analyze real-time data. Data visualization tools will evolve in the direction of real-time data visualization, which can update and display data in real-time to help users identify trends and anomalies in data promptly. That will significantly enhance the ability of decision-makers to understand and respond to data in real-time. In summary, the future development direction of data visualization will focus on combining technological innovation, user experience, and storytelling capabilities to provide users with a more intelligent, immersive, and convincing data visualization experience. Through continuous innovation and development, data visualization will play an increasingly important role in the future of data analysis and decision-making.

References

[1] Gjerloev J W. The SuperMAG data processing technique[J]. Journal of Geophysical Research: Space Physics, 2012, 117(A9).

[2] Thomas D B, Oenning N S X, Goulart B N G. Essential aspects in the design of data collection instruments in primary health research[J]. Revista Cefac, 2018, 20(5): 657-664.

[3] Howard J H, Kazar M L, Menees S G, et al. Scale and performance in a distributed file system[J]. ACM Transactions on Computer Systems (TOCS), 1988, 6(1): 51-81.

[4] Jones S. Text and context: document storage and processing[M]. Springer Science & Business Media, 2012.

[5] Zeng X, Hui Y, Shen J, et al. An empirical evaluation of columnar storage formats[J]. Proceedings of the VLDB Endowment, 2023, 17(2): 148-161.

[6] Robinson I, Webber J, Eifrem E. Graph databases: new opportunities for connected data[M]. " O'Reilly Media, Inc.", 2015.

[7] Fang J, Mulder Y T B, Hidders J, et al. In-memory database acceleration on FPGAs: a survey[J]. The VLDB Journal, 2020, 29: 33-59.

[8] Pundt H, Brinkkötter-Runde K. Visualization of spatial data for field-based GIS[J]. Computers & Geosciences, 2000, 26(1): 51-56.

[9] Few S, Edge P. Introduction to geographical data visualization[J]. Visual Business Intelligence Newsletter, 2009, 2.

[10] Jones R H. Fitting a continuous time autoregression to discrete data[M]//Applied time series analysis II. Academic Press, 1981: 651-682.

[11] Johnson N P, Denes P. The ladder diagram (a 100+ year history)[J]. The American journal of cardiology, 2008, 101(12): 1801-1804.

[12] Moritz D, Fisher D. Visualizing a million time series with the density line chart[J]. arxiv preprint arxiv:1808.06019, 2018.

[13] Caceci M S, Cacheris W P. Fitting curves to data[J]. Byte, 1984, 9(5):

[14] Wardell D G, Moskowitz H, Plante R D. Control charts in the presence of data correlation[J]. Management Science, 1992, 38(8): 1084-1105.

[15] Meng X L, Rosenthal R, Rubin D B. Comparing correlated correlation coefficients[J]. Psychological Bulletin, 1992, 111(1): 172.

[16] Carr D B, Littlefield R J, Nicholson W L, et al. Scatterplot matrix techniques for large N[J]. Journal of the American Statistical Association, 1987, 82(398): 424-436.

[17] Fortino V, Alenius H, Greco D. BACA: bubble chart to compare annotations[J]. BMC Bioinformatics, 2015, 16: 1-5.

[18] Scott D W. Histogram[J]. Wiley Interdisciplinary Reviews: Computational Statistics, 2010, 2(1): 44-48.

[19] Jung K, Kim K I, Jain A K. Text information extraction in images and video: a survey[J]. Pattern recognition, 2004, 37(5): 977-997.

[20] Kausar M A, Dhaka V S, Singh S K. Web crawler: a review[J]. International Journal of Computer Applications, 2013, 63(2): 31-36.

[21] Thomas D M, Mathur S. Data analysis by web scraping using Python [C]//2019 3rd International Conference on Electronics, Communication and Aerospace Technology (ICECA). IEEE, 2019: 450-454.

[22] Huang C H, Yin J, Hou F. A text similarity Measurement combining word semantic information with the TF-IDF method[J]. Jisuanji Xuebao(Chinese Journal of Computers), 2011, 34(5): 856-864.

[23] Havre S, Hetzler E, Whitney P, et al. Themeriver: Visualizing thematic changes in extensive document collections[J]. IEEE Transactions on Visualization and Computer Graphics, 2002, 8(1): 9-20.