Research on Computer Information Processing Technology in the Age of Big Data

Yingli Wang[a], Zhicai Hu[*, b]

Guangzhou Huali Science and Technology Vocational College, Guangzhou 510000, China.

[a]105624725@qq.com, [b]1296566925@qq.com

*Corresponding author

Keywords: Big data; Computer information technology; Data unit; primitive relationship; Data unit; Visualization

Abstract: In the era of big data, users' access to data information has changed from simple to complex, and from passive to active. A large amount of data covers a large number of valuable information. However, the amount of data is large, and the display area area limitation often causes effective information to be often concealed, cannot be directly obtained, and increases the difficulty of searching. In particular, information with high dimensions, multiple levels, and complex transformations, how to use computer information processing technology to help users quickly obtain information and analyze data is a problem that needs to be solved in the era of big data. To this end, the paper uses data units, primitive relationships and data units to construct data structure methods, establish information models, and propose innovative thinking.

1. Introduction

The "big data era" has great value for use because of its own characteristics. A large number of data carriers and new technologies for big data forecasting of cloud computing, sensing solutions and distributed processing technologies have fully promoted new changes in productivity, technology and human activities. Use big data to maximize wind and power forecasts and optimize grid performance; Pretence uses big data to predict lost users and provide developers with optimal operating options for potential paying users: Baidu will play tens of thousands of games the event data is used to predict Brazil's World Cup prediction model. In order to better present data, visualization methods have emerged. With the expansion of application platforms, the increase of application fields, dynamic effects, user interaction, etc., data visualization requires innovation and usability. However, the massive amount of information resources for users, if there is no efficient and convenient analysis and processing means, it is like a mess, no effect [1]. Different from the data analysis and processing perspective in the computer field, from the perspective of the interaction design field, how to make users use the data better, how to make the data have better rendering effect on the interface platform is the main content of our concern. Therefore, we need to find a reasonable and effective data structure split method and data visualization structure model construction method.

2. User-oriented big data high-dimensional information

The complexity of high-dimensional data makes existing statistical and analytical methods difficult to cope with, and the visual presentation technology of information also encounters challenges. Due to different data collection sources and different types of structures, traditional and single visualization methods cannot cope with the analysis of such data; the high complexity of data puts pressure on processing technology, and the traditional single-machine, external-storage model and even small computer group processing capabilities It cannot be processed effectively; the fidelity problem in the data processing process becomes more and more difficult as the data complexity increases [2]. Dimension reduction, visual optimization, and different transmission methods all have an impact on the quality of the data that ultimately reaches the user's hand. Therefore, it is essential to fully understand the content and characteristics of high-dimensional data, to explore appropriate visual presentations, and to conduct appropriate user research for the audience of such data. However, compared to list-type numerical information, users are different from professional researchers, and they need finished products that are easy to understand and operate.

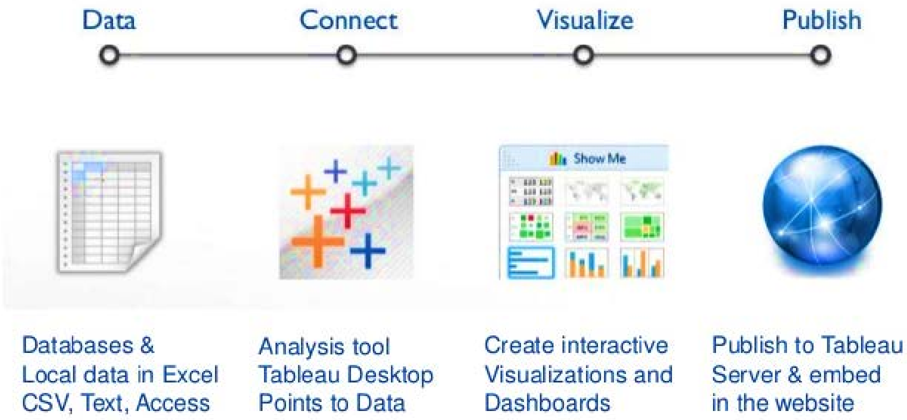

Fig. 1 Data visualization process

2.1 Interpretation and characteristics of high-dimensional information

High-dimensional multivariate data refers to data with two or more independent or related attributes per data. Therefore, the high-dimensional data we study refers to data with two or more independent or related attributes for each data object. Data with independence and relevance is more objective. The concept of high-dimensional data is a good explanation of the data characteristics of the current big data era [3]. However, the high Dimension of such data also causes Dimension disasters. That is, as the number of dimensions increases, the algorithms that can be used by the machine cannot be directly processed directly, and the resulting data processing is incapable. In addition to the dimensional disaster, the cumbersome order of missing data, duplicate data, invalid data, and massive data in the data brings trouble to the data processing. However, if the existing data visualization method does not filter, simplify, and highlight the data well, Such a presentation can cause visual confusion for the user. Therefore, in addition to being an effective means of presentation and interaction, data visualization itself can also achieve the role of data cleansing and simplification, and truly present appropriate data information to users in complex high-dimensional data sources.

2.2 Structured and unstructured data definitions and characteristics

The nature of the structured information is relatively stable, and the position of the magnitude is relatively fixed. For example, form information such as e-commerce information, such data is easy to manage and readily available; semi-structured headings and body grammar are very standardized, and the scope of keywords is quite limited. Although semi-structured data is more versatile and more flexible than the former, there are still many limitations; unstructured information is more common in today's Internet age, such as blogs and forum information [4]. All content is unpredictable, and this high degree of variability and flexibility makes it difficult to accurately obtain and efficiently process such information.

3. Unit selection and architecture method of data structure

3.1 Primitive relationship

Meta data is data used to describe data, that is, structured encoded data, used to describe features that carry information entities in order to identify, discover, evaluate, and manage the entities being described. Typically, we refer to the description of multiple attributes of a data element as Meta data for that data element. A data structure refers to a collection of data elements that have one or more specific relationships with each other and is a way for a computer to store and organize data. A primitive diagram is a diagram used to characterize a data structure, a view that represents the relationship between elements and elements in dotted lines. Optimizing the data structure to make it scientifically sound is critical to improving the efficiency of data analysis and processing.

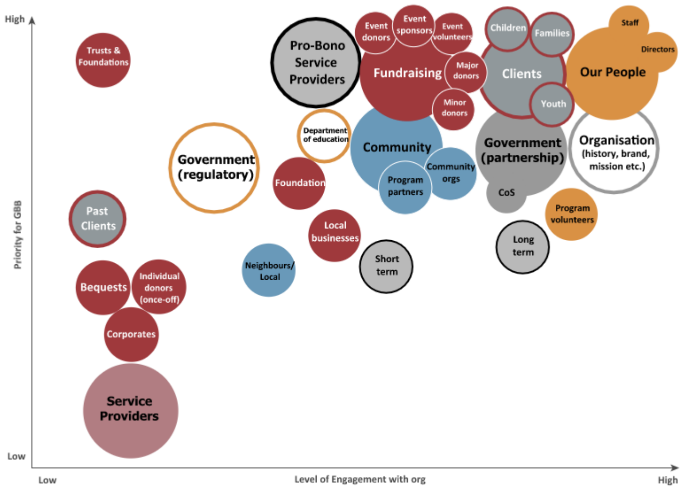

Fig. 2 Primitive relationship

3.2 Visualization Structure and Modeling Experimental Background

Meta data is the basic unit of the data structure architecture, and the data structure is the building unit of the data visualization structure model. The architectural phase of the data visualization structure model is set before the data visualization visual processing stage, helping the designer to better understand the relationship and characteristics of the processed data, so that the design expression is more reasonable [5]. The main purpose of this evaluation experiment is to verify the availability of the “task-level discrete data visualization structure model”. It is verified that the availability of the model is mainly from verifying the effectiveness of its two innovations, namely, verifying the validity of “information layering” and the effectiveness of “visual point” from a cognitive perspective. Therefore, in the next chapter, this verification evaluation experiment sets a total experimental content, that is, the “stratified processing performance evaluation” experiment: comparing the “multi-user information co-planar” and “information hierarchical paging”, Try to complete the performance of the same visual search task to evaluate the build availability of the "task-level discrete data visualization structure model."

4. Experimental study on data visualization structure model of high-dimensional information interface

4.1 Experimental methods

The main purpose is to verify the "task hierarchy separation" by comparing the eye movement search situation and the target node positioning time when the user completes the same visual search task on the "multi-user co-planar" information page and the "hierarchical facet" information page. Effectiveness. The visual search efficiency of the subject is judged by measuring the amount of time before the subject is positioned to the target node, and whether the search of the subject is valid is determined by comparing the number of points of interest around the specific node. These all require a lot of eye movement data to support [6]. Eye movement tracking technology is a process evaluate technique that helps students understand how image information is understood. The movement of the eyeball represents the direction and speed of the line of sight. By capturing the information of the trajectory, we can understand the order, habits and points of interest of the observer. At the same time, the gaze point, Muscadet process and scan path experimental parameters in the experimental data, combined with the actual observer behavior, can help construct the cognitive behavior model of the subject and analyze the characteristics of the cognitive process of the subject and even explore the subject's cognition. Potential needs. In addition, through eye tracking technology, we can classify the observer's visual information capture process according to the trajectory, analyze the characteristics of each stage and find out the key links that we can optimize. Eye tracking technology combined with the design of scientific experimental methods can help us get the real-time scheduling of the observer's attention, and these rich data is the main basis for us to study the visual behavior characteristics of users. Therefore, the availability of this “task-level discrete data visualization structure model” was evaluated using an eye tracking experiment.

4.2 Experimental design

4.2.1 Experimental object

The total number of subjects in this experiment was 82. The number of people who successfully completed the experiment was 80. Participants were divided into four groups of 20 people each, which participated in the experimental group and the control group of the group A experiment, the experimental group and the control group of the group B experiment. The subjects had normal or corrected visual acuity and no color blindness or color weakness. The basic information is as follows.

4.2.2 Experimental instruments

The X2-30Compact portable eye tracker produced by Tobit Technology in Sweden was used in this experiment. The device is connected via a single USB cable and can be easily attached to a notebook, computer screen or even a tablet for a compact, highly portable eye tracking solution, usability testing and market research in the field. Ideal product. Compared with other eye trackers, the device can allow a large range of motion of the subject's head without affecting the accuracy of the test results, and adapt to different experimental conditions and different subjects. In addition, the tracking and calibration effect is stable and reliable, which can create a natural experimental research environment for the subjects, thus ensuring the naturalness of the test behavior and the validity of the research conclusions. The windows7 system is equipped with a computer host; the screen ratio is 4:3 hp display.

4.3 Experimental procedure

The purpose of the “hierarchical processing performance evaluation” experiment is to verify that the hierarchical division of complex information can effectively reduce the information complexity at the unit level, improve the information processing performance within the hierarchy and the overall task completion efficiency. This experimental task is set to sequentially search for target nodes of each level, such as A43, B12, C12. By setting the same visual target search task, the observer's search behavior characteristics and performance are compared under the hierarchical presentation of information and the co-planar presentation of information [7].

4.3.1 Experimental material

The number of high-dimensional information nodes is large, and when the number of information dimensions reaches 3 or more, it is high-dimensional data. Therefore, according to the minimum dimension number 3 of the high-dimensional information, the number of nodes of the experimental graph node is also 3, that is, the node type is set to three hierarchical nodes of A, B, and C, which are the first hierarchical node and the second hierarchical layer. Node and third level node. At the same time, the structural relationship mode between the node levels is set to a tree structure relationship, the node C is a child node of B, and B is a child node of A. According to the rule of tree node relationship, the setting of the B-level node is within the range of the A-level node, and the setting of the C-level node is within the range of the B-level node. Therefore, the experimental set A, B, C three levels of node representation size is reduced in size, the ratio is 16:8:1, and the number increases. According to the short-term memory capacity of the person is 7±2 chunks, the ratio of the number of A, B, and C nodes in the experiment is 1:5:25, totaling 2170 nodes [8].

According to the design idea of the experimental group and the control group, each task map contains three node pages, which are A-level node page, B-level node page and C-level node page. At the same time, when the next level page appears, the upper level node is hidden except the target node. The retention of the target node helps to prompt the hierarchical attribute of the current node page, the relationship between the current node page and the upper and lower level pages, and the search path of the current task. As shown in FIG. 3, in order to control the irrelevant variables, the spatial positions of the three hierarchical nodes A, B, and C remain the same and unchanged in the experimental group and the control group.

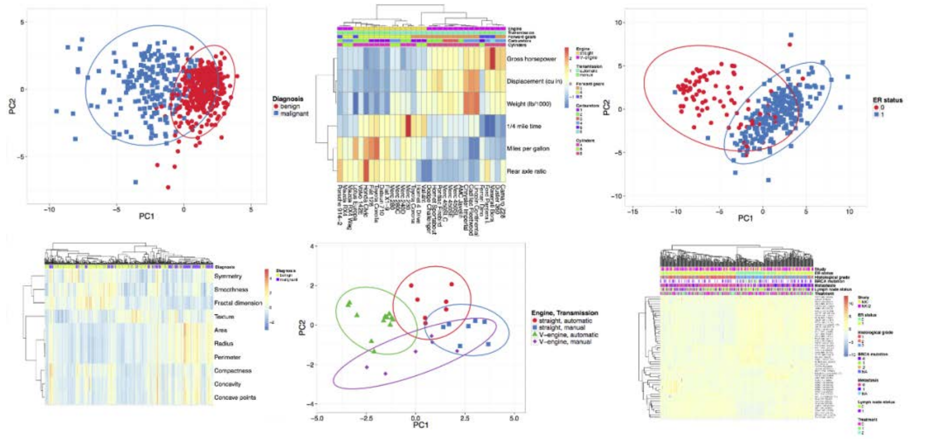

Fig. 3 The spatial position of the three hierarchical nodes in the experimental group and the control group are consistent.

4.3.2 Experimental data processing and analysis

This group of experiments was divided into experimental group and control group, each group was arranged to participate in the name and not repeated participants. According to the task arrangement, each participant completed three search tasks, and analyzed the operation performance of the two groups by evaluating the eye track, the number of viewpoints, and the operation time. Thus, the effects of the experimental group and the control group on the visual search process were compared [9]. According to the comparison chart of the total number of fixation points, it can be seen that the visual search efficiency of the experimental group is higher, and the attention level of the subjects is better. Taking the A-level point search of task one as an example, there is a significant difference in the number of fixation points between the experimental group and the control group in the range of the target node. Among them, the experimental group is divided into layers according to the information of each page, so the number of nodes in the page is reduced, so the visual target of the subject on the current page is clearer, and the search difficulty is reduced.

5. Conclusion

It can be seen that the control group shows that the number of viewpoints around the target node is large, but they are all in a glance state, that is, multiple scans pass through the target node but are not successfully located. Therefore, the multi-user co-planar information node page interferes greatly with the visual search activity of the subject, resulting in the difficulty of node search: GB>A, so the multi-user node is co-planar, due to the increased interference, C-level The search location of the target node is more difficult. In the above figure, (1) in the search positioning of the same hierarchical node, the number of viewpoints of the experimental group in the area where the target node is located is more than that of the control group; (2) The target nodes around the target nodes of the experimental group A, B, and C The number is more average than the control group, indicating that the accumulation of visual cognitive load is eased due to the hierarchical division of nodes.

References

[1] Song Xueqing, Liu Yu. Big Data: A Transformation of Information Technology and Information Management. Information Science, Vol. 9 (2014) No.28, p. 148-151.

[2] Fu Yan, Zhang Jianxun. Reflections on the Deep Integration of Information Technology and Higher Education in the Big Data Era. China Light Industry Education, Vol. 4 (2014) No.15, p. 9-11.

[3] The New Path of Personal Information Protection under the Background of Big Data——the Renewal of Forgotten Right and the Innovation of Information Technology. Library and Information Science, Vol. 6(2016) No.18, p. 107-113.

[4] Hou Xiaoling. Thoughts on the Expansion of Information Technology Work in University Libraries from the Perspective of Big Data. Curriculum Education Research, Vol. 20 (2016) No.22, p. 132-136.

[5] YANG Zhen, ZHONG Cheng, HE Qiuyan. Research on Personalized Teaching of University Information Technology Course Based on Big Data. New campus (first), Vol. 9(2015) No.17, p. 45-48.

[6] Wang Bingli. Visualization Analysis of International Big Data Research Based on Scientometrics. Journal of Information, Vol. 2(2015) No.13, p. 131-136.

[7] Su Nan. Visualization Analysis of Big Data Research in China Based on CSSCI. Information Science, Vol. 3(2016) No.19, p. 112-116.

[8] Zhou Lin, Kong Lei, Zhao Fangqing. Current Status and Challenges of Bio-Big Data Visualization. Chinese Science Bulletin, Vol. 4(2015) No.21, p. 547-557.

[9] He Qun, Yang Mingchuan. Research and Optimization of Big Data Visualization Based on WebGS. Telecommunications Technology, Vol. 6(2015) No.38, p. 37-40.