Makrufa Sh. Hajirahimova[1], **Marziya I. Ismayilova[2]**   DOI: 10.25045/jpit.v09.i1.07
Institute of Information Technology of ANAS, Baku, Azerbaijan
[1]makrufa@science.az, [2]imarziya@gmail.com

BIG DATA VISUALIZATION: EXISTING APPROACHES AND PROBLEMS

One of the biggest problems of the century we live is a big data problem. One of its main problems is the visualization of the results of the analysis. The article reviews and interprets the history and phases of visualization, classification of visualization methods, existing approaches, problems of big data visualization, visualization tools.

Keywords: *visualization, big data, interactive visualization, scientific visualization, tag cloud, motion charts.*

Introduction

Since the early 21st century, digital data generated by advanced technologies have grown with a geometric progression every year resulting in the "big data" phenomenon. "Big data" is a huge amount of digital data collection unable to be processed through traditional databases and existing tools. The problem here is not just related to its volume, but also its different formats of the information, namely unstructured data generated at great velocity. Rapidly increasing data flow generated from various sources, such as social networks, artificial satellites of the Earth, sensor networks and so on [1, 2], makes the process of data management, including information and knowledge acquisition even more difficult. The application of data visualization tools plays an important role in eliminating the problem of data processing and analysis. The visualization emerged in the late 1980s as the main area of information science. It covers and integrates the areas such as computer graphics, image processing, computer visualization, signal processing, automated designing systems and human-computer interaction. Visualization also plays a role of bridge connecting a computer with a human vision system. It identifies descriptions, creates hypotheses, and develops ideas from large volumes contributing to scientific research and prediction. Although the use of computer technologies in the visualization began in the 1990s, visualization has an ancient history. For example, geographical maps, Mendeleev's periodic table, and so on [3, 4]. Data visualization is a process of interpreting the analysis results in different ways to provide a more efficient decision-making process. The expert on data visualization *Edward Tufte* says: "The world is complex, dynamic, multidimensional, the paper is static, flat. How are we to represent the rich visual world of experience and measurement on mere flatland?" Indeed, in modern times, data is easy to access, however, it is difficult to comprehend and understand it. That is why the improved visualization tools have become the main component of the modern society. Thus, the visualization of information can play a very important role in this process. A combination of properly selected illustrations, words, and figures can create a complete picture of the data perception. With regards to the big data processing, this is not a simple matter, and it requires exceptional methods and approaches. Graphic description is one of the most effective ways of objective evaluations and making the right decision in the problem solution.

However, in the case of big data, furthermost classic methods are ineffective, even impossible to apply in specific problems. Collection and management, storage, security, search, analysis of terabytes and exabytes of unstructured digital data generated every second, including the visualization and assessment of their results have become a serious problem.

Recently, a large number of visualization methods have been developed for rapid presentation of processed data. Nevertheless, new big data visualization methods form new research issues and solutions. At the same time, the characteristics of large volumes of data, such as volume, velocity, variety [5, 6], value and veracity [7], require flexible decisions. Therefore, it is important and relevant to study this area as a subject matter and explore its scientific and theoretical problems.

The concept, history, essence and stages of visualization

The term **Visualization** comes from the Latin word "visualis" and refers to the imagining, observation, and presenting the visual results of proper observation and analysis of digital information or physical events. Visualization has become one of the key characteristics of big data, such as volume, velocity, variety, value, veracity [8]. Recently, many international conferences, symposiums, workshops and forums dedicated to various aspects of information processing have been organized by authoritative international organizations and academic institutions. One of the main topics of these events is the visualization. It should be noted that the first visualization conference was held in 1990 by the Institute of Electrical and Electronics Engineers (IEEE) [3, 4]. Big data visualization has become the discussion topic of the international conferences held since 2013 within the framework of "Big data". One of the special releases of the 11th International Symposium on the Visual Computing held in the United States and organized by Springer in 2015, was devoted to "big data" visualization and analytics [9]. The issue presents the results of research, ideas, projects and programs in the field of "big data" visualization and analytics for the experts of industrial and research institutions. Big data visualization, big data processing for visual calculations, visual big data analytics, scientific and information visualization, image processing and computer visualization and so forth were the symposium topics.

In general, data visualization is not a new field. Thus, visual transmission of information is very ancient (for example, ancient carvings on stones, Egyptian hieroglyphs, Greek geometry, revolutionary painting techniques by Leonardo da Vinci, technical drawings for engineering and scientific purposes, etc.). American psychologist Michael Friendly structures the history of data visualization into the following stages [10-12]:

- until XVII century – early maps and diagrams;
- 1600 -1699 - measurements and theories;
- 1700-1799 - new graphic forms;
- 1800 -1850 - beginnings of modern graphics;
- 1850 -1900 - golden age of statistical;
- 1900 -1950 – dark ages;
- 1950 - 1975 - Rebirth of data visualization;
- 1975 - present times - interactive and dynamic data visualization.

Early maps and diagrams - The starting point of visualization is the illustration of the stars' positions in the geometric diagrams and tables. The early images display a graph of the displacement of stars in the two-dimensional coordinate system of the planets. This graph illustrates a clear imagination of that period.

During the period of *measurements and theories*, scientists focused on measuring time, distance, and space. Coordinate system emerged and the probability theory and demographic statistics were developed during that period. The illustration of that period refers to the work by Christopher Scheiner (1630).

In the period of *new graphic forms*, the geographical locations on maps were depicted not just as point-shaped, but also as contours. Moreover, those maps highlighted thematic, geological, economic, medical illustrations. As the abstract visualization expanded, more information was gathered, and consecutively, new visual forms were required for their description.

During that period, "chronological maps" based on the time scale were proposed by *Jacques Barbeu-Dubourg*. In parallel, *Joseph Priestley* described the biography of 2000 well-known people on the time scale. Shortly afterwards, he introduced his own time development version. Later, he used the geometric figures such as squares, triangles, circles to compare the fields and demographic sets. At the same time, tricolored publications enabled color-coding. Another successful example of that period was the foundation of millimeter papers to make graphs easier.

In the *beginning of modern graphics*, circular, columned and field diagrams have occurred. The starting point of modern graphics is the visualization of the cholera of the famous streets of London by

John Snow. The pie chart was created in 1858 by the British public figure Florence Nightingale to present the data about the soldiers fighting in the Crimea war, who died from illness (highlighted in blue), at the battlefield (highlighted in red), and for other reasons (highlighted in black).

Golden age of statistical graphics. In the mid-1800s, the "ideal storm", i.e. favorable condition was created for data graphics and a rapid increase in the visualization. Official state statistics bodies were established throughout Europe due to the significant rise of digital information for social projects, industrialization, trade and transport. Established by Gauss and Laplace and spread in social media by Guerry and Quetelet, the statistical theory includes the methods for the perception of large databases. In fact, this century was called the golden age with many innovations, creating a spirit of excitement for graphics and thematic cartography.

Dark ages. The visualization of the early 1900s can be called a "dark age" in the background of the previous "golden age" in statistical graphics and thematic cartography. Compared to innovations in the field of graphics, the quantitative and formal models rapidly developed in those years. During that period (1920-1930), the founder of the "isotype" *Otto Neurath* as a visual analyst and philosopher visualized the statistics of modern economies for visual demonstration of social inequality. After that, the interest in the *infographer* increased again, which became a springboard for future growth.

In 1950s and 1975s, scientists and writers began to popularize the visualization ideas. In 1962, *John Tukey* calls for recognizing data analysis as a legitimate area of statistics that is different from mathematical statistics in his article titled "The Future of Data Analysis". Later, he discovers a number of new, simple, and effective graphic views (stem-leaf charts, box plots, binary table, etc.) in his work "Exploratory Data Analysis". At this point, the first interactive visualization has emerged.

Robert Kosara describes the next stage of visualization as a fairy-tale in the article "Storytelling: The Next Step for Visualization". The author believes that it is always characteristic for people to co-ordinate historical facts, as it helps to better present and store information. This also indicates the future of visualization. *Daniel Keefe* and *Tobias Isenberg*, in the scientific article "Re-imagining the Scientific Visualization Interaction Paradigm", imagine the future of virtual visual data as an area in which perfect computer techniques can be combined with natural human relationships and visualization communications.

Today, a wide range of data analysis and visualization tools has set the stage for interactive and dynamic visualization. All of this is due to the availability of interactive systems, the capabilities of the 3D models, the increased computing power and, in particular, the data accessibility due to the Internet availability.

At present, fundamental studies are conducted in the field of big data visualization in well-known scientific centers accross the world. For example, White Paper by Intel, and studies conducted by the computerized visualization and human-computer expert B.B. Shneiderman and professor of the University of California, leading specialist in big data visualization Kwan-Liu Ma [13-17].

As a result of our research on a number of scientific bases, the dynamics of studies by years is presented in the following graph (Figure 1).
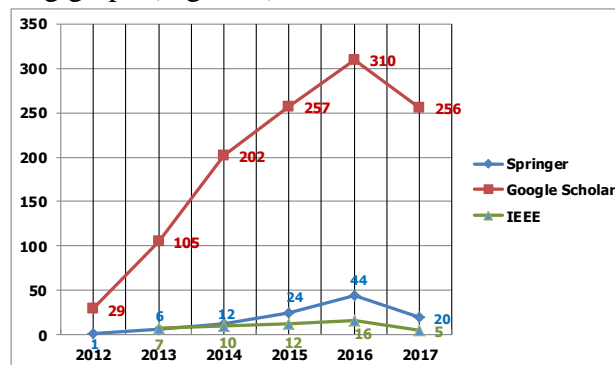

Figure 1. Distribution of Big Data visualization research in Springer, Google Scholar and IEEE databases.

Traditional data visualization methods and existing approaches

Information is now an integral part of human life. A large number of information or data is generated from different sources, and every year people have to deal with this information. Presenting the results of data analysis is very important for better understanding. Traditionally, various methods have been used to visualize the data. These visualization methods are classified by three major groups, as static, dynamic, and interactive. There are many approaches to these methods. It should be noted that some of these approaches may be used in all three groups. Here are some of them:

The line graph shows the interaction of one variable with another. Such graph is often used to track the changes that have taken place over a long period of time, and to compare several objects at the same time. The line graph is used for the visual description of one or several variables and the rate of change of information about the content of these variables.

It should be noted that, if certain data points are known, it is not necessary to select only linear graphs for them. In this case, rather simpler image type can also be chosen, for example, listing as a table in a certain order. Whilst using a line graph, the connection between points must be considered [17, 18].

Bar charts are often used to compare quantitative indicators of data of different groups and categories. The contents of the categories are displayed in bars, they can be vertical and horizontal, and their height and length correspond to certain data values. If these prices are fairly different, the difference in bars will be obvious, and a simple bar chart can also be used. When the values are very close to one another or the numbers are high, their difference becomes difficult. For this purpose, the bars can be presented in different colors.

The other form of the histogram is the progressive chart or *waterfall* chart. This type of charts shows how the initial value of the data changes (increases or decreases) during a number of operations. The first bar starts with an initial value, and each subsequent bar starts at the end of the previous one. The length and direction of the bar indicate the amount and type of the process. The resulting cascade shows how the process or operation leads to the final outcome [17-19].

Scatter plot uses the cartesian coordinate system to describe the data set of mostly two variables of mathematical diagrams. In some cases, the number of variables can be increased to 3 by using color icons. With this, the direction and linearity of the dependence between the variables shown in the diagram can be determined. The more the number of points in the chart increases, the higher the correlation rate is. Scatter plot is mainly used for the visualization of multidimensional data. However, 2D descriptions have a large number of problems. The 3D expansion in the visualization of interactive data of the scatter plot is called regressive cube. They dramatically increase 3D diagrams from three aspects. Here are correlations with the help of sensitive lines and sensitive flow line between the two variables. Regression cube enables the study of interactive visualization of multidimensional data sets through various classification and information search [17, 18].

Bubble plots are a type of dots diagram, in which the dots are shown in the form of water bubbles and differ from one to another by their size and position. Here, along with *X* and *Y* axes, *Z* is also used. Bubble plot is often used to present financial data. The difference in the sizes of bubbles visually separate the actual value [17, 18].

Pie chart describes the data as a circle segment or as a percentage. With the help of it, the data is presented in only one order and indicates the ratio of the size of the elements to their sum [17, 19].

Many approaches are offered to improve the pie charts. Since it is not so easy to compare the diagram parts by the human eye, and there are problems with the interpretation of the results.

Tree map displays all data as the components of a hierarchical tree. It presents rectangles associated with the branches of the tree, which differ from each other by color and size. This depends on the given parameters. This method accurately illustrates the relation of the data; however, it often happens over a certain period of time [17, 20].

Sunburst is relatively newer method rather and often used as an alternative to "tree maps". The main difference between these methods is that the data differ not only by the height and width but also by the radius and length of the branches. Accordingly, a sector with new data can be changed by changing the radius, while the whole scheme is remained unchanged. Due to this feature, the method can be adjusted to show the dynamics of data with the use of animation [20].

The most commonly used types of *maps* are geographical, photographic, road and e-thematic. Obviously, geographical maps are used in the schematic representation of the geographical objects. Photographic maps are the photo representation of the geographical objects from the satellite. Road maps show the schemes of highways, railways and other roads [21].

Depending on the application areas, the visualization methods are grouped as follows [3-4, 17, 22, 23]:

- Scientific visualization;
- Information visualization;
- Software visualization.

The intensive development of computational techniques and mathematical modeling methods has created a new domain namely, scientific visualization, which was consecutively emerged as a separate discipline in 1980. Scientific visualization helps scientists and engineers more effectively comprehend the physical events hidden in large volumes of data. The data can be acquired from complex modeling or from a medical scanner, telescope, satellite system. The distinctive feature of the scientific visualization is related to the physical properties of the visualized objects with the natural similarities. For example, the Earth, human body, molecule, deoxyribonucleic acid, and so forth. The development of the mathematical models for the description of physical objects plays an important role in the representation of information. Color and other visual signs are used to describe different attributes of the physical objects. "*Isosurfaces*", "*volume rendering*" and "*glups*" are the most widely used methods for providing scientific visualization attributes. "*Isosurfaces*" refers to the distribution of attributes, for example, in geographic maps, colored contours show the ranges of the change in temperature. "*Volume rendering*" presents three-dimensional data in full volume through interaction. For example, the medical data generated from the magnetic resonance tomography, the data of automated design systems, distance learning systems, and so forth. *Glups* (description of special characters) present many attributes through a combination of different visual signals. The symbols here are usually used to describe the information flood. The most common character is an arrow. For example, the arrows illustrate the direction and strength of the wind on the map.

Information visualization was shaped as part of the human-computer interaction in late 1980. Information visualization is a field of interdisciplinary research and based on the knowledge about visualized data and processes, and studies visual perception of information and the acquisition of mathematical methods of data analysis by human. Information visualization helps users to identify descriptions, correlations, or clusters. Information visualization is applied either to structured or non-structured information. Structured information is presented in the form of numbers with predefined variables. For example, business transaction data, Internet traffic data, and data on Internet access, etc. Initially, standard static graphics such as line graphs, scatter plots, bar charts, and pie charts were used to improve data compression.

At present, big data visualization applies various data mining to maximize the identification of common images out of data. Computerized visualization is a tool for delivering images or structures identified by data mining algorithms. Non-structured information does not have precisely defined variables and attributes. For example, office documents, web sites or email archives, and other data sets. In this case, it is required to define and analyze the variables before the graphic design is formed.

Software visualization is defined as follows: Software visualization is a discipline that uses various forms of images to minimize the complexity of existing software system and provide mutual comprehension and perception.

Software visualization usually consists of the visual presentation of either the software code or the algorithm animated. This also helps engineers manage software development, configuration, and optimization.

Big data visualization problems

One of the key issues in big data analysis is the presentation of results, i.e. visualization. In general, data visualization is one of the simplest and most regular methods in data processing and analysis. It helps to comprehend presented information more quickly and to make the optimal decisions by evaluating the results [2, 3].

The use of traditional methods to visualize large-scaled and amount of data is ineffective. Studies show that the foremost scientific and theoretical problems of big data visualization, which has the features such as volume, velocity and variety are as follows [20, 24-27]:

- *Visual noise*. This problem occurs when the objects are over-connected with one another in the data set. Visual noise is not a distortion or damage of data, but a minimization or a loss of images on the screen. This complicates the extraction of useful information out of whole image and requires additional processing;

- *Large image perception*. The above problem can be solved by distributing the data on larger screens. However, sometimes the problem of perception of a large image may occur. The human brain can perceive a visual image at a certain level. Though the level of perception of the graphical visualization of data is higher than the table visualization, there are certain restrictions. Thus, after a certain level, human loses the ability to percept the information in addition to extremely overloaded visual data. The capabilities of all visualization methods are limited to the capabilities of the technical devices that provide the presentation of the data. Unquestionably, although we use a more modern device, we still face the limitations of comprehension abilities of a human. Hence, the data visualization is limited not only to the capabilities of the devices, but also to the physical perception of a human. Data filtering, i.e. diminution approach is applied for the solution of the mentioned problem;

- *Information loss*. This problem arises from the solution of the problem of visual noise and large image perception. The approaches to the problem solution reduce the end data, thus creating a new problem of information loss. Since the visual information reduction methods are based on the proximity of the objects, they data are aggregated and filtered according to one or more criteria. These approaches can mislead the analysts, remaining the hidden and more important objects unnoticed. Furthermore, the aggregation of data may require a great deal of time and computing resources to get accurate and relevant information;

- *High performance requirements*. Graphic analysis is not limited only to the static visualization of images; it is also used for dynamic visualization, which leads to unnoticeable problem in the static visualization. At the certain rate of the visualization, there arises a need for high performance. Because the process of analysis requires a large amount of computing resources and time for filtering large numbers of data;

- *High rate of image change*. As it seen from the name, this problem is due to the rapid change of images. That is, during the observation, a person simply can not react to the rapid change of data or their intensity on the screen. Reduction of changed data can not provide the desired effectiveness of the process. However, the rate of human reactions causes certain limitations in this process.

For the solution of these problems, new methods and technologies should be developed and a qualified staff should be organized.

New methods and tools for the visualization of large data

The visual representation of large data principally aims at:
1. Detecting the confidentiality and anomalies of data;
2. Increasing flexibility in the search of certain values;
3. Comparing the various blocks in order to get the relevant difference in the volume;
4. Establishing human relationships for real-time mode (excursion, scaling, etc.).

However, many of the classic visualization methods are unable to solve specific problems of large data, which has led to the emergence of new visualization tools and techniques.

Here are some of the most commonly used methods for the visualization of large data:
- *Tag Cloud* – a term describing the content of the document or set, or a visual description of popular Web 2.0, which consists of short phrases. Typically, the keywords or referenced objects derived from the content of the document are described using natural language processing techniques. The size, color, and position of words are governed by their importance, aesthetic and comfort criteria. This method adopts a weight coefficient for each element. The higher the coefficient, the higher the font size will be. Weight coefficient depends on the importance of the element determined by the expert, the frequency of its condition and other factors [28, 29].
- *Clustergram* - a visualization method used to demonstrate how the elements of data are integrated into the clusters as the number of clusters increases in the course of analysis. The choice of the number of clusters is an important parameter for the cluster analysis [22, 30].
- *Motion charts* - enables to effectively explore and interact with large and multidimensional data using two-dimensional bubble charts. Bubbles (main objects of this method) can be controlled in accordance with the considered changing image [22]. Graphic data tools for motion charts are presented in Google [31], amCharts [32] and IBM Many Eyes [33].
- *Dashboard* - displays log files of different formats and filters them based on selected data ranges. The dashboard consists of three layers, namely data (raw data), analysis (data transmitted from formulas and data layers to tables) and presentation (graphical presentations based on analysis layer) [22].
- *History flow* - describes the evolution of a document that is modified by many authors. It enables to monitor which author the document is edited by, what is added to the document, and the time spent. Here, the horizontal axis denotes the time, while the vertical axis indicates the text inserted. Each author has a different color code, and the vertical length of the string shows the number of text written by each author [28].

Visual representation of large data analysis is very important for its interpretation. As noted, human perception is limited. Effective visualization tools should take into account the characteristics of the human brain, such as comprehension and perception. Advanced methods for data visualization aim at the improvement of images, charts and animations. Though, big data recovery tools are capable to process petabytes (PB) and zetabytes (ZB) of data, sometimes they can not visualize them. Since large data is expanding at an extremely high speed, the lack of scalable tools for visualization complicates the process of acquiring confidential information. Various visualization tools are used to overcome the above-mentioned problems. Some of them are described below:

Tableau provides extensive opportunities for interactive visualization for business analytics-oriented data visualization. It is fast and flexible, the user interface is intuitive and many charts are available. No coding skills are required for simple calculations and statistics. However, the programming language and environment are enabled in R only for model analysis for challenging analytics, data analysis, statistical calculations and data visualization, and then the results it can be switched to Tableau. Tableau consists of three basic tools, namely Tableau Desktop, Tableau Server, and Tableau Public:

- Tableau Desktop easily visualizes and views data in a different and intuitive way;
- Tableau Server is a business analytics system that provides analytics based on a browser;
- Tableau Public is used for building interactive effective visualization [8, 24-26, 34].

For example, the eBay online trade center uses Tableau as one of the major data visualization tools to provide all large and complicated set of data in an understandable format [5].

Microsoft Power **Business Intelligence (BI).** It is a powerful cloud-based business analytics service with interactive and rich visualization opportunities. Power BI consists of three elements: Power BI Desktop, software as a service (SaaS) and Apps. Each service is accessible and therefore flexible and handy. Power BI incorporates up to 60 types of sources, and may start visualizing within minutes. It also combines Microsoft Office, SharePoint, and SQL. This visualization tool does not need programming skills for surveys: it supports the natural language [8, 34].

Plotly, also known as Plot.ly uses Python programming language and the Django web application for data analysis and visualization. With some restricted options, Plotly is freely available for users. These options require professional membership. Charts and dashboards are created online through these options. The charts may be statistical, scientific, 3D, information panels, etc. Moreover, Plotly uses the DWP (Digitizer Web Plot) tool to automatically retrieve statistical image data [8, 34, 35].

Excel 2016, produced by Microsoft, is not only an online table for large data and statistical analysis, but also a powerful tool for visualization. Using the query capability, Excel incorporates HDFS, SaaS, and other services and manages semi-structured data. Together with visualization methods, such as conditional formatting and interactive graphics, Excel 2016 is a powerful competitor among the Big Data visualization tools [34].

ManyEyes is a product of the International Business Machines (IBM). Through this website, users can perform interactive visualization by downloading data [24].

Raw is a simple tool that deploys data in a simple way. A very simple step is required to create a graphic. Raw uses the D3.js library [8].

Leaflet runs with mobile and desktop computers and visualizes the data generated from speech and high traffic. The tool includes a JavaScript library that enables users to create interactive maps [8].

Dygraphs is a flexible and flexible open-source Javascript graphics collection, which is used to visualize large volumes of data. This visualization tool requires programming skills [8, 24].

Gephi is an open-source network analysis tool written in Java and OpenGl. This network analysis includes social network analysis, connection and biological network analysis. This tool, designed for a large amount of complex data, provides interactive visualization and the study of dynamic data [8, 35].

Data-Driven Documents (D3) is a small free JavaScript library for processing HTML documents based on data. D3 performs high-speed visualization of html or svg data and manages interactivity [36].

Conclusion

Many studies hetherto have shown that visualization is one of the most promising aspects of data analysis. There are technical, technological and scientific-theoretical problems in this field. Another problem is the selection of corresponding visualization methods and tools to solve a certain problem.

Generally, the problem of data visualization is related not only to the development of existing technologies, but also to human-oriented problems. The data visualization method can almost be improved through basic psychological principles and interactions that are more natural. Data visualization should be intuitively comprehended in terms of flexibility and design. The user must understand not only the process, but also the scalability of any visualization issue. The current developments in the field of big data visualization focus on the creation of tools to enable users to achieve fast and effective results when working with large volumes of data.

Consequently, we can conclude that computer graphics widely used for the visualization of analyzed data should provide clearer, faster and low-cost solution through different methods and algorithms. Graphics is not just a way of visualizing the results, but a perfect tool for working with the results. Today, the results of many methods can be represented in a three-dimensional field (e.g. associative rules) more clearly.

Large volumes begin to grow at very high speeds as tsunami, and existing visualization tools may not be used in the near future. In this regard, new technologies and tools will be required for the solution of big and unexpected problems for the storage and visualization of extremely large amount of data as of zetaabytes or yotabytes in the future.

Finally, we can note that some large data visualization problems, such as perception of large images, information loss of information, high performance requirements, high speed image change still remain unsolved, and are of the future subjects of research.

References

1. Gants J., Reinsel D. The digital universe in 2020: Big Data, Bigger Digital Shadows, and Biggest Growth in the Far East. Study report, IDC, 2013, 7 p.
2. Alguliyev R.M., Hacirahimova M.Sh. "Big Data" phenomenon: problems and opportunities // Problems of Information Technologies, 2014, No2, pp.3-16.
3. Information visualization (INFOVIS) // IEEE Symposium. http://ieeexplore.ieee.org
4. Peskova O.V. About the data visualization // Vestnik MSTU after. Bauman N.E., "Instrumentation", 2012, pp.158-173.
5. Chen C.L., Zhang C.Y. Data-intensive applications, challenges, techniques and technologies: A survey on Big Data // Information Sciences, 2014, vol.275, pp.314–347.
6. Beyer M.A, Laney D. The importance of "Big Data": A definition. Gartner, 2012, 7 p.
7. Demchenko Y., Ngo C., Membrey P. Architecture framework and components for the big data ecosystem // Journal System and Network Engineering. SNE Technical Report, 2013, 31 p.
8. Mashkoor A., Ahamad M.V. Visualization, Security and Privacy Challenges of Big Data // International Journal of Advanced Technology in Engineering and Science, 2017, vol.5, no.6, pp.394–400.
9. A Special Track of the 11th International Symposium on Visual Computing, 2015. www.isvc.net/15/ST6.pdf.
10. Friendly M. A. Brief history of data visualization, Springer, 2006, 41 p.
11. Kumar S. A review of recent trends and issues in visualization // International Journal on Computer Science and Engineering (IJCSE), 2016, vol.8, no.3, pp.41–54.
12. Azzam T., Evergreen S., Germuth A.A., Kistler S.J. Data visualization and evaluation // New Directions for Evaluation, 2013, no.139, pp.7–32.
13. Averbukh V.L., Manakov D.V. Analysis and visualization of "big data" / Proceedings of the Conference "Parallel Evaluation Technologies", 2015, pp.333-340.
14. Big Data Visualization: Turning Big Data into Big Insights. The Rise of Visualization-based Data Discovery Tools, White Paper. Intel IT Center, 2013, 14 p.
15. Shneiderman B. The big picture for big data: Visualization // Science, 2014, vol.343, pp.730.
16. Keim D., Qu H., Ma K-L. Big Data Visualization // IEEE Computer Graphics and Applications, 2013, pp.20–21.
17. Khan M., Khan S. Data and Information Visualization Methods, and Interactive Mechanisms: A Survey // International Journal of Computer Applications, 2011, vol.34, no.1, pp.1–13.
18. Data Visualization Techniques. White Paper, SAS Institute, 2014, 17 p.
19. Kaushik A., Naithan S. An Anatomy of Data Visualization // International Journal of Computer Science and Network Security, 2016, vol.16, no.2, pp.77–82.
20. Gorodov E., Gubarev V. Analytical Rewiew of Data Visalization Methods in Application to Big Data // Journal of Electrical and Computer Engineering, 2013, pp.1–7.

21. Goranson C., Huang X., Bevington W., Kang J. Data Visualization for Big Data. 2014, 26 p.
22. Olshannikova E. et al. Visualizing Big Data with augmented and virtual reality: Challenges and research agenda // Journal of Big Data, 2015, vol.2, pp.2–22.
23. Hacirahimova M.S., Ismayilova M.I. About the problems of visualization of software versions / I Republican Conference on the Actual Scientific-Practical Problems of Software Engineering, 17 May 2017, pp.254-258.
24. Wang L., Wang G., Alexander C. Big Data and Visualization: Methods, Challenges and Technology Progress // Digital Technologies, 2015, vol.1, no.1, pp.33–38.
25. Patil S.S. Overview of Big Data Visualization // International Journal of Advanced Networking & Applications (IJANA), 2016, pp.436–438.
26. Chen C.L., Zhang C.Y. Data-intensive applications, challenges, techniques and technologies: A survey on Big Data // Journal Information Sciences, 2014, pp.314–347.
27. Alguliyev R.M., Hacirahimova M.S., Aliyeva A.S. Actual scientific and theoretical problems of Big data // Problems of Information Society, 2016, No2, pp.37-49.
28. Manyika J., Chui M., Brown B. et al. Big data: The next frontier for innovation, competition, and productivity. Analyst report, McKinsey Global Institute, May 2011, 143 p.
29. Seifert C., Sabol V., Kienreich W., Lex E., Granitzer M. Visual analysis and knowledge discovery for text. In Large-Scale Data Analytics, Springer, 2014, pp.189–218.
30. Schonlau M. Visualizing non-hierarchical and hierarchical cluster analyses with clustergrams // Journal of Computational Statistics, 2004, vol.19, no.11, pp.95–111.
31. http://developers.google.com.
32. www.amcharts.com.
33. Vigas F, Wattenberg M. IBM—Many Eyes Project, 2013, 7 p.
    http://hint.fm/papers/viegasinfovis07.pdf.
34. Ali S.M., Gupta N., Nayak G.K., Lenka R.K. Big Data Visualization: Tools and Challenges / 2nd International Conference on Contemporary Computing and Informatics (IC3I), 2016, pp.656–660.
35. Automatically Grab Data From an Image with WebPlotDigitizer.
    https://plotlyblog.tumblr.com/post/70293893434/automatically-grab-data-from-an-image-with.
36. Agrawal R., Kadadi A., Dai X., Andres F. Challenges and oppotunities with big data visualization / Proc. of the 7th International Conference on Management of computational and collective intElligence in Digital EcoSystems, 2015, pp.169–173.