Knowledge Organization Model Construction and Application of Oral Archives Based on Knowledge Graph

LU Yonghe, XIAO Ying
School of Information Management,
Sun Yat-sen University, Guangzhou, China

Abstract. Oral archives are not only records of social history, but also the valuable resources in digital humanities research. With the development of digitalization of archives, the traditional coarse-grained organization methods are not conducive enough to the mining and development of archives knowledge. Hence, based on the knowledge graph, combining automatic extraction methods and expert review, the research constructs the knowledge organization model of oral archives. Firstly, the research constructs the ontology of oral archives based on protégé, which includes 6 core concepts, 20 sub-class concepts and 32 concept properties. Besides, taking the oral archive database "Hong Kong Voices" as an example, extracting the knowledge units of oral archives, the ontology model is applied to instantiate the knowledge graph of oral archives. Finally, based on the oral archives knowledge graph, analyze the application scenarios of the knowledge organization model. The research focuses on features of oral archives and apply knowledge graph construction on archive work, so as to optimize the archives knowledge organization method and the development of archives knowledge services in the times of digital humanities.

Key words: Oral archives, Knowledge organization, Knowledge graph, Digital humanities

Introduction

As the original record of social history, oral archives are of great value in humanities research and also the important resources of digital humanities. With the popularity of social memory rescue awareness and the development of information technologies, such as multimedia and big data, the researches on collection, management, development and utilization of oral archives are deepening gradually. Most of the existing oral history projects adopt digital technology to save and manage oral archives by building oral archives database (Nie, & Li, 2020). However, the coarse-grained organization methods at document level are difficult to adapt to the diversified needs of archives users in knowledge organization. Facing the massive digital archives, it is urgent to develop the knowledge organization methods regarding fine-grained knowledge elements as organization object, so as to provide more intelligent and convenient archive services. With the advent of knowledge organization era, the combination of digital humanities theories and technologies is the trend of archives research and practice. Knowledge graph, as an important technology of archive knowledge organization, has been widely concerned and gradually applied in practice.

Knowledge graph is an important direction of development in the field of knowledge organization, which is the structured semantic knowledge base (Pujara, et al., 2013), transforming the extracted entities and relations between entities into the form of graph. At present, the researches and practices of knowledge organization on oral archives from the perspective of digital humanities are still in the very initial stage. The researches mainly focus on the construction, development and dissemination of digital archives (Pan, 2012). However, few studies have applied digital technologies such as knowledge graph to organization, management and utilization of oral archives so as to exert greater value of archives. Therefore, in era of digital humanities, this study regards knowledge graph as the breakthrough point to knowledge organization of oral archives, and analyzes the theme, content, background and

relations of oral archives. Additionally, taking the "Hong Kong Voices" oral archive database as an example, this paper based on knowledge graph constructs the oral archive knowledge organization model. It not only provides theoretical and practical support for improving the efficiency to acquire oral archive knowledge, but also helps to optimize the collection and construction of oral archives knowledge organization.

Literature Review

In the era of digitalization and intellectualization, archives organization need to adapt to the knowledge organization based on knowledge elements to meet the diversified utilization needs of users. Existed researches can be generally divided into three aspects: basic theories of archives knowledge organization, methods of archives knowledge organization, and tools and applications of knowledge organization. In terms of theories, Wang proposed that knowledge organization is the basis of realizing knowledge management and service of archives. It needs to go through three stages: knowledge elements indexing, ordering and structured organization, and knowledge presentation (Wang, 2008). In the research of knowledge organization methods, scholars mainly put forward the methods, such as topic classification, metadata and knowledge base (Li, 2013). Wang proposed that archives knowledge organization needs to go through several steps: knowledge definition, metadata management, knowledge mining and knowledge storage (Wang, 2008). In view of specific knowledge organization tools and applications, Lv constructed domain ontology of archives from the perspective of knowledge relation to improve the quality of linked data of archives (Lv, 2012). Mao focused on the construction of archive knowledge base and analyzed it from the perspectives of classification indexing, relation clustering and reasoning of knowledge elements (Mao, 2015).

Knowledge graph is an important tool for knowledge organization, management, development of archives, and researches on archives knowledge graph are increasing gradually. In current researches, scholars constructed knowledge graph of overall knowledge structure of archives (Guo, 2019), or for specific archive domains, such as celebrity archives (Sun, et al., 2020), scientific research archives (Lei, et al., 2020), and student archives (Shu, 2020), etc. Based on knowledge graph, Yang constructed the subject clustering knowledge graph of China Unicom archives to realize intelligent semantic reasoning and retrieval of archive resources (Yang, 2018). Researchers also built a big data management system for scientific research archives based on knowledge graph to realize the intelligent collection, archiving and processing (Lei, et al., 2020). In addition, some scholars have carried out researches on ontology construction, semantic organization and other related aspects. Duan and other scholars put forward the standardized construction method of archives description ontology, and also studied the concept modeling and instance extraction of the ontology (Duan, et al., 2018). Lv used semantic association technology to realize the cross-media aggregation of digital archives and promote the semantic construction and development of archive resources (Lv, 2015).

Oral archive refers to the various records of a particular social practice which failed to form after it was carried out due to some subjective and objective reasons, and were formed after the interview to the parties involved in the event, so as to reproduce the historical facts as true as possible (Wu, & Chu, 2012). Oral archives have the characteristics of individuality, subjectivity, richness and correction (Liu, & Xue, 2007). In recent years, more attention has been paid to the study of oral history in China. With the development of information technology and mass media technology, researches on the collection, development and utilization of oral archives in the digital age is helpful to rescue and protect the social history. As for the knowledge organization and management of oral archives, Liu first put forward that under the background of knowledge management, enterprises regard knowledge sharing and internalized of oral archives as the goal, to implement the externalization of tacit knowledge and

internalization of explicit knowledge (Liu, 2006). Hou established oral archive database, and used MARC to index oral archive metadata, to refine the granularity of oral archive organization (Hou, 2018). Pang studied the construction of oral archive platform and discussed the integration and development of oral archive resources (Pang, 2017). Nie explored a universally applicable metadata scheme of oral archives to assist the subsequent development, utilization and sharing (Nie, 2020). However, generally the existed researches on the knowledge organization and management of oral archives are still in their infancy.

On the whole, the existed studies rarely considered on the construction of archives knowledge organization models in the application of knowledge graph. Hence, it is urgent to carry out the systematic researches on the construction and application of knowledge graph of oral archives. This study builds a knowledge organization model of oral archives based on knowledge graph, and discusses the specific application scenarios, aiming to explore the innovative method of knowledge organization on oral archives and deeply organize and integrate massive archive resources.

Materials and Methods

In order to ensure the reliability of the knowledge organization model of oral archives, the research takes the relevant archive ontologies, thesauruses and specifications for reference. Besides, expert knowledge is also introduced to review the oral archive knowledge ontology. Taking the "Hong Kong Voices" oral archives database as an example, knowledge elements of oral archive are collected and the knowledge graph of oral archive is instantiated. In summary, the top-down approach is adopted to build the knowledge organization model of oral archive based on knowledge graph, as shown in Figure 1.
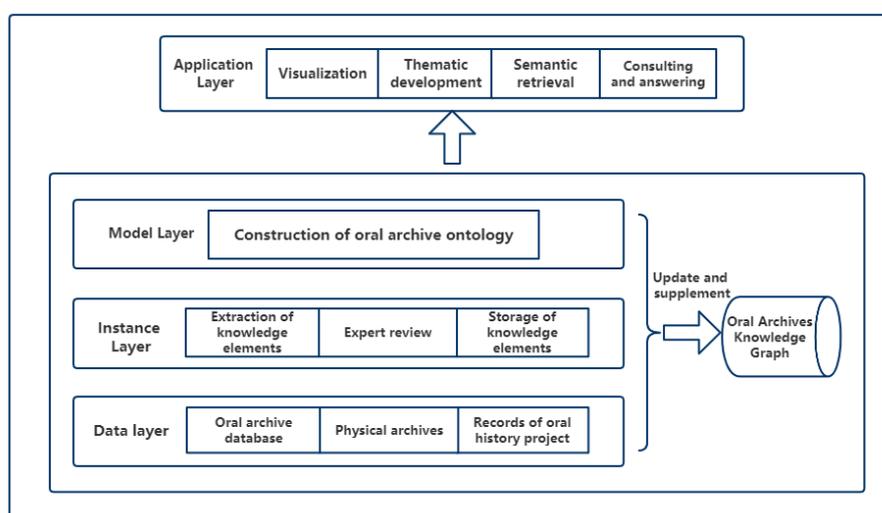
Figure 1. Knowledge organization model of oral archives based on knowledge graph

Firstly, structured, semi-structured or unstructured oral archives are collected from oral archive database, physical archives and oral history project, as the basis of knowledge elements extracted in the next step. Based on the existing archive ontologies, thesaurus and other specification, with the expert knowledge, the oral archive ontology is constructed as the model layer of knowledge graph. Besides, according to the oral archive ontology, entity extraction methods are adopted to extract the oral archive knowledge elements. Therefore the instance layer of the oral archive knowledge organization model is constructed to realize the instantiation of oral archive knowledge graph. Finally, according to the research and practice needs, the application layer is constructed to provide theoretical and practical references for the optimization of oral archive retrieval, thematic development and consulting service.

Model Layer Construction of Oral Archives Knowledge Graph

The subjectivity of oral archives determines that oral archives are easily affected by the factors such as personal memory of narrators, inaccurate record in the oral process and imprecise transcription of audio and video recording materials. Such factors may lead to the situation of inaccurate, incomplete or logical confusion and contradiction of archive content. Therefore, it is necessary to build up the standardized ontology model with strict structure as the model layer to ensure the controllability of the knowledge graph.

Model layer, also known as ontology layer, refers to the norms of concrete or abstract things and their relations in specific domain. Methods of ontology construction include Seven Step (Noy, & MC, 2001), TOVE (Tham, Fox, & Gruninger, 1994), Skeleton (Uschold, & Gruninger, 1996) and so on. Considering about the research needs and features of oral archives, this paper proposes a six-step method for domain ontology construction of oral archives. The specific steps are shown in Figure 2.

Figure 2. Oral archives ontology construction process

(1) Determine the nature and scope of oral archive ontology. Comparing with other types of archives, features and management process of oral archives have certain differences. In order to define properties, characteristics and ranges of oral archive ontology model, not only need to use the general legal norms and standards, such as Archives law of the People's Republic of China and Chinese Archives Subject Thesaurus, but also combine with professional reference books in the field of oral archives, such as Oral History in China, Study of Oral History, Everyone to Do the Oral History.

(2) Refer to relevant archive ontologies and oral archive metadata schemes. Qiu et al. (2014) used the mixed method to construct the archives ontology, which consists of author, journal, title of paper, institution, fund, year and volume and indexing keywords. Wang et al. (2015) established the ontology of archive science. The first layer is divided into three categories: representative figure, archive science and archives. Duan et al. (2018) put forward four categories of archive description ontology, including laws and regulations, documents, responsible persons and business. Zhou et al. (2009) established the ontology of e-government archives, which includes four subcategories: people, organization, document and event.

On the other hand, the Archives Bureau of Ningxia Hui Autonomous Region proposed the Metadata Standard for Digital Photographic Records, in which the entity types include archive entity, business entity, agency personnel entity and authorized entity. Specification for the Collection and Management of Oral History Materials issued by the State Archives Bureau sets relevant standards from the perspective of collection planning, collection process and historical data management of oral archives. Hence, it can be seen that there are still few ontology models specifically for oral archives at present. This study will learn from relevant ontologies and metadata scheme, and combine with the features of oral archives as well as the research and practice needs to construct the oral archives ontology model.

(3) Determine the core concepts of oral archives ontology. It is based on the relevant glossaries, terms and standards such as Chinese Archives Subject Thesaurus and Metadata Standard for Digital Photographic Records, and the ontologies widely used in the field of library and information science (Gao, & Huang, 2020), such as FOAF and Event. The core concepts reuse two FOAF ontology elements: "FOAF: People", "FOAF: Organization"; an Event ontology element: "Event: Event". According to Metadata Standard for Digital Photographic Records, three core concepts are customized, including "Oral history project",

"Oral archive" and "Archive business". The core concepts as well as the relationships among the core concepts are shown in Figure 3 below.

"Oral history project" refers to the oral history project of collecting and developing oral archives; "People" refers to the personnel involved in the process of collecting and developing oral archives and in the oral history project; "Organization" refers to the institutions involved in the collection and development of oral archives; "Event" is the description of event content recorded in oral archives; "Oral archive" covers the description of external and content features of oral archives. "Archive business" node is the description of a series of management business activities such as the formation, collection, archiving, management and utilization of archives.
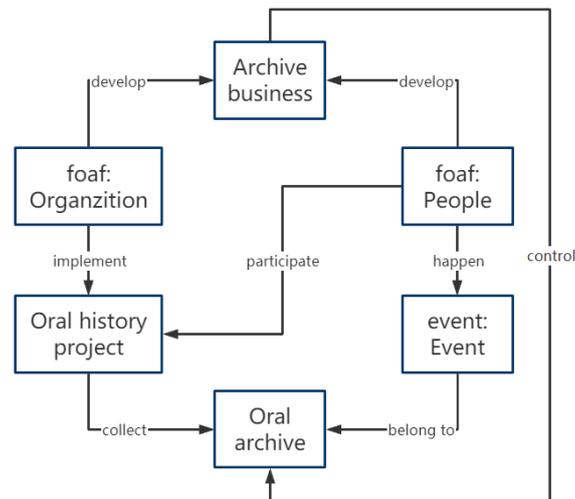
Figure 3. Core concepts of oral archives ontology

(4) Establish the concept structure of oral archives. Based on the identified core concepts of oral archives ontology, the sub-classes of each core concepts are gradually refined from top to bottom. The concepts at the same level are mutually exclusive, which cover the parent class as many ranges as possible. According to the research and practice needs of oral archives, "Oral history project" is divided into domestic and foreign oral project, and further divided into people, groups, history and social oral project according to the project theme. "People" concept is divided into narrator, interviewer, recorder, event related person and other people. "Organization" is divided into government, enterprise, private organization and other organizations. According to General Terminology for Archives Work (DA/T1-2000), oral archives can be divided into five types: paper archive, photo archive, audio archive, video archive and electronic archive. According to Metadata Standard for Digital Photographic Records, archives business can be divided into four types: affair, activity, function and joint function. Using protégé ontology development tool to edit oral archives ontology model, the concept structure is shown in Figure 4.
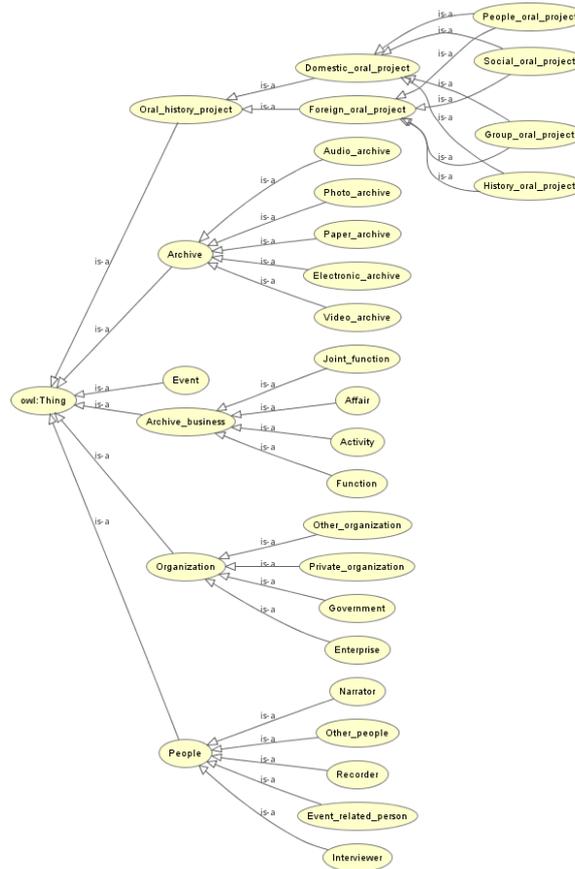
Figure 4. Concept structure of oral archives ontology

(5) Define the properties of concepts in oral archives ontology. On the basis of concept hierarchy, defines the properties of every concept, and its sub-classes can inherit the properties of the parent class. Concept properties include object properties and data properties. Object properties describe the relationships between concepts, such as the object properties of "Oral history project", including the "implement", which can associate with "Organization", and the "collect" associated with "Oral archive". Data properties describe the features of the concept, and the properties of "Oral history project" include "project name", "project time", and "project fund". As the subclass of "oral history project", both domestic oral project and foreign oral project can inherit the above properties. The specific data properties and object properties of oral file ontology model are shown in Table 1.

Table 1. Concept attributes of oral archives ontology

Data Property		Object Property		
Core concept	Properties	Properties	Domain	Range
People	People name	Participate	People	Oral history project
	Gender	Happen	People	Event
	Introduction	Implement	Organization	Oral history project
	Occupation	Belong to	Event	Oral archive
	Education level	Develop	People	Archive business
	Age	Control	Archive business	Oral archive
Organization	Organization name	Collect	Oral history project	Oral archive
	Address	Event element	Event	Element

Oral archive	Establish time
	Archive theme
	Archive name
	Archive date
	Copyright
	Archive kind
	Archive number
	Language
Oral history project	Project name
	Project fund
	Project time
Event	Event theme
	Event content
Archive Business	Business name
	Business people
	Business time

(6) Expert evaluation and inspection of oral archives ontology. Oral archives have the features of subjectivity. According to ontology construction principle (Karp, & Gruber, 1995), expert knowledge should be introduced to evaluate and test oral archives ontology to ensure the quality of ontology. The principles of ontology construction include clarity and integrity, which means a clear and complete definition of oral archives terms; consistency, which means the logical consistency between oral archives ontology concepts and knowledge reasoning, without contradiction or error; extensibility, that is, the ontology can be expanded with the development of theory, which is constantly improved and expanded. According to the construction principle and the performance of ontology model in practice, the oral archives knowledge ontology is gradually adjusted and improved.

Instance Layer Construction of Oral Archives Knowledge Graph

Based on the model layer which includes 6 core concepts, 20 subclass concepts and 32 concept properties, the instantiation of ontology model is studied, and the real knowledge elements are collected and linked to the ontology base. Taking "Hong Kong Voices" oral archives database as an example, this paper mainly collects elements from semi-structured and unstructured data, and constructs oral archives knowledge graph instance layer.

The interviewees of "Hong Kong Voices" oral project come from different ages, native places and industries. The interviews cover the topics including industry, education, community, culture, art, and people's livelihood (Hong Kong Voice, 2012). The copyright of the database belongs to the Hong Kong Memory Project. Hence, this paper only takes "Huang Lisong (Hong Kong Voice, 2010)" and "Liang Peizhen (Hong Kong Voice, 2010)" two pieces of oral archives as examples to elaborate the process and method of constructing the knowledge graph of oral archives, as shown in Figure 5.
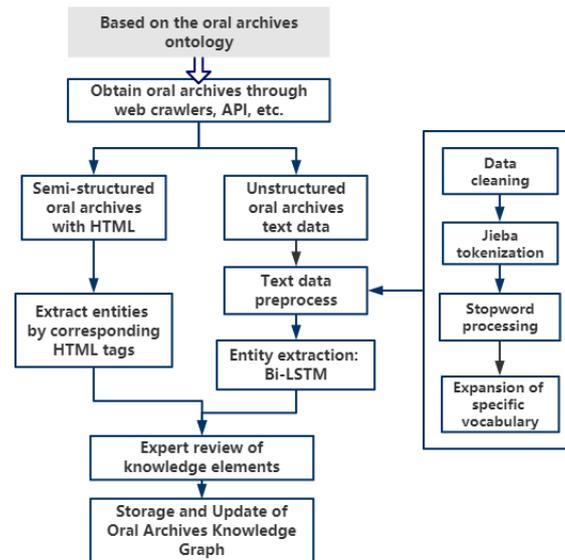
Figure 5. Process and method of constructing knowledge graph

The content of oral archives has strong subjectivity and low information density. This is because in the interview, the oral narrator will narrate with background introduction, mental activities and conversation related to the theme. Hence, important information which need to be collected as knowledge elements, is relatively sparse. Therefore, it is necessary to introduce expert knowledge, to check and screen if knowledge elements are closely related to the themes or not. Secondly, compared with other texts, oral archives are easily influenced by education level and expression ability, and the context coherence and semantic logic differ greatly among different narrators. In conclusion, after the knowledge extraction, still need to further clarify the context and logic to ensure the quality of relevance among knowledge elements.

(1) Knowledge element extraction of semi-structured data

The semi-structured data in the oral archives database of "Hong Kong Voices" refers to the archives information that has been indexed by HTML, such as person name, age, occupation, education level and so on. For this kind of semi-structured data with HTML tags, they can be extracted by using the HTML tags. Then transform the information into structured data, which is more helpful to the construction of oral archives knowledge graph. As shown in Figure 6, the profile information of "Huang Lisong" is mainly presented with hierarchical HTML tags, including <divclass><tr><td></td></TR> and so on. With the help of corresponding tags, the required elements such as "gender: male", "date of birth: 1920" (Hong Kong Voice, 2010) can be extracted and stored in a structured form like table, as the element resources of the instance layer. Using D2R mapping method, the data in relational database can be transformed into semantic data in the form of RDF triples and stored in the knowledge graph of oral archives.

```
▼<div class="TRS_Editor"> == $0
    "前香港大學校長 "
    <br>
    <br>
  ▼<table border="0" cellspacing="0" cellpadding="2" width="100%">
    ▼<tbody>
      ▼<tr>
          <td width="45%">性別:</td>
          <td width="55%">男</td>
        </tr>
      ▼<tr>
          <td>出生年:</td>
          <td>1920</td>
        </tr>
      ▼<tr>
          <td>受訪時年齡:</td>
          <td>90</td>
        </tr>
      ▼<tr>
          <td>教育程度:</td>
          <td>大學</td>
        </tr>
      ▼<tr>
          <td>主要職業:</td>
          <td>大學校長</td>
        </tr>
      ▼<tr>
          <td>訪談主題:</td>
          <td>教育,日治時期</td>
        </tr>
      </tbody>
    </table>
  </div>
</div>
```

Figure 6. HTML data of oral archives

Taking "Liang Peizhen (Hong Kong Voice, 2010)" in "Hong Kong Voices" database as an example, based on semi-structured data indexed with HTML, the properties, such as person name "Liang Peizhen", education level "primary school", occupation "retired cloth factory workers" are extracted to improve the properties of "People"-"Liang Peizhen" in the knowledge graph instance layer. Taking the "Oral archive"-"Family background and rural life before war" as an example, the archive name "Family background and rural life before the war", theme "social life", archive number "TW-LPC-LIFE-001", and the archive person "Liang Peizhen" are extracted from the HTML tag too, as shown in Figure 7. Through the "Event"-"Liang Peizhen's family background introduction", the relationship between "People"-"Liang Peizhen" and "Oral archive"-"family background and rural life before war" is established, which is helpful for the subsequent management and development of the related properties and other connections between the two entities.
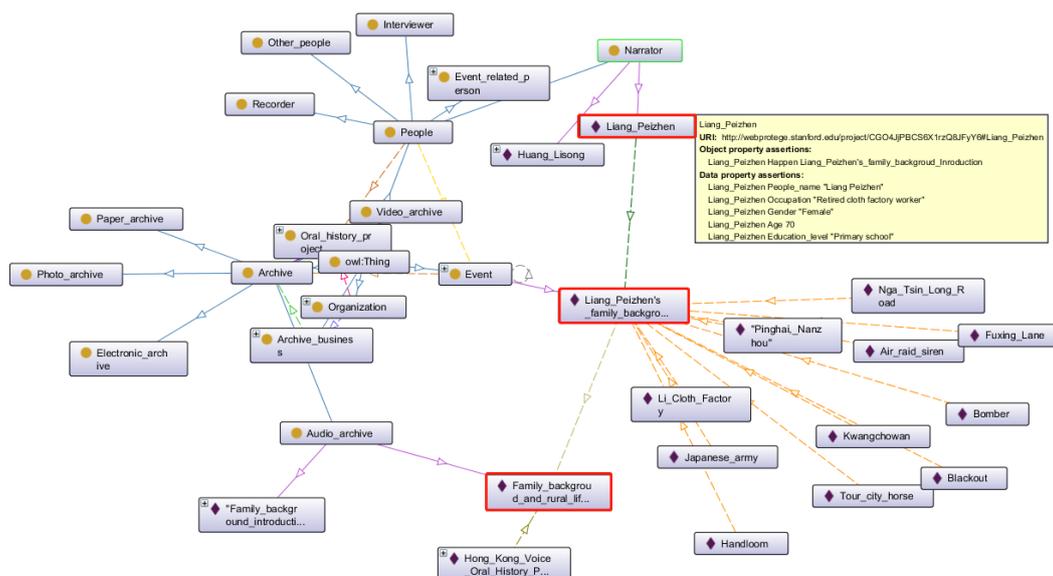
Figure 7. "People"-"Liang Peizhen" and "Event" elements

(2) Knowledge element extraction of unstructured data

As for unstructured oral archive data, it need to extract the knowledge elements through more complex procedures, and then store them in the knowledge graph. Knowledge elements extraction from unstructured data includes two sub tasks: entity recognition and relationship extraction. Entity recognition is to automatically identify named entities in oral archive, and relationship extraction is to extract the relationship between entities. At present, there are mainly three kinds of method of entity recognition, which include rule and dictionary based (Rau, 1991), statistical and machine learning based (Zheng, Li, & Tan, 2000) and deep learning based methods (Wu, Lv, & Yu, 2019).

This study mainly uses the method combining natural language processing and deep learning to extract knowledge elements. Based on the Python environment, Jieba and FoolNLTK toolkit, the transcription of audio-visual archives in "Hong Kong Voices" oral archives database is segmented, with stop word processing and oral archives specific vocabulary expansion. Bidirectional Long Short-Term Memory network (Bi-LSTM) which can effectively reduce the dependence on manually labeled data is used in knowledge element extraction. Bi-LSTM combines the LSTM model from the start of the text sequence and the LSTM model from the end of the text sequence to capture bidirectional semantic information. Thus, the problems of long-term dependence of recurrent neural network and the disappearance of gradient are solved. Each hidden layer of Bi-LSTM includes forget gate, update gate and output gate. The semantic information extracted from forward state and backward state are spliced together to get the final output of Bi-LSTM (Wu, Lv, & Yu, 2019).
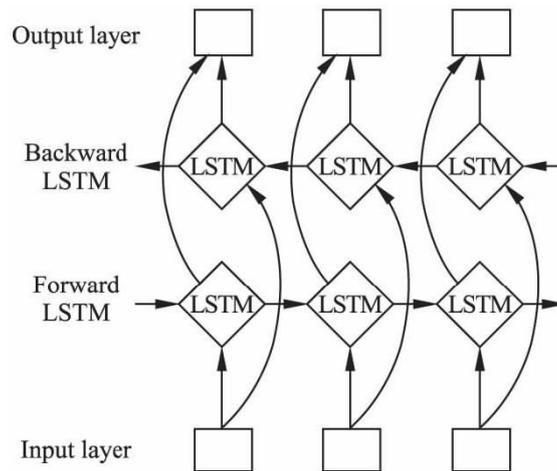
Figure 8. Bi-LSTM model structure

Specifically, it identifies the entities and relationships of oral archive contents without HTML tags in "Hong Kong Voices" database, and relates them to the concept of "Event" in the oral archives knowledge ontology. Taking "Oral archive"-"Family background introduction (Huang Lisong)" as an example, the extraction algorithm is used to collect "Event" knowledge elements, including "Huang Yingran", "Shantou", "England", "Shanghai St. John's University", "Shantou Dishi middle school", "vice principal", "Xiamen Normal University", etc. On the other hand, in order to ensure the quality and relevance of extracted knowledge elements, it is necessary to introduce domain expert knowledge to further check, screen and associate the knowledge elements. By linking and comparing the extracted knowledge elements with external knowledge and existing historical materials, such as relevant archives or historical evidences, the knowledge elements which are not consistent with actual situations can be found. Therefore, after examination and screening, the knowledge elements that do not have specific meaning or do not meet the needs of the extraction, such as "Hong Kong" and "people's livelihood", are screened out. After extraction, examination and screening, the "Event" example of "Huang Lisong's family background introduction" is shown in Figure 9.
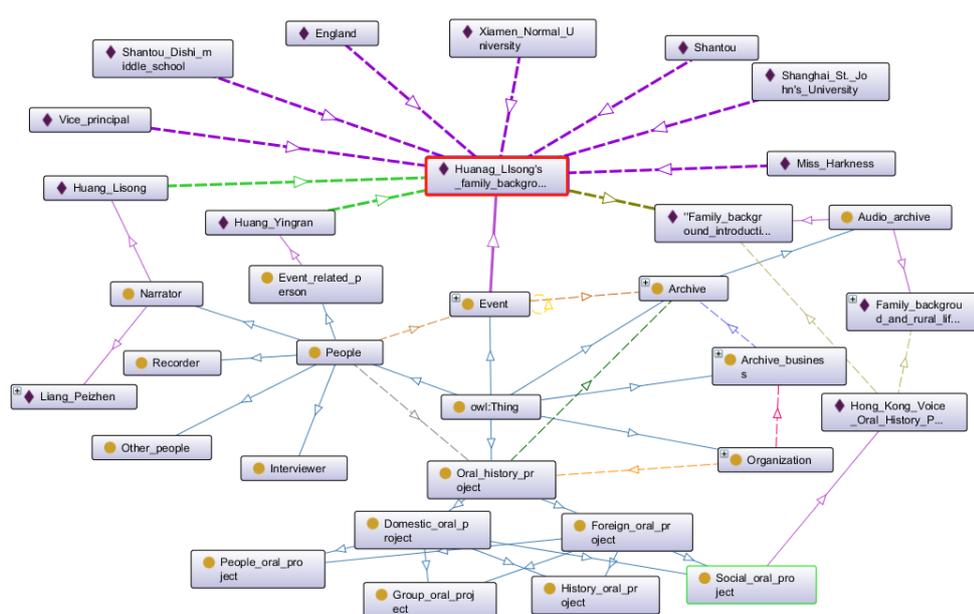

Figure 9. "Event" example of "Huang Lisong's family background introduction"

To sum up, based on the standard set by model layer of oral archives knowledge graph, knowledge elements are extracted from semi-structured and unstructured archive data, and the instance layer is constructed. On one hand, knowledge elements in semi-structured oral archives is obtained by parsing HTML tags. On the other hand, the extraction algorithm is adopted to extract the "Event" knowledge elements from unstructured oral archives, so as to transform the semi-structured and unstructured archives knowledge into hierarchical knowledge elements. After knowledge element extraction, in order to ensure the consistency and accuracy of the content in knowledge graph, entity alignment is carried out to fuse knowledge elements from different sources. In order to ensure the originality, scientificity and reliability of oral archives knowledge elements, this study mainly adopts the method of expert examination, according to the standard of oral archives ontology model layer. Finally, the reviewed and screened knowledge elements of oral archives are stored in Neo4j, protégé or other graph databases and ontology tools for dynamic maintenance and updating, which is convenient for knowledge retrieval and query. So far, the oral archives knowledge elements are collected and linked to the ontology model layer, and the oral archives knowledge graph is instantiated and constructed.

Application Scenarios of Oral Archives Knowledge Graph

With the impact of the digital environment, the traditional appointment access and other archive information services and utilization methods are difficult to meet the diversified needs of users. Therefore, considering the practical work and research needs of oral archives, knowledge graph is used to realize the intelligent utilization of oral archives, including the visualization of archives resources, thematic development of archives, semantic retrieval of archives, and consulting service of archives.

The application methods of oral archives knowledge graph can mainly be divided into two categories. One is directly regarding knowledge graph as the result of visualization, which helps to transform the massive oral archives text into more understandable images. The other is based on the knowledge graph embedding methods, which embed entities and relationships into the continuous space vector, and uses the scoring function to measure the similarity of each fact triplet (Wang, et al., 2017). The models of knowledge graph embedding include distance based methods, such as TransE (Bordes, et al., 2013) and TransH (Wang, et al., 2014), and similarity based methods, such as RESCAL (Nickel, Tresp, & Kriegel, 2011) and DisMult (Yang, et al., 2014). The knowledge graph is embedded in the process of task training as vectors to realize the application of relationship extraction, question answering system and recommendation system.

(1) Visualization of archival resources

Oral archive knowledge graph represents the knowledge entities and their relationships in the form of dots and lines, which reveals the internal relationship between the knowledge elements of oral archives. It also helps users understand and analyze the contents of oral archives, and finds the deeper semantic relationship of archives resources, which provides the basis for the future organization and development. Taking the "People"-"Liang Peizhen" as an instance, it not only contains person's properties, such as age, occupation, education level and so on, but also presents the relationship between the "Event"-"Liang Peizhen's family background introduction", "Oral archives"-"family background and rural life before war". This will help to relate people with their archives and events more clearly, and form the knowledge network of oral archives, which is convenient for users to discover the internal relationship of different knowledge elements.
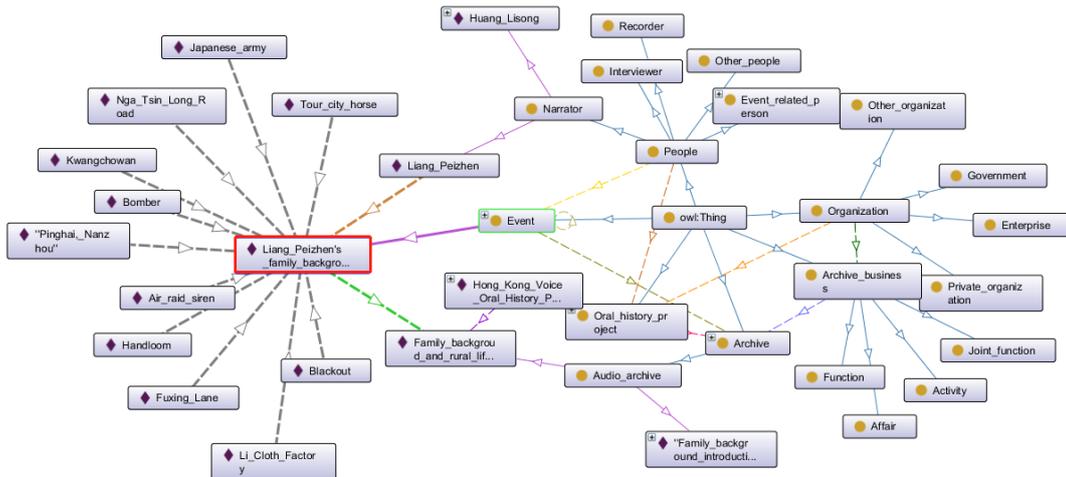
Figure 10. Example of oral archives knowledge graph visualization

(2) Thematic development of archives

Based on the content associations revealed by the triplets in knowledge graph, the thematic development of archival projects are carried out by integrating special topics, which helps to avoid the islands of knowledge. On the other hand, it also reduces the repeated development and waste of oral archive resources, as well as providing archive users with richer archive knowledge and more complete archive services. For example, Hong Kong place names are involved in the "Event" concept, such as "Jordan Road", "Sha Tsui Road", "central", etc. Based on the association between the elements of place names that co-exist in the archives, the elements with internal semantic association is linked. Hence, a number of independent archives are collected to form the thematic content based on place names, which enriches the viewpoints of archive resources development.
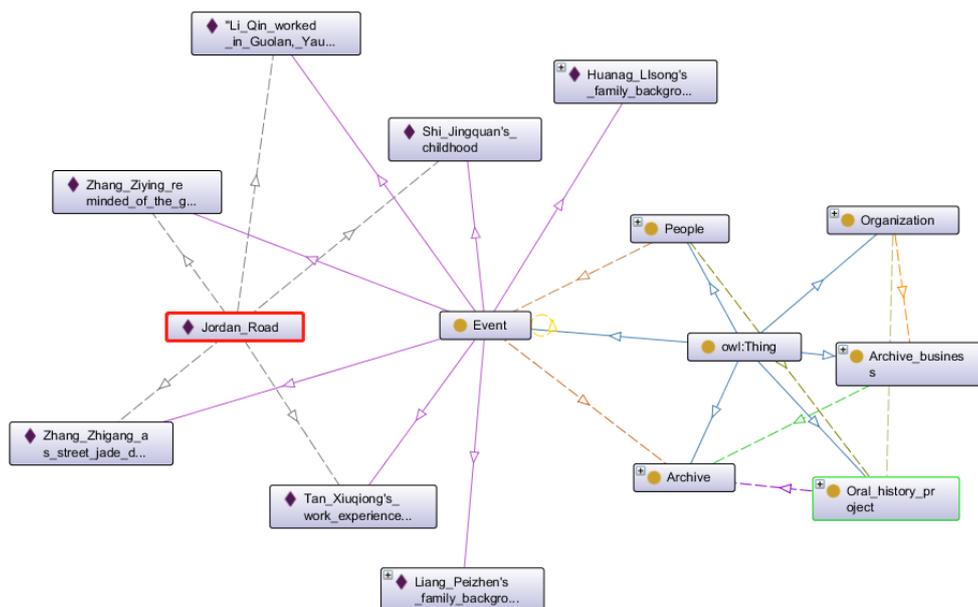

Figure 11. Event element "Jordan Road" and its related events

(3) Optimization of archive retrieval service

Oral archives knowledge graph links the information of archives from the semantic perspective, so that the retrieval service is no longer limited to the traditional keyword matching methods. Based on the clear structure and semantic relationship of knowledge graph, computers can automatically understand the semantic information contained in the user retrieval queries, and retrieve the relevant knowledge elements through semantic association. According to the retrieval queries and corresponding related knowledge elements, the retrieval results are sorted intelligently by relevance, and contents related to the retrieval subject are also recommended automatically. On the other hand, based on the oral archive ontology model, retrieval system can provide multiple retrieval entrances, such as "People", "Organization", "Oral archive" and so on. Receiving the retrieval queries, the system will analyze and infer them by semantic interpretation and reasoning, and match them with the concepts and instances by means of machine learning. If the matching similarity exceeds the threshold, the relevant concepts are added to the retrieval queries to enrich the results, and the results and new queries are also provided for users' reference.

(4) Auxiliary of archive consulting service

It is difficult to clearly describe the required archives, especially when users are not familiar with the contents of archives, which has become the dilemma of optimizing traditional archives service. With the help of semantic association of knowledge graph, it will be easier to analyze the semantic interpretation of users' expression, and refine the users' requirements for oral archives, helping to realize more mature and intelligent oral archive services. For example, taking the place element "Jordan Road" as the center word, the relationship between different instances of "People" in the archives is built up. With the help of semantic analysis and reasoning ability of knowledge graph, the automatic consulting system of archives is constructed. The interactive service of human-computer question and answer is provided, and the efficiency of archives automation service can be realized. Combined with the methods of knowledge graph embedding, oral archives knowledge graph is transformed into feature vectors, helping the automatic question answering. Thus, the service of inquiry and answer are transformed into a task of searching and answering a given question. Based on the triples in the knowledge graph, the question is associated with the answer, which will be returned to users.

Conclusion

Oral archives, as historical memory and research resources, have great potential of development in the digital humanities background. However, there are few researches focused on in-depth development or knowledge mining of oral archives, which is not conducive to the intelligent management and service of oral archives. Hence, this paper discusses the construction of knowledge organization model of oral archives based on the method of knowledge graph, considering the semantic relation of knowledge elements. The knowledge organization model of oral archives based on knowledge graph has the following four layers: the basic one is the data layer composed of oral archive database, physical archives and records of oral history project, and then following the model layer constructed by the knowledge ontology of oral archives. The instance layer composed of knowledge elements of oral archives extracted based on the data layer and model layer. Finally the application layer of knowledge organization model focuses on the application scenarios of oral archives. Considering the subjectivity of oral archives, this study combines standardized ontology structure, intelligent knowledge extraction and expert examination to ensure the normalization and controllability of oral archives knowledge graph.

In the digital environment, requirements of archives knowledge organization are constantly improving, and the public demands of archive usage also presents a diversified trend.

The knowledge organization model of oral archives, to provide effective organization and management based on digital technology, can help to meet the knowledge acquisition needs and supply knowledge-based and intelligent services. In the future, the research will further enrich and optimize the ontology structure and instances of oral archives knowledge graph, to improve the scientificity and coverage of oral archives knowledge graph. Besides, the knowledge representation and application of oral archives knowledge graph also need to be improved, which help to transform knowledge elements in oral archives into semantic relevance and knowledge retrieval resource base.

References
Bordes, A., Usunier, N., Garcia-Duran, A., et al. (2013). Translating embeddings for modeling multi-relational data. *Advances in Neural Information Processing Systems*, Lake Tahoe, Nevada, USA, 2787-2795.
Duan, R., Ma, Y., & Li, Z. (2018). Study on the Standardization of Archival Description Ontology's Construction. *Archives Science Study*, *2*, 63-71.
Gao, C., & Huang, X. (2020). Knowledge Graph Construction and Visualization of Regional Government Microblog. *Modern Information*, *40*(12), 90-99+113.
Guo, X. (2019). *Research on Archives Knowledge Graph Construction Technology*. Master thesis, China Academic of Electronics and Information Technology.
(Note: According to the reference format provided by the Hong Kong Voices database)Family Background and Rural Life Before War, Liang Peizhen, 01/03/2010, Hong Kong Voices Oral History Archives, Hong Kong Memory website: www.hkmemory.hk.
Family Background, Huang Lisong, 19/03/2010, Hong Kong Voices Oral History Archives, Hong Kong Memory website: www.hkmemory.hk.
Hong Kong Voices Introduction, Hong Kong Voices, 2012, Hong Kong Voices Oral History Archives, Hong Kong Memory website: www.hkmemory.hk.
Hou, T. (2018). Highlights and Enlightenment of Construction of Veterans' Oral Archives Database in American Folklore Center. *China Archives*, *9*, 74-75.
Karp, P. D. & Gruber, T. R. (1995). A Generic Knowledge-base Access Protocol. *Proceedings of the International Joint Conferences on Artificial Intelligence, Montreal. An ontology-based knowledge sharing API for AI people*.
Lei, J., Li, S., Zhao, R., et al. (2020). Research on the Construction and Application of Knowledge Graph Oriented to Scientific Research Archives Management. *Digital Library Forum*, *5*, 8-15.
Lei, J., Zhao, R., Li, S., et al. (2020). Construction of Knowledge Graph for Scientific Research Archives Big Data Management System. *Digital Library Forum*, *2*, 19-27.
Li, J. (2013). Discussion on Knowledge Organization and Service Mode of Archival Information Resources. *Archives Management, 1,* 49-50.
Liu, J. (2006). Management of Enterprise Oral Archives in Background of Information Management. *Archives of Shanxi*, *4*, 21-23.
Liu, X., & Xue, H. (2007). Discussion of the Value of Oral Archives. *Archives Science Bulletin*, *4*, 88-91.
Lv, Y. (2012). The Knowledge Organization of Digital Archival Resources Based on Linking Data. *Archives Science Study*, *6*, 44-48.

Lv, Y. (2015). Research on the Strategies of Realizing Cross-media Semantic Correlation Integration of Digital Archives Resources. *Archives Science Study*, *5*, 60-65.

Mao, T. (2015). Research on Construction of Archives Base Based on Knowledge Organization. *Shandong Archives*, *2*, 16-18.

Nickel, M., Tresp, V., & Kriegel, H. P. (2011). A three-way model for collective learning on multi-relational data. *International Conference on Machine Learning*, Washington, USA, 809-816.

Nie, Y. & Li, R. (2020). A Metadata Schema for Oral Archives Based on Dublin Core. *Archives Science Study, 34*(3), 129-136.

Noy, N. F., & Mc Guinness, D. L. (2001). Ontology development 101: A guide to creating your first ontology. *Knowledge Systems Laboratory*, *32*(1).

Pan, Y. (2012). Understanding and Action: Discussion on Construction of Oral History Archives resources. *Archives Science Bulletin, 1*, 101-104.

Pang, X. (2017). *Research on the Construction of Oral History Archives Platform in China*. Master thesis, Wuhan University.

Pujara, J., Miao, H., Getoor, L., et al. (2013). Knowledge Graph Identification. *International Semantic Web Conference*. Springer, Berlin, Heidelberg.

Qiu, J., Yang, Q, & Lou, W. (2014). Theory and Empirical Study of Resource Ontology Construction. *Information Studies: Theory & Application*, *37*(5), 1-6.

Rau, L. F. (1991). Extracting company names from text. Artificial Intelligence Applications. *Proceedings. Seventh IEEE Conference on*. IEEE.

Shu, Z. (2020). Research on the construction of archival knowledge graph under the context of digital humanities. *Archives of Shanxi*, *2*, 53-60.

Sun, M., Fang, X., & Chen, X. (2020). Research on the Construction of Knowledge Graph of Celebrity Archives from the Perspective of Digital Humanities—Taking the Archives of Writers as an Example. *Archives of Shanxi*, *6*, 79-88.

Tham, K. D., Fox, M. S., & Gruninger, M. (1994). A cost ontology for enterprise modeling department of industrial engineering. *Proceedings of third Workshop on Enabling Technologies: Infrastructure for Collaborative Enterprises, Morgantown, WVWorkshop on Enabling Technologies*, 111-117.

Uschold, M., & Gruninger, M. (1996). Ontologies: principles, methods and applications. *Knowledge Engineering Review*, *11*(2), 14-17.

Wang, L. (2008). Discussion on Archives Information Organization and Searching Development under the Environment of Knowledge Integration. *Archives Science Study*, *5*, 45-50.

Wang, Q., Mao, Z., Wang, B., et al. (2017). Knowledge Graph Embedding: A Survey of Approaches and Applications. *IEEE Transactions on Knowledge and Data Engineering*, *29*(12), 2724-2743.

Wang, Y. (2008). Discussion on Organization of Archival Knowledge. *Archives Science Bulletin, 2*, 23-27.

Wang, Y., Lv, Y., & Nie, L. (2015). Discussion on Construction of Archival Ontology. *Archives Science Bulletin*, *6*, 19-25.

Wang, Z., Zhang, J., Feng, J., et al. (2014). Knowledge graph embedding by translating on hyperplanes. *Proceedings of the Twenty-Eighth AAAI Conference on Artificial Intelligence*, Québec, Canada, 1112-1119.

Wu, H., Lv, L., & Yu, B. (2019). Chinese Named Entity Recognition Based on Transfer Learning and BiLSTM-CRF. *Journal of Chinese Computer Systems*, *40*(6), 1142-1147.

Wu, P., & Chu, L. (2012). Theoretical Basis of Archival Protection of Intangible Cultural Heritage. *Archives Science Bulletin*, *5*, 75-77.

Yang, B., Yih, W., He, X., et al. (2014). Embedding entities and relations for learning and inference in knowledge bases. *arXiv preprint* arXiv:1412.6575.

Yang, Q. (2018). Research on data mining and intelligent utilization of electronic archives of China Unicom. *Archives Science Study*, *6*, 105-109.

Zheng, J., Li, X., & Tan, H. (2000). Research of Chinese Names Recognition Method Based on Corpus. *Journal of Chinese Information Processing*, *1*, 7-12.

Zhou, Y., & Dong, H. (2009). Construction of Digital Archives Ontology in E-government Domain. *Library and Information Service*, *53*(19), 112-116.