Advances in Social Science, Education and Humanities Research, volume 610
Proceedings of the 7th Annual International Conference on Social Science and Contemporary Humanity Development (SSCHD 2021)

Study on the Construction of Digital Humanities Group Portraits

Tingting Duan[1], Ping Qin[2,*] and Ping Jiang[2]

[1] *College of Humanities and Social Sciences, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China*
[2]*Library of Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China*
** Corresponding author. Email: qplib@nuaa.edu.cn*

ABSTRACT
Digital humanities research has received great attention from scholars in recent years, and its research field has been expanding with the advancement of technology, and the team of digital humanities scholars has gradually grown. In this paper, through the study of user portrait construction, we deeply explore the group characteristics and academic behaviors of China's digital humanities researchers in the context of disciplinary crossover. Through the construction of scholars' portraits, the research fields, institutional cooperation networks and scholars' relationship networks of scholars are explored, and the group portrait of China's digital humanities research scholars is presented. The study points out that the development of China's digital humanities field requires cultivating a young generation of scholars, strengthening inter-institutional connections and cooperation, and focusing on the development, application and training of tools and technologies.

Keywords: *User portrait, Digital humanities, CiteSpace, Group characteristics.*

1. INTRODUCTION

With the development of big data, cloud computing and artificial intelligence, digital humanities, an emerging field formed by the cross-fertilization of digital technology and the humanities, has shown an accelerating trend of integration, and digital humanities research has attracted the attention of scholars, and there is an urgent need to strengthen the construction of this field in China. According to statistics from scholars in the Sci-Fund Database, 229 digital humanities projects have been established in China in the past ten years[1], and digital humanities has become a hot spot of concern for scholars. As an emerging field, how to identify the hotspots and problems of digital humanities research through data mining and provide relevant suggestions for the construction of the digital humanities field is an urgent problem to be solved.The construction of user portraits based on data collection provides a new direction for solving this problem.

Digital humanities originate from humanities computing, which refers to the application of digital technology and digital methods and tools to the study of the humanities. Digital humanities research usually requires the collaborative efforts of scholars from different disciplinary fields, and has already been joined by scholars from the fields of history, literature, linguistics, archaeology, library and intelligence studies, and computer science. Research topics include construction of special databases, digitization of resources, digital publishing, resource mining, etc. For example, the Shanghai Library relies on the number of its own libraries and genealogical resources, and has constructed a Chinese genealogical knowledge service platform (http://jiapu.library.sh.cn) by meticulously refining knowledge units about genealogy through semantic technology [2]. The Dunhuang Academy, in conjunction with the State Key Laboratory of Remote Sensing Information Engineering, Wuhan University, the Information Resources Research Centre and the School of History, formed a project team to investigate high-precision digital acquisition techniques for Dunhuang murals, mural knowledge mapping construction techniques and the evolutionary history of mural styles [3]. Ai Yuxi et al. used canonical text data and related canonical dictionaries as research objects, combined with conceptual-semantic information to organize and describe features of canonical content data, and constructed a classification system on click content analysis [4]. The above research provides digital humanities scholars and ordinary users with knowledge services related to the field of digital humanities to varying degrees.

User portrait refers to a data processing method for personalised services that has emerged in the context of big data. By collecting a large amount of relevant data from target users and analysing it, a structured representation of user characteristics and preferences is made using a labelling approach [5]. The concept of user portrait was first introduced by Alan Cooper, the "father of interaction design", and has been rapidly developed and widely used in various fields, such as tourism to accurately profile tourists to achieve personalised recommendations for tourists, retail to build customer profiles and take decisions such as changing sales strategies to achieve a win-win situation for both customers and merchants.In scientific research, scholars build a portrait model of a discipline research group to depict the academic portrait of scholars, so as to realize the accurate recommendation of scholars' academic resources. Li Bao constructed a multi-dimensional portrait model of users from four dimensions: basic user attributes, reading readiness, learning style, and reading interest preference, and proposed a personalized recommendation service strategy based on co9 ld-start user portraits and Yudu process user portraits [6]. Cheng Dongling used tags to paint an accurate "user portrait" of the poor target, compared research data to more accurately filter out effective data, and then conducted a more complete data analysis to establish a foundation for precise poverty

alleviation and consolidation of poverty alleviation effects [7]. Guo Jianlong used a cluster analysis method to process and handle the labels using a deployment environment platform, and eventually used actual cases to build a talent portrait of enterprises in the manufacturing sector, providing a basis for enterprises in the manufacturing sector to introduce outstanding talents in their field [8]. Digital humanities research has received widespread attention from various disciplines, and the number of both native digital humanities scholars and interdisciplinary digital humanities scholars involved in research across multiple disciplines will continue to grow. How to extract the characteristics of the group of digital humanities researchers has rarely been studied. In this paper, we propose to build a portrait of digital humanities scholars by drawing on the ideas and methods of constructing user portraits, and put forward relevant suggestions and countermeasures for the construction of the digital humanities field.

2. RESEARCH IDEAS AND RESOURCES

Based on the portrait construction technique, the four dimensions of scholars' physiological attributes, social attributes, published papers and involved projects are analyzed from two aspects: basic information and academic information, as detailed in Figure 1. Scholars' physiological attributes, social attributes and involved projects are presented through statistical charts, while academic attributes are analyzed through CiteSpace's high-frequency word statistics table, keyword co-occurrence network mapping, and The author and institution collaboration network graphs are presented. By analysing the frequency and mediated centrality of the nodes in the graphs, the research content preferences of the research subjects in the academic field can be found.

For the data sources, domestic humanities scholars were sourced from the CNKI database and the top 40 scholars in the field of digital humanities were selected for the number of articles published. The scholars' information was obtained through the official websites of their units, library websites, scholars' websites, academic results, CNKI database and project information database, etc, as shown in table Fig. 1.

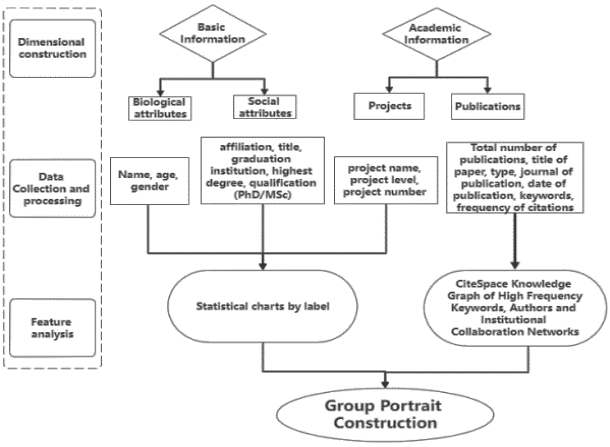


Figure 1 Research idea map for building a portrait of scholars in the digital humanities

2.1. Design system of Portrait Labels

The portrait of digital humanities scholars is constructed from two aspects: basic information and academic information, as shown in Table 1; basic information includes physical attributes and social attributes; physical attributes include labels of name, age and gender; social attributes include labels of affiliation, title, graduation institution, highest degree and qualification (doctoral/master's degree); academic information is mainly developed from two dimensions: information of published papers and information of participating projects. The information of published papers includes the tags of total number of papers, title, type, journal, publication time, keywords and frequency of citations, and the information of participating projects includes the tags of project name, project level and project number.

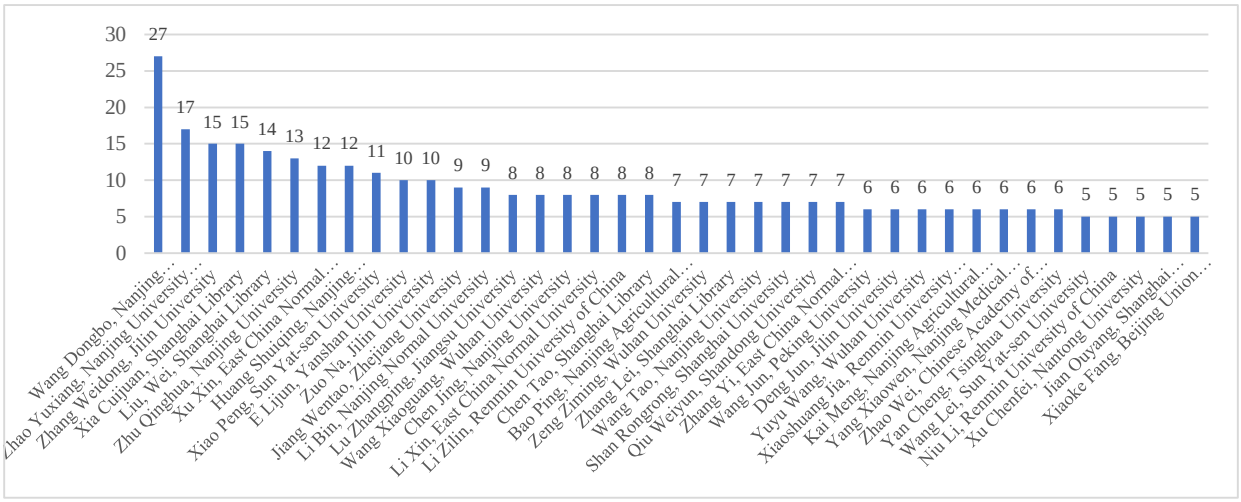


Figure 2 Top 40 scholars in digital humanities publishing

Table 1. Digital humanities scholar portrait tagging system

Attributes of Digital Humanities Scholar Portrait	TAGS	Data sources
Basic Information	**Biological attributes:** Name, age, gender **Social attributes:** affiliation, title, graduation institution, highest degree, qualification (PhD/MSc)	Official websites of affiliated units, library websites, scholars' networks, academic results
Academic Information	**Publications:** Total number of publications, title of paper, type, journal of publication, date of publication, keywords, frequency of citations **Projects:** project name, project level, project number	Official websites of affiliated institutions, CNKI database, academic results, project information base, etc.

2.2. Identification of Digital Humanities Scholars

The first problem to be solved is how to identify digital humanities research scholars from the large number of interdisciplinary scholars. Since only a few schools have set up digital humanities majors, and most digital humanities scholars currently belong to different departments and disciplines, it is difficult to identify digital humanities researchers through departmental and disciplinary divisions. So a search was conducted in the CNKI database using the subject term "digital humanities". After analyzing the distribution of Chinese authors through the result visualization function of CNKI, the top 40 scholars in terms of literature volume were selected as the target of the study, and since two of them with the same name in different units, the two pieces of information were combined to obtain information on 39 scholars. as shown in Fig.2.

3. CONSTRUCTION OF DIGITAL HUMANITIES SCHOLAR PORTRAIT

3.1. Basic Information about Digital Humanities Scholars

In terms of physical attributes, the group of digital humanities scholars is rich in age levels, forming an age echelon of old, middle-aged and young, as detailed in Figure 3. Most digital humanities researchers are distributed in the age range of 30 to 40 years old. A smaller proportion of digital humanities researchers are in the 20-30 age bracket, accounting for only 6%. From the results, it is clear that there is an urgent need to replenish the digital humanities field with young researchers in their 20s and 30s, as well as a certain number of middle-aged researchers to lead the digital humanities field. In addition, the proportion of male researchers is much higher than that of female researchers, with men accounting for 62% while women only account for 38%,as Figure 4 shown.

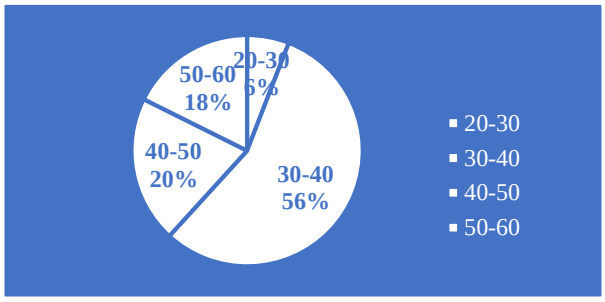


Figure 3 Age Distribution Chart

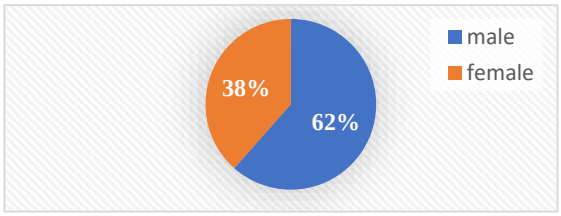
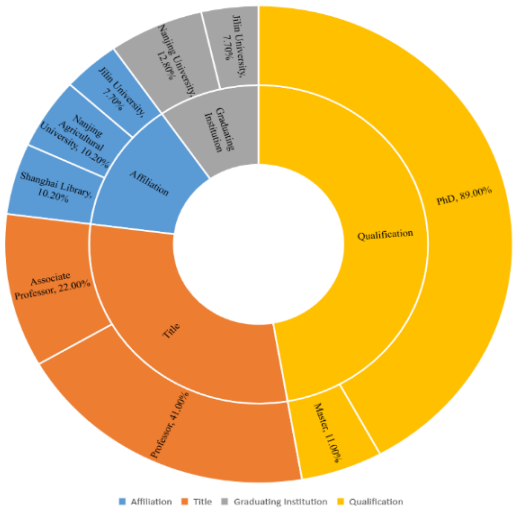


Figure 4 Gender Distribution Chart

After data cleaning, the scholars' affiliations, titles, graduation institutions, highest degrees and academic qualifications were statistically analysed, as shown in Figure 5, where the affiliations were selected from the top 3 institutions in terms of the number of digital humanities scholars distributed.

Figure 5 Social attributes statistical chart



As shown in Figure 5, digital humanities researchers are located in two types of institutions, namely universities and research institutes. At the secondary level, institutions with a large number of digital humanities scholars have set up digital humanities research centers or support platforms to support their digital humanities research, with the exception of Jilin University. The titles of digital humanities researchers are mainly professors or associate professors, with some research librarians and associate research librarians working in libraries. The number of PhD researchers in digital humanities is 89%, and these scholars come from different disciplines, including intelligence, management, engineering, science, history and literature, etc. As scholars in the field of graphical intelligence were the first to come into contact with digital humanities, scholars from intelligence and management disciplines make up a larger proportion. Nanjing University is one of the institutions that incubates the largest number of digital humanities scholars, followed by Jilin University, Sun Yat-sen University, Renmin University of China, Wuhan University and Tsinghua University. These schools pay more attention to digital humanities and, with the exception of Jilin University and Sun Yat-sen University, several other schools have established digital humanities centers, and although Jilin University and Sun Yat-sen University have not established digital humanities centers, seminars and presentations related to digital humanities are held more frequently, which also reates a good academic atmosphere for digital humanities.

3.2. Academic Information of Digital Humanities Scholars

In terms of scholars' published papers, we collected scholars' education and work experiences through the official websites of scholars' affiliated units and scholars' websites, downloaded scholars' papers from the CNKI database, integrated and cleaned the data after downloading, eliminated duplicate data, and finally obtained a total of 274 valid data. In order to discover the hotspots and interests of scholars in the field of digital humanities, CiteSpace was used to analyze the knowledge graphs of scholars' published papers to show the hotspots and frontiers of scholars' research in the field of digital humanities, in order to provide reference for digital humanities research.

3.2.1 Keyword Frequency Analysis

The high-frequency keywords with mediated centrality greater than 0 in the published papers of 39 domestic digital humanities scholars are shown in Table 2. The higher the frequency of keywords, the more they can reflect the research hotspots in the field. The higher the frequency of the keyword, the more it reflects the importance of the term in the network [9].

The frequency and mediating centrality of the keywords in Table 2 show that ‘digital humanities’, ‘digital scholarship’, ‘curatorial resources’, ‘big Data’ are the key areas of research for digital humanities scholars. The year in which the keywords first appear shows that digital humanities scholars have been focusing on research related to digital humanities since

1988, when the term ‘digital humanities’ not only appeared, but also the origin of digital humanities, which is generally accepted by scholars, namely ‘humanities computing’. The term ‘digital humanities’ emerged in 1988, along with what scholars generally agree is the origin of digital humanities, namely ‘humanistic computing’. In addition, high-frequency keywords such as ‘ancient books’, ‘artificial intelligence’, ‘curatorial resources’ and ‘digital historiography’ indicate that, with the development of the Internet and high technology, the field of digital humanities research is becoming increasingly broad.

Table 2. List of keyword terms with betweenness centrality values greater than 0

Keywords	frequency	intermediary	year of first
Digital	581	0.34	1988
Theme model	3	0.07	2017
Digital scholarship	46	0.04	1988
Big Data	5	0.04	2016
Value co-creation	35	0.02	1988
Libraries	9	0.02	2017
Collection	3	0.02	2018
Gi	3	0.02	2018
Augmented reality	3	0.02	2016
North America	3	0.02	2017
Digital	3	0.02	2017
Linked Data	2	0.01	1988
Visualisation	47	0.01	1988
Humanistic	41	0.01	1988
Knowledge	41	0.01	1988
Entity recognition	40	0.01	1988
Ontologies	34	0.01	1988
Artificial Intelligence	31	0.01	2020

3.2.2 Keyword co-occurrence Network Analysis

The keyword co-occurrence network is shown in Fig.6, which is a cluster analysis of the papers published by 39 domestic scholars in the field of digital humanities. The keywords are a high degree of condensation of the content of the papers, and the keyword analysis can quickly understand the research hotspots and frontiers of the research field. The size of the nodes in Figure 6 represents the frequency of keywords, and the lines between the nodes represent the strength of keyword co-occurrence. The smaller the tag number, the more keywords are included under the tag, which shows that 0# ‘Linked Data’ is a key area of interest for scholars. In addition, the role of libraries in the digital humanities and the role they play has also received a great deal of attention from scholars. With the development of digital humanities tools, ‘digital scholarship’, ‘historical codification’ and ‘ancient Chinese language markers’ have come into the forefront of researchers' minds. According to relevant scholars, foreign research in digital humanities focuses on image retrieval, natural language processing, corpus construction, semantic web services, etc[10]. In contrast, domestic digital humanities research is more focused on the application level, and the technologies applied are relatively limited and single, lacking more effective digital humanities research tools and methods.

Figure 6 Keyword co-occurrence network analysis diagram

3.2.3. Author Collaboration Network Analysis

The collaboration network of the 39 digital humanities scholars studied was analysed by CiteSpace, and the final author collaboration network map was generated, as shown in Figure 7. The social relationships of scholars in a certain field can be seen in the authors’ collaborative network analysis map, and the analysis of authors can understand the composition and distribution characteristics of the author team in a certain field. The results show that the number of nodes N=158, the number of connections E=314, and the density Density=0.0253, indicating that the cooperation among domestic digital humanities scholars is relatively close. Digital humanities research scholars have formed two larger collaborative networks centred on Wang Dongbo and Liu Wei, with the largest collaborative network centred on Wang Dongbo scholars, with whom the authors who work closely include Huang Shuiqing, Meng Kai and Hu Haotian, all of whom come from Nanjing Agricultural University, and the group's research in the field of digital humanities focuses on the application of digital humanities to antiquities and historiography. The main co-authors of this group include Chen Tao, Zhao Yuxiang, Ye Ying, Zhu Qinghua and Shan Rongrong. The research themes of this group focus on the theoretical construction and development of digital humanities and libraries.

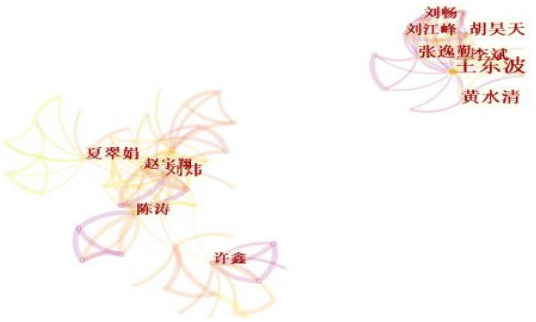


Figure 7 Author collaboration **network** analysis chart

Figure 8 Institutional cooperation network analysis chart

3.2.4. Analysis of Institutional Cooperation Networks

The analysis of institutional cooperation in the field of digital humanities was carried out to generate a network analysis diagram of institutional cooperation, as shown in Figure 8. The research sample contains a total of 123 nodes, and the number of connections and density are both zero, indicating that there are more digital humanities research institutions.however, the strength of cooperation between research institutions is weak, which means that research institutions in the field of digital humanities are relatively scattered and have not formed extensive cooperative relationships. Not only is the cooperation between institutional institutions poor, as shown in Figure 8, there are also no connecting lines between different faculties within the same institutional institution, and it can be inferred that close cooperation between different faculties within the same institutional institution has not been formed either.

3.2.5. Analysis of Author Projects

An analysis of the project fees of 39 digital humanities scholars in China (Figure 9) shows that national projects account for the largest proportion, 37.94%, and provincial projects 32.15%, which shows that the state and relevant provincial and ministerial units attach importance to digital humanities research. In the future, scholars should strengthen international cooperation, and at the same time should cooperate with enterprises with the support of local governments to realize the transformation of results.

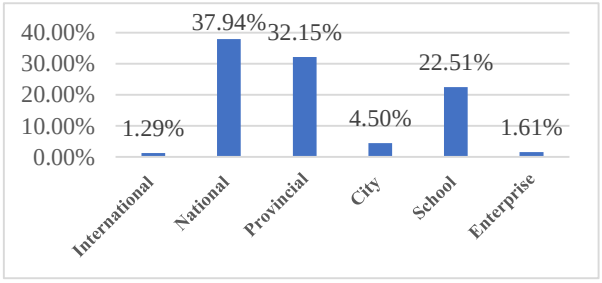


Figure 9 Project level ratio

3.3 Portrait of Digital Humanities Scholars

Through the analysis of the above data, a portrait of digital humanities scholars is formed, which is shown in Figure 10. Through the above analysis of the data on both basic information and academic information of digital humanities scholars, a portrait of scholars can be constructed in four dimensions: physical attributes social attributes, published papers, and participation in projects. Basic information is a statistic of the overall characteristics of the researcher, while academic information is a discovery of the research interests and preferences of the researcher.

Figure 10 A sample portrait of a scholar in the digital humanities

4. CONCLUSIONS AND RECOMMENDATIONS

This paper constructs a portrait of the basic and academic information of China's digital humanities scholars from four dimensions, and the research results show that the number of young scholars in the field of digital humanities in China is relatively small at present; the hotspots of scholars' research focus on knowledge organization and pulse combing such as canonization, archival management, digitization of ancient books, historical research, semantic annotation, etc.And there is a lack of research on the development and technical application of digital humanities tools; in terms of communication In terms of cooperation, there is less close cooperation between different institutions and within institutions. Early researchers in the field of digital humanities research can actively seek help or cooperation from senior members of the group [11]; by exploring the changing paths of the academic research themes of scholars in the group, they can understand the development of the field; by exploring the cooperation networks between authors and institutions, they can obtain the latest academic conferences and reports in the field and grasp the latest trends in the field. In view of the current situation of group research in the field of digital humanities in China, the following suggestions are made.

First, cultivate and nurture young scholars. The gender composition of the sample shows that the proportion of young scholars in the field of digital humanities research is currently much lower than that of middle-aged scholars, while young talents are the group with the most innovative ability and enthusiasm for innovation. The lack of young talents will make it difficult for the field of digital humanities to develop in depth and sustainably in China, and without the promotion of young talents, the research on digital humanities in China will remain in the construction of theories and frameworks for a long time. As a field that has emerged from the development of the times and is multidisciplinary, the development of digital humanities needs to absorb the latest technological achievements of the times, and young talents who are the bearers of new technologies are exactly what the field needs.

Secondly, we need to strengthen inter-institutional links and cooperation. Judging from the author cooperation network map, China's digital humanities scholars work closely, but usually the authors who work closely belong to the same institution. Judging from the institutional cooperation network map, China's digital humanities research institutions have a low degree of cooperation, are not closely connected, and have a single core institution. Institutions and other institutional units should strengthen cooperation and ties, while enhancing the plurality of cooperating institutions and making full use of the distinctive resources of each institution to achieve complementary advantages and resource sharing, thus promoting the expansion of the influence of research in the field of digital humanities in China and attracting more scholars to join.

Thirdly, the development, application and training of tools and techniques should become the focus of attention for digital humanities scholars. As can be seen from the keyword co-occurrence network mapping, China's digital humanities research is still largely limited to the application level of knowledge organization, canonization and digitization, and archive management. There is an urgent need to develop more tools and methods suitable for the humanities research perspective at the development level, and to strengthen training in their application. The key to breakthroughs in digital humanities research also lies in how to apply tools, techniques and methods to the humanities and produce research results that are different from traditional humanities.

ACKNOWLEDGMENT

Fund project: 2020 Jiangsu Provincial Higher Education Digital Library Project, the research on the new generation of library innovation activities under the digital humanities background (2020KT13);

2020 Central University Basic Research Fund Special Funding Project (Think Tank Innovation Fund), based on multi-source data Analyze the research results of the scholar portrait construction research (NK2020009).

REFERENCES

[1] Yan Y X. A study of digital humanities projects in China in the last decade [J]. Media Forum, 2021, 4(19):173-174.

[2] Zhou D M. Shanghai library: opening the volume to search for roots and promoting the digitalization of genealogy [J]. Chinese children, 2020(3):80-81.

[3] Digital Humanities Research Center, Wuhan University. Understanding, organizing and displaying the content of the Dunhuang Mogao Caves murals [EB/OL]. [2021-10-09]. http://dh.whu.edu.cn.

[4] Ai Y X, Xu J, He L, Qi Y. Research on the construction method of classification system for content analysis of canonical books [J]. Journal of Agricultural Library Information, 2021, 33(9):18-26.

[5] Wang X P, Wang D L, Zhao H Y. Research on the evaluation of disciplines based on user portraits: an example of first-level aerospace science and

technology [J]. Journal of Library and Information Studies, 2021, 6(7):53-61.

[6] Li B. Design of personalized resource recommendation service for university libraries based on user portraits [J]. New Century Library (4):8.

[7] Cheng D L, NIE Y M, HAN S F. Research on precise poverty alleviation services based on big data portraits [J]. Wireless Connected Technology, 2019, 16(22):160-161.

[8] Guo J L, Fang Y Sh, Huang Q W,et al. Research on the application of talent portrait technology based on big data[J]. Microcomputer Applications, 2021, 37(6):166-170.

[9] Zhang Q, Sun Ch P. Knowledge mapping analysis of business environment research priorities and trends in China based on CiteSpace [J]. Statistics and Management, 2021, 36(11):124-128.

[10] Cui Q Q. Visual analysis of domestic and international digital humanities research based on knowledge graph [J]. Library World, 2020(6):48-54, 65.

[11] Xu T Y, Wang Sh Y. Diverse cooperation and institutional guidance: An exploration of the characteristics and insights of digital humanities projects in Germany [J/OL]. Library Construction: 1-15 [2021-10-04].