Tech Science Press

Provenance Method of Electronic Archives Based on Knowledge Graph in Big Data Environment

Chun Xu[1] and Jiang Xu[1,2,*]

[1]Nanjing University of Information Science & Technology, Nanjing, 210044, China
[2]Engineering Research Center of Digital Forensics, Ministry of Education, Nanjing, 210044, China
*Corresponding Author: Jiang Xu. Email: renyj100@126. com

Abstract: With the advent of the era of big data, the Provenance Method of electronic archives based on knowledge graph under the environment of big data has produced a large number of electronic archives due to the development of science and technology. How to guarantee the credential characteristics of electronic archives in the big data environment has attracted wide attention of the academic community. Provenance is an important technical means to guarantee the certification of electronic archives. In this paper, knowledge graph technology is used to provide the concept provenance of electronic archives in large data environment. It not only enriches the provenance method, but also guarantees the certification of electronic archives in the large data environment.

Keywords: Knowledge graph; big data; electronic archives; conceptual provenance; quality assessment

1 Introduction

In the era of big data, the connotation and extension of data have undergone tremendous changes [1]. The living environment and life cycle of data are also different from those of information age, which brings various management and governance risks. Therefore, "data management" has become a hot topic in many disciplines from the perspective of big data [2]. As a historical record carefully preserved by the state and society, electronic archives are of great value. The arrival of the era of big data has given new features to the data resources of electronic archives. The main participants in the management of electronic archives are increasingly aware of the value of electronic archives [3].

Under the same technical conditions, the common characteristics of large archives data and big data are huge amount of data, various types of data and fast processing speed; while slightly different characteristics are: relatively scattered data, high value density of single type of electronic archives data [4], high reliability and fidelity of data.

The data resources of electronic archives are large and growing fast as a whole. At present, the total amount of electronic archives resources in single archives in China has not reached the level of PB, but the total amount of archives resources in each archive (office) can be regarded as massive. At the same time, with the digitalization of existing archives and the increase of new electronic archives received each year, the number of electronic archives has formed a huge scale.

The data types of electronic archives are various, and the structure is complex. Electronic archives involve government agencies, finance, education and other industries. Each industry produces different archival data, not only traditional paper archives, but also electronic archives such as pictures, charts, audio, video, etc. The formats and features are various, thus forming a large number of heterogeneous data [5]. So, there are a lot of structured and unstructured data in archives data resources.

High speed of data processing of electronic archives. In the era of big data, the timeliness of electronic archives is more obvious, and it is more necessary to predict what will happen in time [6], which requires fast processing speed, which is also the most remarkable feature different from traditional data. Secondly, the speed and efficiency of electronic archives processing is the life of industry or enterprise.

The value of archival data resources is high. Compared with the characteristics of high total value and low value density of most data resources, electronic archives are the most authentic, reliable, authoritative and credential primary information resources [7]. The more people they use, the greater their value. Electronic archives are relatively scattered, which means that due to the separation effect of if2012.

 scope and storage period, electronic archives in space and time range are often stored in different physical spaces [8], although the numbers can be enormous, but it is difficult to complete. The high value density of single kind of electronic archives means that the value density of electronic archives is higher than that of general data without identification because of archival archiving and identification [9]. The high density of data value comes from the value of archives and is proportional to the value of archives.

Archives objectively record the historical situation and retain the true historical markers, which is convincing historical evidence [10]. The value of vouchers is the basic value of electronic archives, and the real electronic archives have the value of vouchers. One of the reasons for maintaining the veracity of electronic archives is to give full play to the evidential value of electronic archives [11]. Data provenance is an important technical means to ensure the veracity of electronic archives. However, in the current large data environment, the number of electronic archives has increased dramatically [12], and the traditional provenance method cannot meet the current needs. Therefore, this paper studies the Provenance Method of electronic archives in the large data environment.

2 Related Work

2.1 Knowledge Graph

In essence, knowledge graph is a semantic network that reveals the relationship be-tween entities, and it can formally describe the real-world things and their relationships [13]. Now the knowledge graph has been used to refer to a variety of large-scale knowledge base.

Triple is a general representation of knowledge graph, that is, is the entity aggregate in the knowledge base, the number of different entities is |E|. Is a set of relationship in it, which contains |R| different relationships, represents the triple set in the knowledge base [14] The basic form of triple includes Entity 1, relationship, Entity 2 and concept, attribute and attribute value. Entity is the most basic element in knowledge graph, and different entities have different relationships [15]. Concepts mainly refer to collections, categories, object types [16], and the types of things, such as people and geography; Attributes mainly refer to the attributes, characteristics, characteristics and parameters that an object may have, such as nationality, birthday, etc. Attribute values mainly refer to the values of the attributes specified by the object, such as China, 1988-09-08, etc. Each entity can be identified by a globally unique ID, each pair (attribute-attribute value) can be used to describe the intrinsic characteristics of the entity, and the relationship can be used to connect two entities and describe the relationship between them.

2.1.1 Logical Structure

The structure of knowledge graph of logical structure includes the logical structure of knowledge graph itself and the technology (system) framework used to construct knowledge graph.

Logically, the knowledge graph can be divided into two layers: data layer and pat-tern layer. In the data layer of the knowledge graph, knowledge is stored in units of fact in the graph database [17]. If the "entity-relation-entity" or "entity-attribute-property value" triples are taken as the basic expression of fact, all the data stored in the graph database will form a huge entity-relationship network and form the graph of knowledge [18]. The schema layer is the core of the knowledge graph. What is stored in the pattern layer is the extracted knowledge. Ontology library is usually used to manage the pattern layer of knowledge graph, and the support ability of ontology library to axioms, rules and constraints is used to

regulate the relations between entities, relationships, types and attributes of entities and other objects.

2.1.2 Architecture

The architecture of knowledge graph refers to the construction of pattern structure, as shown in Fig. 1.
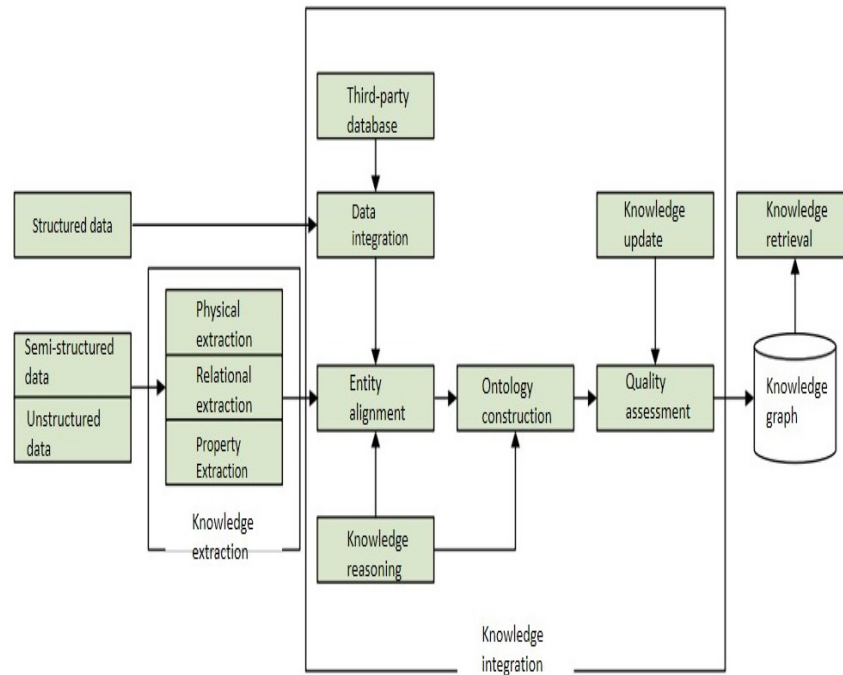
Figure 1: The architecture of knowledge graph

The part of the dotted line frame is the process of constructing knowledge graph, which needs to be updated and iterated with people's cognitive ability.

There are two main ways to construct knowledge graph: top-down and bottom-up. Top-down refers to defining ontology and data patterns for knowledge graph, and then adding entities to the knowledge base [19]. This method needs to use some existing structured knowledge base as its basic knowledge base, such as Freebase project. Bottom-up refers to extracting entities from some open linked data, choosing those entities with high confidence to join the knowledge base, and then building the top-level ontology model.

3 Problem Statement

At the 13th International Archives Congress, Terry Cook, a Canadian archivist, developed the traditional entity provenance into a highly abstract and generalized concept provenance on the basis of summing up the experience and lessons of archival practice in various countries. Conceptual provenance holds that provenance is not only the originator of electronic archives, but also an abstract and broad-ranging relationship [20]. It is an abstract concept provenance focusing on the functions and business activities of the archivists within the organization. This theoretical innovation sublimates the principle of provenance in theory and deepens the understanding of the essence and value of electronic archives.

3.1 The Position of Provenance Principle in the Electronic Age

Under the current big data environment, a large number of electronic archives have been produced. We should go deep into the concrete formation process of electronic archives and understand the origin and development of document information [21]. In essence, the principle of provenance requires us to fully understand and maintain the traceable information of documents, while in the era of electronic archives, it has risen to the sense of background information acquisition. The content and emphasis of the

description have changed. The description of electronic archives should involve various elements in the formation of documents, including the background, content and structure. The emphasis of the description should be shifted from the content of documents to the background of the formation of documents.

The principle of provenance embodies the interrelationship between the formations of electronic archives and ensures the evidential value of electronic archives. Electronic archives do not exist in isolation and without connection [22]. Each electronic archive exists as an integral part of the whole archive. In order to find out whether an electronic file is of great significance, it is necessary to know exactly who the file is, under what conditions, for what purpose, and in what form it is structured.

3.2 The Role of Provenance Principle in the Electronic Age

The role of provenance principle in electronic archives filing. On the one hand, the traditional archives work still exists objectively. As a sort-out principle, the principle of provenance classifies archives and libraries from the perspective of provenance so as to distinguish them from each other in basic methods [23]. The principle of provenance fully guarantees the integrity of the whole file and facilitates the filing staff to sort out the electronic files. The principle of provenance requires that electronic archives be classified according to their sources. It is required to classify electronic archives according to their provenance in public administration organs [24], so as to maintain the organic connection of electronic archives and the inherent historical connection between documents. According to the principle of provenance, filing helps to protect the integrity of electronic archives, to explain the importance of documents, to maintain the evidential value of documents, and to provide convenience for the subsequent management links. On the other hand, electronic archives have a series of technical characteristics. These characteristics are different from traditional documents, but all these characteristics have not fundamentally changed the characteristics of electronic archives as archives [25]. Still like traditional documents, they are true records of human activities. Therefore, it is equally important to maintain the historical organic relationship between documents for the collation of electronic archives.

The role of provenance principle in identification [26]. The concept of provenance plays an important guiding role in the field of electronic archives identification. The principle of provenance requires that the original history of electronic archives be maintained. According to the principle of provenance, electronic archives can maintain their complete provenance information and system. Its identification depends on real and reliable provenance. The electronic archives management organization is the starting point of the appraisal work. The identification work based on the formation of electronic archives has increasingly shown its feasibility and reliability.

In the current big data environment, how to provide concept provenance lacks the corresponding technical means. Therefore, this paper proposes to use knowledge graph technology to provide concept provenance [27].

4 Provenance Method of Electronic Archives Based on Knowledge Graph

4.1 Information Extraction of Electronic Archives

Information extraction is the first step in the construction of knowledge graph. Information extraction of electronic archives based on knowledge graph can be obtained by entity extraction, relationship extraction and attribute extraction [28].

With the continuous progress of named entity recognition technology, academia began to pay attention to the problem of information extraction in the open domain, that is, no longer limited to specific knowledge areas, but oriented to the open internet, to study and solve the problem of information extraction in the whole network. Firstly, a scientific and complete named entity classification system is established, which can be used to guide the research of algorithm on the one hand, and to facilitate the management of the extracted entity data on the other hand. Finally, the automatic classification of entity was realized by using adaptive perceptron algorithm.

Statistical machine learning (SML) is widely used to model the relationships between entities,

instead of predefined grammatical and semantic rules. For example, Lexical, syntactic and semantic features of natural language were used to model entity relationships, and the rule-free approach by maximum entropy method was successfully realized. Moreover, a large number of supervised learning methods based on eigenvectors or kernels are used in relation extraction, which makes the accuracy improve continuously.

The goal of attribute extraction is to collect attribute information of a specific entity from different sources. For example, for a public figure, information such as nicknames, birthdays, nationalities, educational backgrounds can be obtained from the public information on the Internet [29–30]. Attribute extraction technology can collect the information from a variety of data sources to realize the entity belongings. Realize the complete sketch of entity attributes.

4.2 Integration of Knowledge in Electronic Archives

Through information extraction, the goal of obtaining entity, relationship and entity attribute information from unstructured and semi-structured electronic archives is achieved. However, these results may contain a lot of redundant and error information [31], and the relationship between electronic archives data is flattened and lacks layers. Secondary and logical, so it is necessary to clean up and integrate them. Knowledge fusion includes two parts: entity link and knowledge merge. Through knowledge fusion, the ambiguity of concepts can be eliminated, redundancy and erroneous concepts can be eliminated, so as to ensure the information quality of electronic archives.

Entity linking refers to the operation of linking the entity object extracted from electronic archives to the corresponding correct entity object in the knowledge base [32]. The basic idea of entity link is to select a group of candidate entity objects from the knowledge base according to the given entity reference items, and then link the reference items to the correct entity objects through similarity calculation. In recent years, academia has begun to pay attention to utilizing the co-occurrence relationship of entities, while linking multiple entities to the knowledge base, called collective entity linking [33].

The general process of entity linking is extracting entity reference terms from electronic archives through entity extraction; disambiguating entity and co-referential resolution, judging whether the same-name entity in the knowledge base represents different meanings and whether there are other named entities in the knowledge base that express the same meanings [34]; confirming that the same-name entity in the knowledge base represents the same meanings. After the correct entity object, the entity reference is linked to the corresponding entity in the knowledge base.

Entity disambiguation is a technology specially used to solve ambiguity problem of the same name entity. In real language environment, we often encounter the problem that an entity reference item corresponds to multiple named entity objects. For example, the term "Li Na" can correspond to the entity Li Na as a singer, or to the entity Li Na as a tennis ball. By entity disambiguation, athlete Li Na can establish entity links accurately according to the current context. Entity disambiguation mainly adopts clustering method.

Entity resolution technology is mainly used to solve the problem that multiple reference items correspond to the same entity object. For example, in a document, many pronouns may point to the same entity object. By using co-referential resolution technology, these referential terms can be merged into the correct entity object [35]. This problem is of special importance in the fields of information retrieval and natural language processing and attracts a lot of research efforts [36]. Academic circles have different expressions on this issue, such as object alignment, entity matching and entity synonyms. In addition to the problem of co-referential resolution as a classification problem, it can also be solved as a clustering problem. The basic idea of clustering method is to focus on entity referential terms and realize the matching between referential terms and entity objects through entity clustering [37]. The key problem is how to define the similarity measure between entities.

4.3 Knowledge Processing of Electronic Archives

Through information extraction, knowledge elements such as entities, relationships and attributes can be extracted from the original corpus. After knowledge fusion, the ambiguity between entity reference items and entity objects can be eliminated and a series of basic factual expressions can be obtained. However, facts themselves are not equal to knowledge [38], and in order to obtain structured and network finally. Knowledge processing includes three aspects: ontology construction, knowledge reasoning and quality evaluation.

Ontology is a norm for modeling concepts and an abstract model for describing the objective world. It defines concepts and their links in a formal way. The greatest feature of ontology is that it is shared, and the knowledge reflected in ontology is a well-defined consensus [39]. In knowledge graph, ontology is located in the pattern layer, which is used to describe the conceptual hierarchy system as the conceptual template of knowledge in the knowledge base. At present, the research on Ontology generation methods is mainly focused on entity clustering. The main challenge is that the entity description obtained by information extraction is very short and lacks the necessary context information, which makes most statistical models unavailable.

Knowledge reasoning refers to expanding and enriching knowledge networks by establishing new relationships among entities based on the existing entity relationship data in knowledge base through computer reasoning [40]. Knowledge reasoning is an important means and key link of knowledge graph construction. Through knowledge reasoning, new knowledge can be found from existing knowledge. The object of cognitive reasoning is not limited to the relationship between entities, but also can be the attribute value of entities and the conceptual hierarchical relationship of ontology. Conceptual reasoning can also be carried out according to the inheritance relationship of concepts in ontology library.

Quality assessment is also an important part of knowledge base construction technology. Due to the limitation of current technology level, there may be errors in knowledge elements obtained by using open domain information extraction technology such as entity recognition errors, relationship extraction errors, etc. The quality of knowledge obtained by knowledge reasoning is also not guaranteed. Therefore, before adding it to the knowledge base, we need a process of quality assessments [41]. With the development of open related data projects, the quality differences among the knowledge base products produced by each sub-project are also increasing, and the conflicts between data are increasing. How to evaluate the quality of the knowledge base and the global knowledge is becoming more and more important. The construction of atlas plays an important role. The significance of introducing quality assessment is that the credibility of knowledge can be quantified, and the quality of knowledge base can be guaranteed by discarding the knowledge with lower confidence.

4.4 Intelligent Search of Electronic Archives Based on Knowledge Graph

After the above processing, we can build an intelligent search tool for electronic archives based on knowledge graph. Intelligent search based on knowledge graph is a search based on long tail. Search engines display search results in the form of knowledge cards. The query request of electronic archives users will go through two stages: query semantic understanding and knowledge retrieval.

Query-based semantic understanding. The semantic analysis of query type in knowledge graph mainly includes: word segmentation, part-of-speech tagging and error correction of query request text; description normalization to match knowledge in knowledge base [42]; and context analysis. In different contexts, the object of user query will be different, so knowledge graph needs to integrate the user's emotions at that time and feedback the user's answers in time query expansion. After defining the query intention and related concepts of the users of electronic archives, the relevant concepts in the current context are added to expand.

Knowledge retrieval. After query analysis, the standard query statements are entered into the knowledge base search engine, which retrieves the corresponding entities in the knowledge base and the entities with high matching degree in category, relationship, correlation and so on.

5 Conclusion

With the advent of the era of big data, the Provenance Method of electronic archives based on knowledge graph under the environment of big data has produced a large number of electronic archives due to the development of science and technology. How to guarantee the credential characteristics of electronic archives in the big data environment has attracted wide attention of the academic community. Provenance is an important technical means to guarantee the certification of electronic archives. In this paper, knowledge graph technology is used to provide the concept provenance of electronic archives in large data environment. It not only enriches the provenance method, but also guarantees the certification of electronic archives in the large data environment.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

[1] Y. Lu, "Research on trusted DNP3-BAE protocol based on hash chain," *EURASIP Journal on Wireless Communications and Networking*, vol. 53, no. 1, pp. 107–108, 2018.

[2] P. Cui, "Provenance and connotation of electronic documents," *Archives and Construction*, vol. 64, no. 2, pp. 4–7, 2013.

[3] D. Chen, G. B. Han, B. Wang and S. X. Yang, "Typical application of knowledge atlas in equipment big data," *Network Security Technology and Application*, vol. 36, no. 1, pp. 69–71, 2019.

[4] L. L. Yan, Y. Chang and S. B. Zhang, "Measure-resend semi-quantum private comparison scheme using GHZ class states," *Computers, Materials & Continua*, vol. 61, no. 2, pp. 877–887, 2019.

[5] Y. P. Liu, Y. J. Ren, C. P. Ge, J. Y. Xia and Q. R. Wang, "A CCA-secure multi-conditional proxy broadcast re-encryption scheme for cloud storage system," *Journal of Information Security and Applications*, vol. 47, no. 2, pp. 125–131, 2019.

[6] Z. J. Chang, K. Bei, Q. Li, Y. Wang and J. Xie, "Knowledge atlas construction model and methodology for large data of science and technology," *Data Analysis and Knowledge Discovery*, vol. 3, no. 1, pp. 15–26, 2019.

[7] Y. Chen, W. Liu and X. B. Ma, "Visual analysis of domestic large data research based on knowledge map," *Journal of University of Electronic Science and Technology (Social Science Edition)*, vol. 20, no. 3, pp. 9–15, 2018.

[8] L. Chen, J. Y. Li and S. C. Xie, "Computer network information security in the era of big data." *Electronic Technology and Software Engineering*, vol. 25, no. 9, pp. 178–179, 2019.

[9] H. Fan and H. Zhu, "Motion vector detection based on local autocorrelation coefficient," *Cluster Computing*, vol. 22, no. 5, pp. 11633–11639, 2019.

[10] C. X. Wang, X. Shao, Z. Gao, C. X. Zhao and J. Gao, "Common network coding condition and traffic matching supported network coding aware routing for wireless multihop network," *International Journal of Distributed Sensor Networks*, vol. 15, no. 6, pp. 1–20, 2019.

[11] J. Y. Ding and X. C. Wang, "Knowledge map of big data: concept, characteristics, application and impact." *Information Science*, vol. 31, no. 9, pp. 10–14+136, 2013.

[12] G. S. Li, Y. C. Liu, J. H. Wu, D. D. Lin and S. S. Zhao, "Methods of resource scheduling based on optimized fuzzy clustering in fog computing," *Sensors*, vol. 19, no. 9, pp. 1–16, 2019.

[13] L. Yin, C. Zhou, L. Li, Y. Su and C. Xia, "A physiological and behavioral feature authentication scheme for medical cloud based on fuzzy-rough core vector machine." *Information Sciences*, vol. 43, no. 5, pp. 143–160, 2020.

[14] X. J. Zhao, X. H. Zhang, P. Wang, S. L. Chen and Z. X. Sun, "A weighted frequent itemset mining algorithm for intelligent decision in smart systems," *IEEE Access*, vol. 25, no. 6, pp. 29271–29282, 2018.

[15] Y. F. Guo, L. Han, L. H. Pi and K. Song, "Application of knowledge atlas in large data." *Telecommunication Technology*, vol. 63, no. 2, pp. 25–29, 2015.

[16] B. Hu, F. Xiang, F. Wu, J. Liu, Z. Sun *et al.*, "Research on time synchronization method under arbitrary network delay in wireless sensor network," *Computers, Materials & Continua*, vol. 19, no. 3, pp. 1323–1344, 2019.

[17] L. Gong, B. Yang, T. Xue, J. Chen and W. Wang, "Secure rational numbers equivalence test based on threshold cryptosystem with rational numbers." *Information Sciences*, vol. 25, no. 4, pp. 44–54, 2018.

[18] W. Liu and J. Guo, "ETC-IoT: Edge-node-assisted transmitting for the cloud-centric Internet of Things," *IEEE Network*, vol. 32, no. 3, pp. 101–107, 2018.

[19] G. S. Li, J. H. Yan, L. Chen, J. H. Wu, Q. Y. Lin *et al.*, "Energy consumption optimization with a delay threshold in cloud-fog cooperation computing," *IEEE Access*, vol. 7, no. 1, pp. 159688–159697, 2019.

[20] Y. H. Xia, C. Ying, G. F. Lin and Z. X. Sun, "A third-party mobile payment scheme based on NTRU against quantum attacks," *IEEE Access*, vol. 34, no. 7, pp. 56070–56080, 2019.

[21] H. B. Liu, "Model exploration of knowledge map in the application of intellectual property data," *Science and Technology and Publishing*, vol. 16, no. 2, pp. 95–98, 2018.

[22] L. Y. Cao, W. M. Ding, L. P. Wang and X. Y. Wu, "Application of knowledge graph based on big data," *Science and Technology Horizons*, vol. 20, no. 2, pp. 89–90+56, 2019.

[23] Y. J. Ren, Y. Leng, F. J. Zhu, J. Wang and H. J. Kim, "Data storage mechanism based on blockchain with privacy protection in wireless body area network," *Sensors*, vol. 19, no. 10, pp. 2395, 2019.

[24] W. N. Wan, J. Chen and S. B Zhang, "A cluster correlation power analysis against double blinding exponentiation," *Journal of Information Security and Applications*, vol. 48, no. 10, pp. 1–8, 2019.

[25] Q. L. Wang and H. Wang, "Hotspots and topic analysis of educational big data research based on knowledge map," *Chinese Medical Education Technology*, vol. 31, no. 4, pp. 367–371, 2017.

[26] Y. Chang, "Practical two-way QKD-Based quantum private query with better performance in user privacy," *International Journal of Theoretical Physic*, vol 58, no. 7, pp. 2069–2080, 2019.

[27] J. Ren. "Research status and trend of publishing in the big data era: knowledge map analysis based on CSSCI journal (2008–2017)," *Publishing and Printing*, vol. 34, no. 1, pp. 29–37, 2018.

[28] Y. J. Ren, F. J. Zhu, K. S. Pradip, W. Tian, J. Wang *et al.,* "Data query mechanism based on hash computing power of blockchain in Internet of Things," *Sensors*, vol. 20, no. 1, pp. 3390, 2020.

[29] Z. Sun, Y. R. Bi, S. L. Chen, B. Hu, F. Xiang *et al.,* "Designing and optimization of fuzzy sliding mode controller for nonlinear systems," *Computer, Material & Continua*, vol. 19, no. 1, pp. 12, 2019

[30] S. L. Chen, L. T. Zheng, Y. Zhang, Z. X. Sun and K. Xu, "VERAM: view-enhanced recurrent attention model for 3D shape classification," *IEEE Transactions on Visualization and Computer Graphics*, vol. 25, no. 12, pp. 3244–3257, 2019.

[31] Y. Chang, "A quantum authorization management protocol based on EPR pairs," *Computer, Material & Continua*, vol. 59, no. 3, pp. 1005–1014, 2019.

[32] Z. Chuan, "Knowledge atlas analysis of big data research," *Library Forum*, vol. 35, no. 4, pp. 84-91, 2015.

[33] Z. G. Qu, Z. Y. Li, G. Xu, S. Y. Wu and X. J. Wang, "Quantum image steganography protocol based on quantum image expansion and grover search algorithm," *IEEE Access*, vol. 25, no. 7, pp. 50849–50857, 2019.

[34] Z. G. Qu, Z. W. Cheng, W. J. Liu and X. J. Wang, "A novel quantum image steganography algorithm based on exploiting modification direction," *Multimedia Tools and Applications*, vol. 36, no. 7, pp. 7981–8001, 2019.

[35] H. Zhu, H. Fan and Z. Shu, "Edge detection with chroma components of video frame based on local autocorrelation," *IEEE Access*, vol. 15, no. 7, pp. 48543–48550, 2019.

[36] H. Z. Wang, Z. L. Jian and X. long, "LKAQ: Large-scale knowledge graph approximate query algorithm," *Information Sciences*, vol. 25, no. 2, pp. 505, 2019.

[37] Z. Hao, F. F. Hou, Y. Q. Liang, H. L. Shao and F. Zhang, "Visual analysis of large data research based on knowledge atlas in the international perspective," *Library Journal*, vol. 35, no. 5, pp. 13–19+31, 2016.

[38] J. Xu, Y. J. Zhang, K. Y. Fu and S. Peng, "SGX-Based secure indexing system," *IEEE Access*, vol. 36, no. 2, pp. 77923–77931, 2019.

[39] S. W. Lei. "Research progress of large data at home and abroad based on knowledge map," *Library and Information Research*, vol. 10, no. 2, pp. 62–69, 2017.

[40] L. Mei and J. W. Zhao, "Bibliometrics and knowledge mapping analysis of intelligence education in China: Based on CNKI 2004–2018," *Adult Education*, vol. 39, no. 9, pp. 14–21, 2019.

[41] S. B. Zhang, Y. Chang, L. L. Yan, Z. W. Sheng, F. Yang *et al.,* "Quantum communication networks and trust management: a survey," *Computer, Material & Continua*, vol. 61, no. 3, pp. 3, 2019.

[42] M. X. Cui, H. Y. Luo, X. Y. Lin and Q. Shao, "Knowledge map analysis of internet public opinion research in China under the big data environment," *Information Exploration*, vol. 34, no. 5, pp. 128–134, 2019.