CSCanada
Management Science and Engineering
Vol. 12, No. 2, 2018, pp. 47-55
DOI:10.3968/10594

ISSN 1913-0341 [Print]
ISSN 1913-035X [Online]
www.cscanada.net
www.cscanada.org

Research on Big Data Collection Mode of Ancient Village Culture Based on Network Crowdsourcing

XIE Zhengxia[a]; LIU Siqi[a],*

[a]School of Business Administration, South China University of Technology, Guangzhou, China
*Corresponding author.

Received 13 March 2018; accepted 20 May 2018
Published online 26 June 2018

Abstract

In the long history of Chinese civilization, the village is the most basic social unit. However, in the wave of urbanization, the village culture is accelerating. Therefore, in the process of modernization, how to protect the village culture is an important issue that cannot be ignored. At the same time, with the advancement of the times, the introduction of modern information technology into the work of folk culture has become an important means of protecting traditional cultural heritage, but the historical and cultural materials of ancient villages are huge in number and form, which brings certain difficulties to the collection and collation of historical and cultural materials in villages. In order to solve the problems of large amount of data and difficulty of collection in ancient village digital culture protection, this paper combines ancient village cultural protection with inheritance and modern information technology, and proposes a big data collection mode based on crowdsourcing and supplemented by cloud service platform. Firstly, aiming at the status quo of cultural protection of ancient villages, we analyze the difficulties in data collection of ancient villages, and explain the necessity of combining cultural protection of ancient villages with network crowdsourcing. Then we design a network crowdsourcing mode and the ancient village cultural big data cloud service platform. At last, through the pilot verification, it verifies that the data collected by the crowdsourcing mode proposed in this paper has the advantages of high user participation, low cost, short time-consuming and quality control. This study effectively solves the problem of time-consuming and costly big data collection, and provides new feasible ideas for the digital inheritance and protection of ancient village culture. At the same time, the cloud platform can provide the deep excavation of ancient village culture, content indexing, literature compilation and publication, and thematic knowledge retrieval and so on. And this platform can serve as a basic tool for future cultural protection, heritage, research and foreign exchange in ancient villages.

Key words: China ancient villages; Crowdsourcing; Big data collection; Cloud service platform

Xie, Z. X., & Liu, S. Q. (2018). Research on Big Data Collection Mode of Ancient Village Culture Based on Network Crowdsourcing. *Management Science and Engineering*, *12*(2), 47-55. Available from: URL: http://www.cscanada.net/index.php/mse/article/view/10594
DOI: http://dx.doi.org/10.3968/10594

INTRODUCTION

The ancient village refers to that village built before the Republic of China, and it still retains historical evolution until now, which is means that its building environment, architectural style and location has not changed greatly, it remains its folkways. And although the time has passed for a long time, people live in the village yet. Due to its long history, ancient Chinese villages have gradually formed a colorful village culture in the process of development. However, in the process of modernization, the village culture is accelerating. How to effectively protect the village culture in the era of urbanization has become an important issue that needs to be solved urgently.

The rapid development of information technology not only brings people a new way of life and thinking, but also provides a new way for the protection of traditional cultural heritage. Among them, the collection and preservation of ancient village culture in a digital way is an effective way. In particular, the development of big data technology has solved the problems of storage and

use of data form diversification in ancient villages, so it provides a powerful support for preservation of ancient village culture.

However, in the digital preservation activities of ancient village culture, collecting its data is still difficulties. As a special historical and cultural carrier, ancient villages have geographically dispersed, large amount of data, complex data sources, and data forms including material culture and intangible cultural heritage data, which makes it difficult to collect. For example, if we want to collect the data of the 205 ancient villages in Guangdong Province of China in a traditional way, it will need a professional team and will take two or three years. It will spend a lot. Therefore, how to collect data quickly and effectively has become a key problem in the digital protection of ancient village culture.

This paper combines the cultural protection and inheritance of ancient villages with modern information technology, and proposes a big data mode of ancient village culture based on network crowdsourcing and cloud service platform. With the participation of the network public, we strive to make the data collection of ancient village become more efficient and less expensive. Finally, we verify the feasibility of the acquisition mode through pilot.

1. LITERATURE REVIEW

1.1 Crowdsourcing Definition

Jeff Howe (2006) first proposed the concept of crowdsourcing, that is, "crowdsourcing provides a new cheap labor resource that enables people to use their leisure time outside of life to help companies solve problems and even help companies develop". Then its concept was continuously improved, and the rise of this mode also made foreign scholars pay attention to crowdsourcing. However, at present, research on crowdsourcing in China is relatively lagging. Ni (2010) defines crowdsourcing from the perspective of enterprises: enterprises rely on their own websites to announce some work or tasks to a large number of external audiences, and finally pay the corresponding remuneration to the self-recommended person who completes the tasks as required, which is a mass contracting mode.

Combined with the mainstream view of academia, this paper believes that crowdsourcing refers to the task which is outsourced to non-specific online public gathered from the Internet (such as a large-scale mass network or virtual community) based on the principle of free resources and publicity. In the process of completing the task, the contractor exerts subjective initiative and creativity to ensure the completion of the task, so as to obtain the corresponding compensation provided by the contractor.

1.2 Big Data and Cloud Computing

The issue of Science magazine published in 2008 defines the big data as "representing the progress of human cognitive processes. The scale of data sets cannot be acquired, managed and processed with current technologies, methods and theories within a tolerable time". Cloud computing is a computing mode that uses the Internet to access the shared resource pools conveniently at anytime, anywhere when you are demand (Luo, Zhou, &Song et al, 2011). And this technology is inseparable from big data. With the rapid development of software-as-a-service, platform-as-a-service, infrastructure-as-a-service and all-as-a-service services advocated in the field of cloud computing, big data analysis has received more and more attention from the society, and analysis-as-a-service has gradually developed and formed (Guan, Meng, &Li, Liu,2015).

Wang and Liu (2013) believe that the solution of big data system will definitely land on the existing cloud computing platform. The distributed file system, distributed computing mode and distributed database management technology of cloud computing platform provide a new idea to solve the problems of big data. Although the development of cloud computing is still not mature enough at present, cloud computing inevitably has various problems in the process of integration with industries, but in general it is still feasible, and the advantages outweigh the disadvantages. Therefore, the research on the integration of cloud computing and specific industries has great theoretical and practical significance (Su, 2012), and the research on the integration of cloud computing and ancient village cloud service platform is also of great significance.

2. INTRODUCTION OF ANCIENT VILLAGE CULTURE BIG DATA COLLECTION

2.1 Content of Big Data Collection

The collection of big data of ancient village culture mainly refers to the aspects of the ancient village culture covered by modern technology through on-the-spot investigation, and recorded as much as possible by pictures, text, audio and video. Due to the collection workload, data scale and data complexity of ancient village cultural data, it is far greater than the conventional scale. At the same time, due to the geographical dispersion of ancient villages, and most of them are far from the city, the information infrastructure is backward, so the data information acquisition and update work is difficult.

The big data preprocessing work mainly focuses on the digital processing of traditional text, pictures, audio and video resources. If the traditional method is used for digital processing, the workload is huge and time consuming. To overcome this problem, this paper considers the network crowdsourcing to collect and update data relatively and efficiently.

2.2 Difficulties in Big Data Collection

2.2.1 The Historical and Cultural Materials of Ancient Villages are Numerous and Diverse

The historical and cultural materials of ancient villages cover all aspects of ancient villages and cover a large number of historical, human and socio-economic information. From the content point of view, including material cultural heritage such as village planning, various buildings, historical sites, etc.; also includes various intangible cultural heritages such as folk customs and national languages.

The data presents large capacity, multiple sources, multiple formats, and multiple types, which are typical features of big data. Therefore, in data collection, it is more difficult to obtain complete and comprehensive historical and cultural materials of ancient villages. In terms of data storage and post-use, it is difficult to achieve the requirements of sharing and inheritance by traditional means.

2.2.2 Historical and Cultural Materials of Ancient Villages Scattered Everywhere and in Various Forms

Due to China's vast territory, based on various factors such as history and geography, the villages in different regions have different forms. Each region and each village has its own special cultural form. These characteristic settlements provide a broad field for humanities and social science research, and also bring difficulties to the collection and collation of village materials.

2.2.3 Data Collection and Pre-Processing Work Needs To Be Sustained For a Long Time

The ancient village is not a cultural relic, it is an ancient community where people still live. The ancient village culture is not only a traditional heritage that has been left behind, but also a special cultural community that is being carried out and developed. The current and future developments mean that the digital work of ancient villages is bound to be a process of long-term accumulation rather than short-term arrival. Therefore, in data processing, an extensible architecture must be built to leave enough room for future expansion.

2.2.4 The Ancient Village Culture is Accelerating With the Village

In the tide of modern society, along with the rapid development of industrialization and urbanization, the village cultures of different styles are experiencing acceleration one by one. Regardless of the material culture level or the non-material cultural level, the inheritance of ancient villages has become increasingly thin, and the status quo is not optimistic.

2.2.5 The Pretreatment of Ancient Village Culture Big Data is Difficult

Digital processing of traditional text, pictures, audio, video and other resources not only requires a lot of labor and time costs, but also needs different processing methods to solve the non-standardization problem of different sources of data.

In summary, the ancient village cultural big data collection and pre-processing work is faced with such problems as large and complex data, scattered and varied forms, accelerated disappearance of ancient village culture, and long time required for collection, if only relying on experts in the field to collect, sort and excavate, the task is very arduous. It will cost a lot of labor and time, and does not meet the principle of benefit. And network crowdsourcing just solves these problems perfectly.

3. DESIGN OF NETWORK CROWSOURCING MODE FOR BIG DATA COLLECTION OF ANCIENT VILLAGE CULTURE

3.1 Crowdsourcing Mode Overall Framework Design

The crowdsourcing mode designed in this paper adopts the combination of the participation of the network public and the subject experts, and belongs to the mode of directed crowdsourcing. For the large number of ordinary tasks with moderate difficulty will be completed by the network public, so as to realize large-scale on-site investigation and general data collection. For the difficult tasks involving expertise are directed by domain experts, such as building components. At the core of the entire crowdsourcing operations are data management, user management, quality management, and payment and feedback management. The platform mode frame diagram is shown below (Figure 1).

3.2 Crowdsourcing Mode Core Process Design

This article divides the main players into three categories, namely general network users, domain experts and managers. In addition, the design of this stage integrates the three stages of existing data collection, field investigation, and post-data compilation. Therefore, the core processes include the ancient village culture big data collection and pre-processing process, the general network user participation process, and the domain expert collaboration process. Among them, the data collection and pre-processing process includes the identification, distribution, recycling, sorting and review of the ancient village culture big data collection task package; the general network user participation process describes how ordinary users participate in tasks, complete tasks, get rewards; domain expert collaboration process describe how domain experts can participate in specific crowdsourcing tasks or collaborate to complete crowdsourcing.

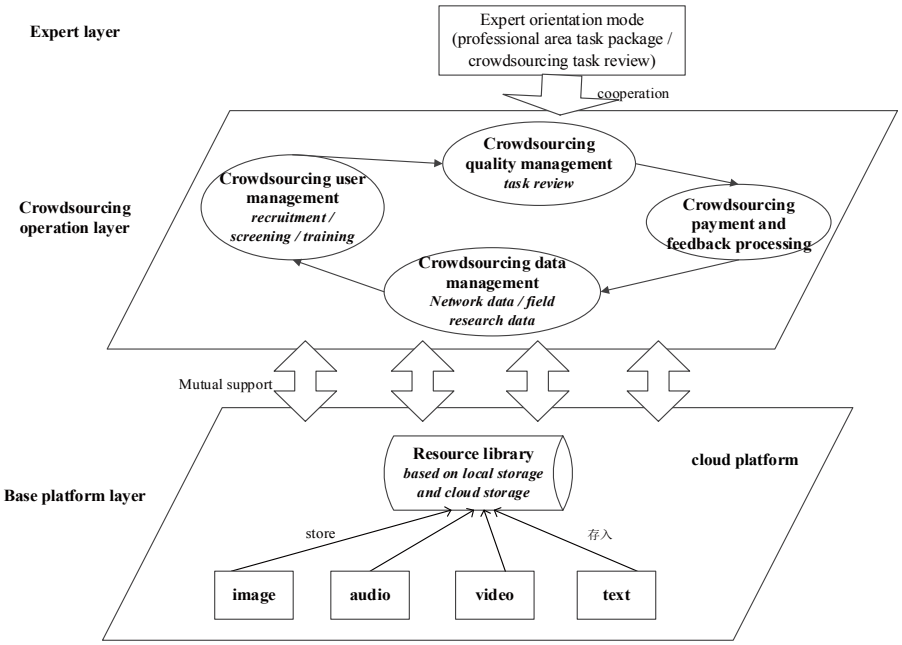

Figure 1
Overall Framework for Directed Crowdsourcing Mode

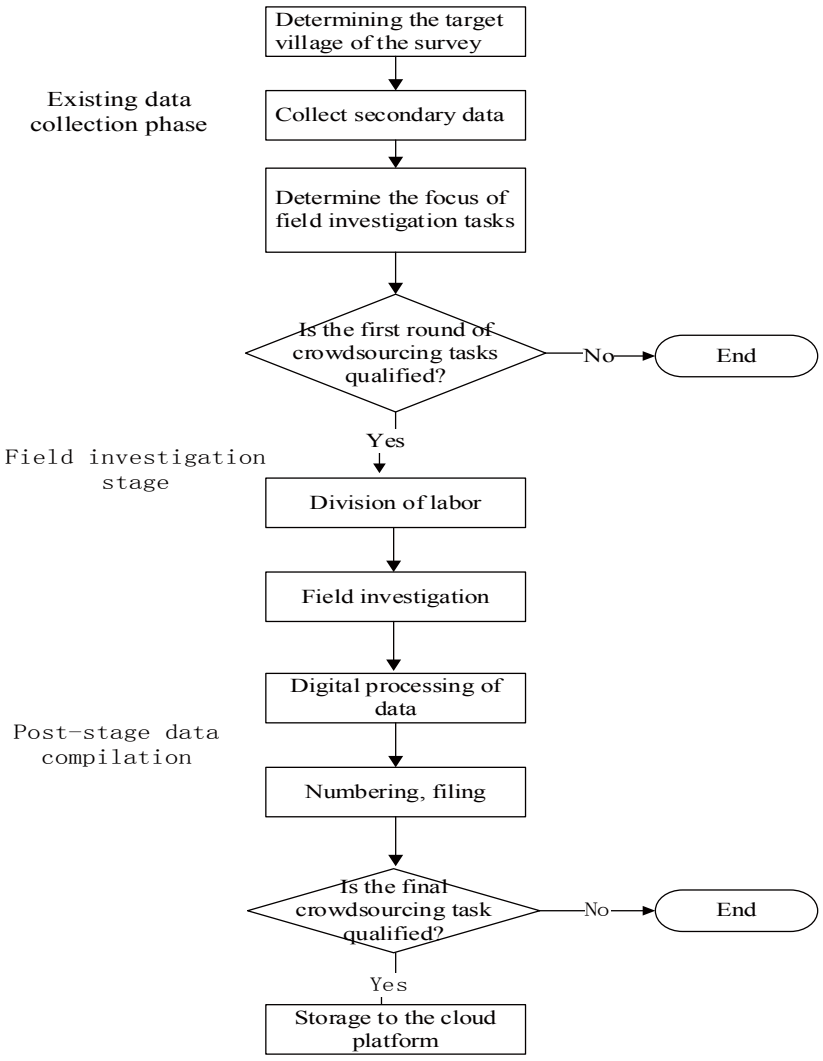

Figure 2
Ancient Village Culture Big Data Collection and Preprocessing Process

3.3 Crowdsourcing Organization Mode Design

Crowdsourcing organization mode design should be based on the overall framework of the crowdsourcing mode and user-centric, to maximize user engagement. Therefore, based on the crowdsourcing platform, this paper is different from the traditional organizational form of crowdsourcing, and builds an online community to enable real-time two-way communication between staff, crowdsourcing users and non-crowdsourcing users. It is not a place to release tasks, but a question exchange and topic discussion place for crowdsourcing tasks, which would make crowdsourcing users, and non-crowd users are more interested in participating tasks.

4. DESIGN OF BIG DATA COLLECTION CLOUD SERVICE PLATFORM FOR ANCIENT VILLAGE CULTURE FOR NETWORK CROWDSOURCING

4.1 Software Architecture Overall Architecture Design

The cloud platform design architecture adopts a combined hybrid mode of B/S and C/S structure, while the software architecture hierarchy uses the most common three-layer structure, from top to bottom: presentation layer (UI), Business logic layer (BLL), data access layer (DAL) (Figure 3).

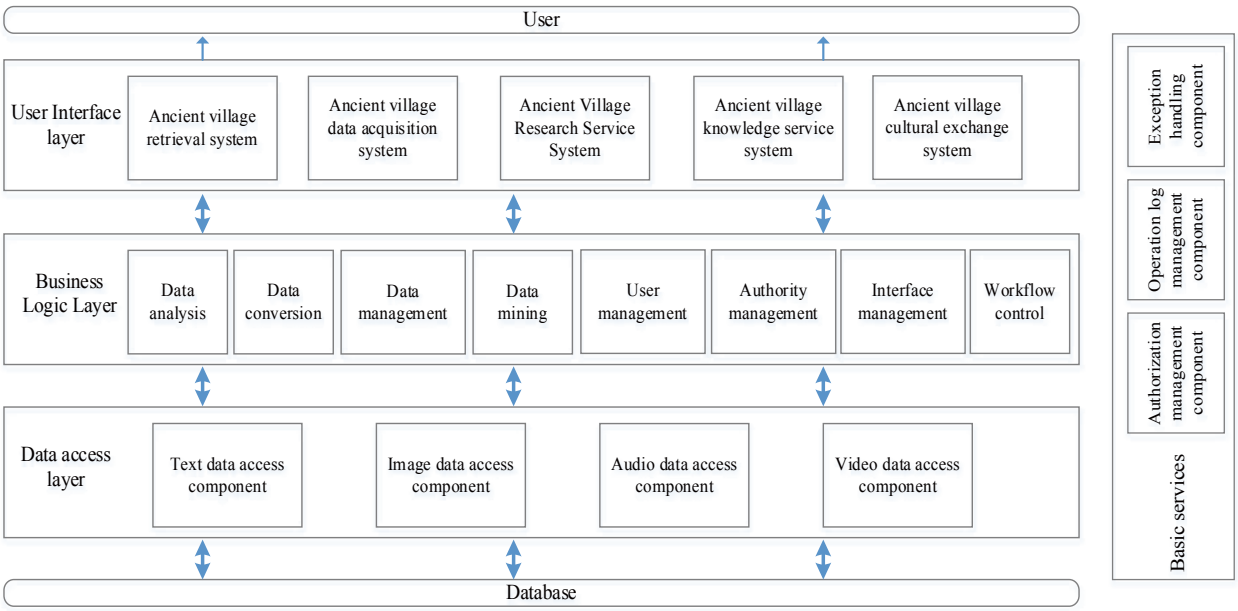

Figure 3
Software System Three-tier Architecture Design

4.2 Cloud Service Platform Design

In the cloud service platform of big data of the ancient village (Figure 4):

The software-as-a-service (SaaS) layer mainly implements functions such as the ancient village retrieval system, data collection system, scientific research service system, knowledge service system, and cultural communication system, and can be collected on WAP, APP, and WEB for users to use.

Platform-as-a-service (PaaS) layer, which mainly implements access control, workflow engine, and decision support functions.

The Infrastructure as a Service (IaaS) layer is the most basic part, including resource scheduling, virtualization/ resource pooling, and cloud infrastructure construction. In other words, it would set up the virtual servers by infrastructure such as servers, storage, and networks, and create a flexible resource pool to achieve automatic allocation of computing resources, storage resources and network resources.

4.3 Cloud Terminal Functional Architecture Design

The main functions of the ancient village cultural protection and inheritance computer client are shown in Figure 5. Basically, all functions such as user information management, search and query, scientific research service, knowledge service, cultural exchange service, and background management can be realized.

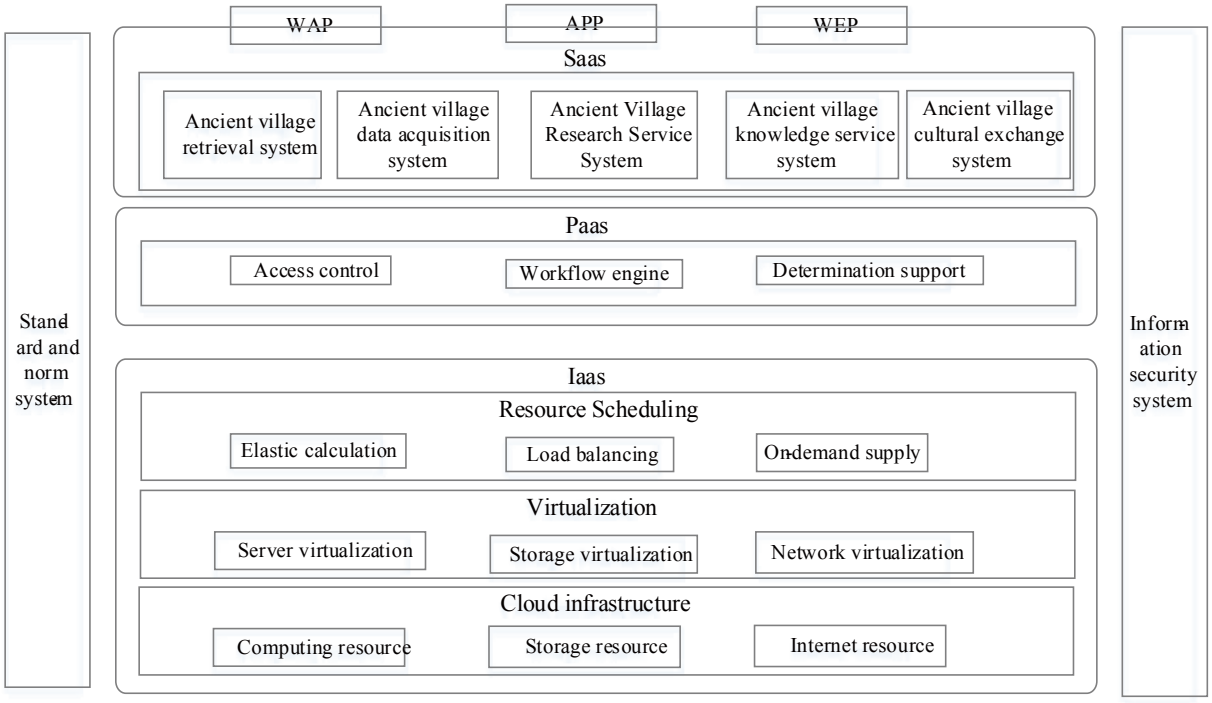

Figure 4
Three-Level Service of the Ancient Village Cultural Big Data Cloud Service Platform

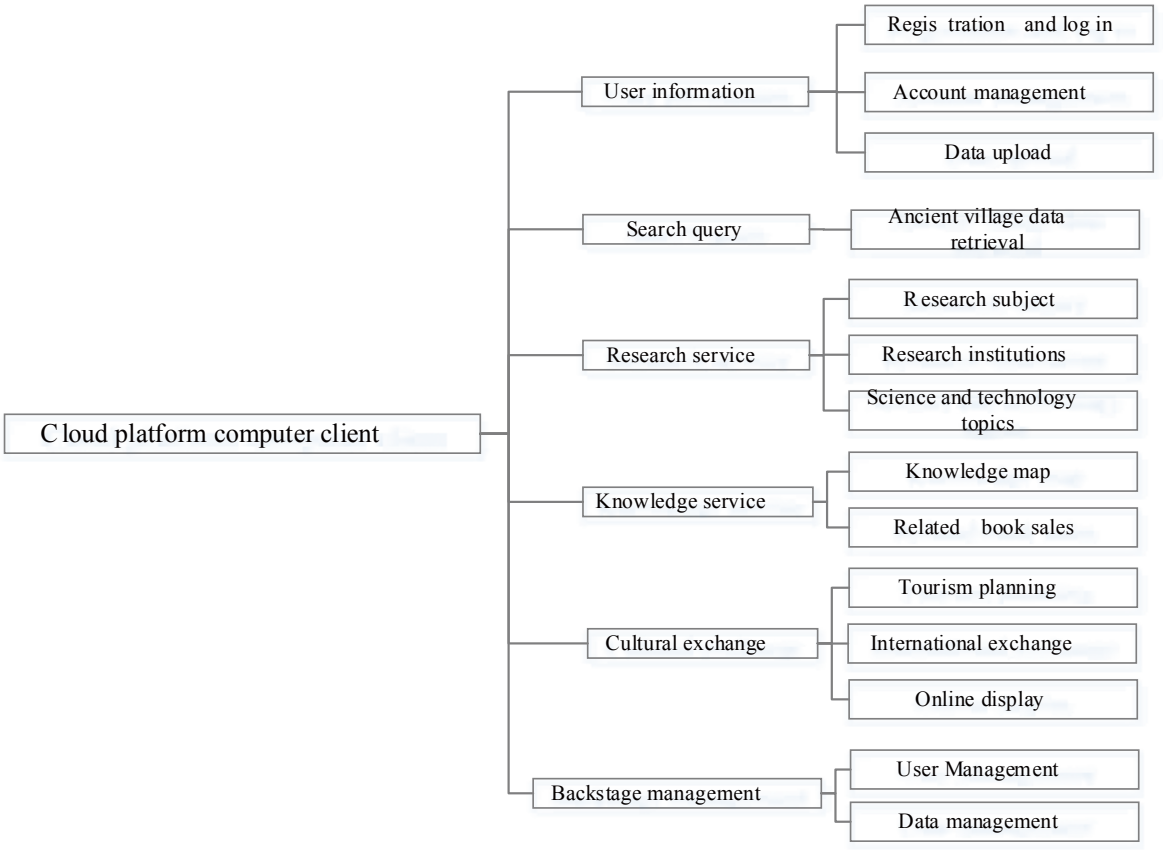

Figure 5
Functional Design of Cloud Platform Computer Client

5. EXPERIMENT AND EFFECT EVALUATION OF CROWSOURCING MODE OF ANCIENT VILLAGE CULTURE BIG DATA COLLECTION NETWORK

5.1 Summary of Pilot Facts

The purpose of this pilot is to verify the feasibility of the crowdsourcing mode, requiring rapid response.

In the implementation process, considering the regional balance problem, we choose 16 ancient villages in the sub-region from the 205 ancient villages preserved in Guangdong Province. The experiment consists of two phases, namely the collection of existing data and the field investigation to evaluate the effect of the crowdsourcing mode designed. Among them, the field survey refers to on-site investigation through interviews, written records, and shooting methods.

According to the crowdsourcing workflow and content mentioned above, the main tasks of this test include: task packaging, data collection work inquiry, task information release, user information review and screening, participant training of crowdsourcing and agreement signing, task delivery, quality audit and remuneration payment.

5.2 Experiment Completion

5.2.1 User Registration and Task Collection

For the first batch of 16 pilot villages, during the 5-day registration period, 48 people were collected from WeChat and email.

5.2.2 The Completion of the Existing Data Collection Phase

A total of 16 ancient villages in Guangdong were opened in the pilot, with a total of 15 participants. In the first phase, all 16 villages completed the collection of existing data, with a completion rate of 100%. Among them, 10 villages have excellent completion ratings, and 4 villages have a good completion rating and can directly enter the field investigation stage; two villages have achieved good ratings after a modification due to repeated data duplication and inaccurate filing, and also enter the field investigation stage.

5.2.3 Field Investigation Phase Completion

Among the 16 pilot ancient villages, 15 villages completed the data collection of field investigations, with a completion rate of 93.75%. Among them, the field survey of 11 ancient villages has been completed as excellent. The field surveys of two ancient villages have been completed with good ratings and can be directly accessed into the payment process. In the two villages, there was a problem that the collected data was not comprehensive enough, but the comprehensive difficulty was considered, so the score was good too.

6. DISCUSSION

Through the crowdsourcing pilot of big data collection in 16 ancient villages, the network crowdsourcing mode designed in this paper is verified, which has practical feasibility. In the following, from the four aspects of user participation, time-consuming, cost and quality, the evaluation of the network crowdsourcing mode of the ancient village cultural big data collection is carried out.

6.1 High User Engagement in the Network Crowdsourcing Mode

During the pilot work, we received 48 people's registration information in just five days. Among the 48 applicants, 11 are from urban planning, architecture, geography and other majors, and have experience in cultural relics surveys and field investigations. Night people are photography and travel enthusiasts. Seventeen people have had relevant work experience in data collation. Three people signed up for the mission of protecting the traditional culture of their hometown. In general, the professional foundation and enthusiasm of the applicants are relatively high. It is foreseeable that if the recruitment time is extended, the number of participants will be more and the staff can choose a wider range.

6.2 Network Crowdsourcing Mode Takes Less Time to Work

One of the great advantages of online crowdsourcing is that subcontracting each village to different people can carry out related work at the same time. The result of the crowdsourcing pilot was that 15 crowdsourcing users completed the data collection of the villages they were responsible for within 10 days of the specified time, which was significantly more efficient than outsourcing to a specific team.

The data collection work of 205 ancient villages in Guangdong is time-consuming. When it is not based on crowdsourcing, it takes at least 205 (village) x 4 (day/village) = 820 days, that is, 2 years and 3 months. If the data collection work of 205 villages is carried out based on network crowdsourcing, it is expected that the data collection of all villages can be completed within 2 months, and the effect is obvious.

6.3 The Network Crowdsourcing Mode Cost Lowly

According to the previous inquiry, in the case of not based on crowdsourcing, the cost of collecting data for each ancient village in a village-by-village configuration is at least 3,000 yuan per village. And if it is outsourced to a professional team, the cost of a single ancient village will be higher. In this crowdsourcing pilot, each village was assigned to a participant, not only the quality is generally good and excellent, but the payment cost of each ancient village is only 1,200 yuan / village, lower than the salary of the same type of part-time work. So the advantages of large-scale implementation are obvious. The comparison of the cost of data collection in ancient villages is shown in Figure 6.

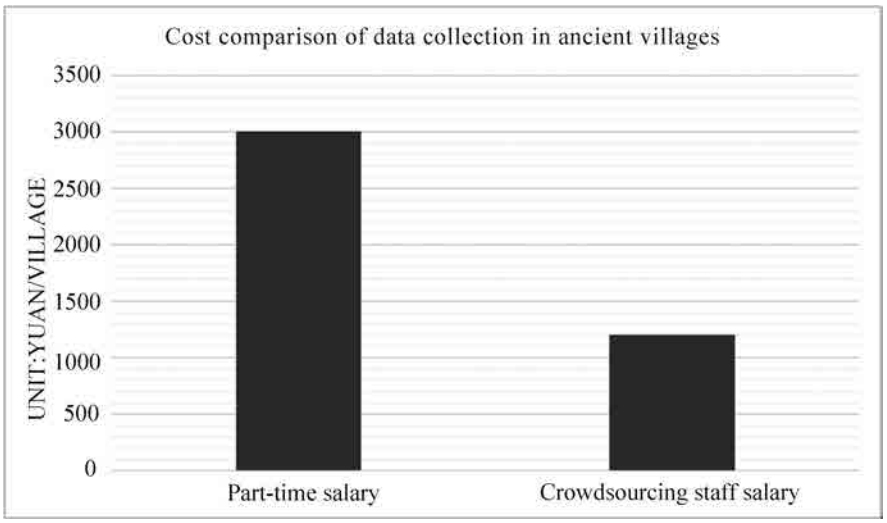

Figure 6
The Comparison of the Cost of Data Collection in Ancient Villages

6.4 The Quality of Work in the Network Crowdsourcing Mode can be Guaranteed

In this crowdsourcing pilot work, the quality rating of the two stages was obtained through the quality audit of the crowdsourced users after the data collection phase and the field investigation phase. Among them, in the existing data collection stage, all 16 villages were completed and both reached the pass level, so the pass rate was 100%. The excellent rate reached 62.5%, and the good rate reached 37.5%. In the field investigation stage, the completion rate of 16 villages was 93.75%, and all of them were at the qualified level. Among them, the excellent rate reached 68.75%, and the good rate reached 25.0% (Figure. 7).

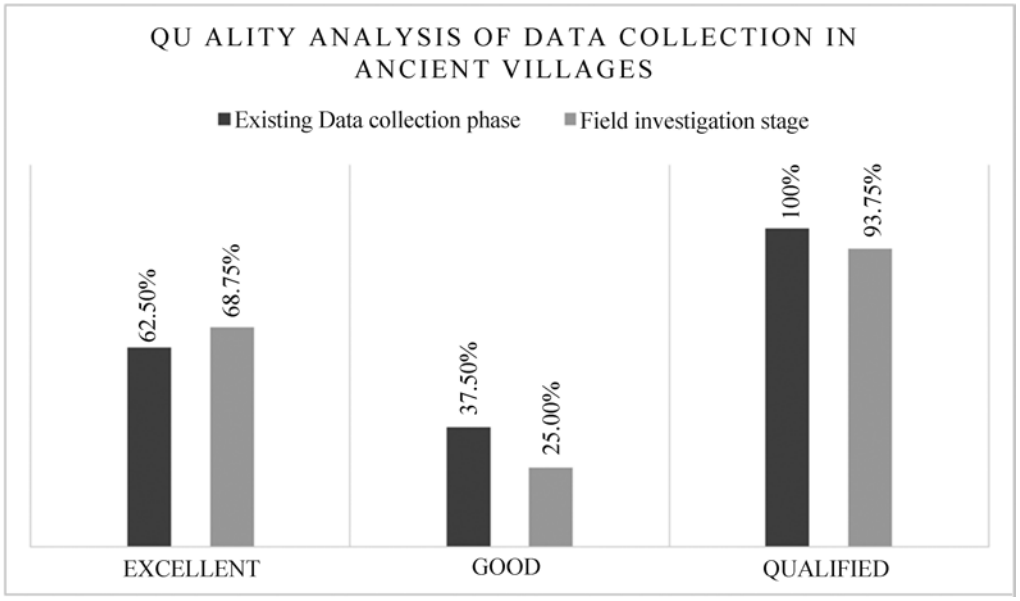

Figure 7
Quality Analysis of Data Collection in Ancient Villages

At the same time, through comparison with the quality of the pre-research of the author's three-person team, it is found that the quality of the completion of the crowdsourcing of the entire pilot is comparable, and the quality of the completion of some students such as architecture and urban planning is not only significantly higher, but also recognized by a professional teacher. Therefore, the crowdsourcing mode can make full use of the professionalism of its users, mobilize the user's enthusiasm and subjective initiative, thus ensuring the quality of project completion.

In summary, the network crowdsourcing mode is feasible in the overall operation. At the same time, it is efficient and low-cost for the data collection of ancient villages. The user participation and enthusiasm are high, and the work quality can be guaranteed too. It can be applied to Guangdong in the future, even in the large data collection work of ancient Chinese villages.

CONCLUSIONS

This paper innovatively introduces network crowdsourcing and cloud platform into the ancient village culture big data collection work, and designs the network crowdsourcing mode and the ancient village culture big data collection cloud service platform for network crowdsourcing, so as to harness the power of the network to solve the problems of large number and variety of big data collection work in ancient villages. Through the pilot project of 16 ancient village data collection, the practical feasibility of the designed network crowdsourcing mode is verified. The network crowdsourcing mode designed in this paper needs to rely on the cloud platform as the basic support. However, the design of the cloud platform in this paper only describes the construction ideas and prototypes, and does not realize the system construction. It needs to be further improved in future research.

REFERENCES

Graham-Rowe, D., Goldston, D., Doctorow, C., et al. (2008) Big data: science in the petabyte era. *Nature*, *455* (7209), 8-9.

Guan, S. F., Meng, X., Li, Z. J., & Liu, Y. (2015). The status quo, problems and countermeasures of big data analysis research. *Journal of Intelligence,34*(05),98-104.

Jeff, H. (2011). *Crowdsourcing: Group power drives the business future.* Beijing: CITIC.

Luo, J. Z., Jin, H. J., Song, A. B., & Dong, F. (2011). Cloud Computing: Architecture and Key Technologies. *Journal of Communications*, *32*(7), 3-21.

Ni, N. (2010). "Crowdsourcing"- A new model for enterprise HR management with external force. *New capital*, 38-40.

Su, K. (2018). *Research on e-government infrastructure under cloud computing platform* (Master's thesis). Shandong Normal University, Jinan.

Wang, X. L., & Liu, P. (2013). Big data key technology. *ZTE Technology,* (04), 17-21.