Postprint: Research on Knowledge Graph and Multidimensional Knowledge Discovery for Oral History Archival Resources

Authors: Deng Jun, Wang Ruan

Date: 2023-04-01T00:00:00+00:00

Abstract

[Objective/Significance] Knowledge discovery in oral history archival resources based on knowledge graphs represents a novel attempt at knowledge discovery in the digital humanities domain, offering new avenues for fine-grained resource association, semantic querying, and personalized exploration. [Method/Process] Utilizing data from the Oral History Data Center for Veterans of the War of Resistance Against Japanese Aggression at Nanjing Normal University as the source, a knowledge graph of oral history archival resources for veterans of the War of Resistance was constructed. Based on this graph instance, multi-dimensional knowledge discovery research was conducted across multiple dimensions, including overall project overview, event-theme relationships, social network relationships, and spatio-temporal network relationships. [Results/Conclusion] Digital humanities technical methods exemplified by knowledge graphs provide robust tool support for knowledge discovery research and inject new impetus into the deep development of humanities resources.

Full Text

Research on Knowledge Graph and Multidimensional Knowledge Discovery of Oral History Archive Resources

Deng Jun, Wang Ruan
School of Business and Management, Jilin University, Changchun 130012

Abstract:
[**Purpose/Significance**] Knowledge discovery of oral history archive resources based on knowledge graphs represents a new attempt at knowledge discovery in the digital humanities field, providing new pathways for fine-grained resource association, semantic querying, and personalized exploration.
[**Method/Process**] Using data from the Anti-Japanese War Veterans Oral

Data Center of Nanjing Normal University as the data source, we constructed a knowledge graph of anti-Japanese war veterans' oral history archive resources. Based on graph examples, we conducted multidimensional knowledge discovery research from the perspectives of overall project overview, event-theme relationships, social network relationships, and spatiotemporal network relationships. **[Result/Conclusion]** Digital humanities technology methods, represented by knowledge graphs, provide powerful tool support for knowledge discovery research and inject new kinetic energy into the deep development of humanities resources.

Keywords: knowledge graph; oral history archives; knowledge discovery
Classification: G270
DOI: 10.13266/j.issn.0252-3116.2022.07.001

In the digital era, technological innovation has profoundly impacted human culture, economy, and various other domains. In this context, the deep interaction between knowledge organization of humanities resources and digital technology has spawned new models of knowledge production and propelled cross-disciplinary integration. Guided by national policy, the "new liberal arts" initiative aims to break down barriers in traditional humanities and social science disciplines and construct an interdisciplinary, multi-faceted system. Digital humanities emphasizes the integration of digital technology and humanities research, leveraging new technologies and tools to achieve novel interdisciplinary fusion, broaden research perspectives, and enrich knowledge content. Consequently, both new liberal arts and digital humanities advocate breaking disciplinary divisions and enabling cross-disciplinary, cross-domain collaborative cooperation.

Under traditional knowledge discovery models, oral history archive resources appear scattered across networks. Users struggle to integrate complex resources, forced to repeatedly click and frequently search through manual browsing, which is time-consuming and labor-intensive. This prevents rapid capture, association, and discovery, hindering effective value realization of resources. In digital environments, digital humanities technologies, particularly knowledge graphs, can break through data silos of oral history archives, reshape information stacks, achieve resource aggregation, and facilitate the transition from "digitalization" to "datafication" to "intelligentization," presenting a progressive process of data reorganization, information association, and knowledge discovery.

Therefore, under the new liberal arts construction background, this paper approaches humanities research from a digital humanities perspective, selecting knowledge graphs as a representative digital humanities knowledge mining method. With the Anti-Japanese War as the thematic backdrop, we introduce oral history archive resources of anti-Japanese war veterans to construct a knowledge graph example. Through visualization, we deeply associate and mine potential knowledge, empowering humanities with technology, exploring the compatibility of "digital technology" and "humanities research," extending traditional research methods, breaking barriers between technology and hu-

manities resources, and providing new pathways for revealing content features, deep semantic association, and multidimensional knowledge discovery.

2 Literature Review

Current knowledge discovery research primarily focuses on four dimensions: basic theory, technical exploration, method application, and research trends.

At the theoretical level, knowledge discovery revolves around conceptual connotations, processes, and steps. Regarding definitions, the widely accepted concept refers to the nontrivial process of identifying valid, novel, potentially useful, and ultimately understandable patterns from datasets [3]. Regarding process steps, Ma Li et al. [4] decomposed it into six steps: data cleaning, data integration, data transformation, data mining, pattern evaluation, and knowledge representation. Jin Xiao'en summarized the process into four layers: data layer, tool layer, mining layer, and application layer [5].

At the technical level, clustering analysis, classification analysis, machine learning, and neural networks are the main methods. In application domains, scholars have focused on constructing knowledge discovery platform systems, analyzing system functions, and conducting specific case studies to explore how to better achieve domain knowledge discovery.

In the library and information science field, scholars such as Zeng Jianxun, Sun Yu, Liu Aiqin, and Lu Bing have conducted relevant research on domain knowledge discovery systems and functions. Song Xueyan et al. [6], from a digital humanities perspective, developed visual graphs of character relationships, regional hotspots, and emotional tendencies using the diaries of Wang Shijie as an example, providing pathways for content analysis and knowledge discovery. In the archives field, Sun Minglei et al. [7] elaborated on the construction process of knowledge graphs for celebrity archives, offering new ideas for celebrity archive development. Yang Xixi [8] summarized and proposed path models and implementation methods for organizing and developing historical archive resources from a digital humanities perspective, further deepening the value realization of historical archive resource knowledge discovery.

The concept of oral history archives was introduced to China in the 1980s, with researchers extensively discussing its connotations and extensions, attempting to demonstrate its legitimacy as archives from aspects of essential characteristics, formation processes, and legal recognition [9]. With the advancement of archive digitization, scholars have focused on practical experiences from foreign oral history archive construction to provide references for developing distinctive Chinese paths. Regarding knowledge discovery research on archive resources, Zhu Lingjun [10] analyzed the basic procedures of archive knowledge discovery driven by data, constructing a content framework for archive knowledge discovery from the data layer, logic layer, application layer, and presentation layer. Gao Chenxiang [11] proposed a knowledge discovery model for regional government microblogs from an archival science perspective, aiming to develop

government information resources with document and archive attributes. H.F. Yu et al. [12] achieved knowledge discovery of archived files based on the Internet Cache Archive System (ICAS). M.C. Pattuelli et al. [13] discussed the generation, processing, and integration processes of linked data for cultural heritage, using oral history archives as the primary source for linked data named entities, describing how the data development process provides new pathways for research queries and engagement with heritage data.

In summary, in the knowledge discovery field, domestic and international efforts have not yet formed a cohesive whole, especially in the oral history archive domain where exploration remains nascent. Therefore, anchored and mined association of oral history archive resources has become a new breakthrough point for knowledge discovery. As knowledge discovery research deepens, it is also developing toward digitalization and intelligence, which will undoubtedly become an important driving force and inspiration point for knowledge discovery in oral history archive resources.

3 Multidimensional Knowledge Discovery Path for Oral History Archive Resources

The excavation of oral history archive resources utilizes digital technology for information extraction, making the mining of themes, characters, events, and background information particularly important [15]. Knowledge graphs reveal semantic information through graph structures, enabling users to capture, associate, and discover knowledge through simple "online information," providing possible pathways for knowledge discovery of oral history archive resources.

This paper uses data from Nanjing Normal University's Anti-Japanese War Veterans Oral Data Center (comprising 1,501 oral history projects from the Kuomintang, Eighth Route Army, and New Fourth Army) as the experimental data source. We adopt a top-down approach to construct the knowledge graph of anti-Japanese war veterans' oral history archive resources: first constructing the schema layer; second, organizing the data layer under the guidance of the schema layer structure; and finally importing the processed data into the Neo4j graph database for knowledge graph visualization, laying the foundation for ultimately achieving multidimensional knowledge discovery of oral history archive resources, as shown in Figure 1 [Figure 1: see original paper].

4 Knowledge Graph of Anti-Japanese War Veterans Oral History Archive Resources

4.1 Schema Layer of the Knowledge Graph

The schema layer serves as the pattern framework for data collection. Based on the data source, we independently extracted required elements to architect the schema layer for the anti-Japanese war veterans' oral history archive resources knowledge graph.

To ensure the integrity and systematicity of the schema layer organization, we adopted a taxonomic perspective, starting from the external features, content features, and formal features of oral history archives to 梳理 class and hierarchical relationships. We established 11 major classes: person, role, identity, event, work experience, time, place, military establishment, project, rights, and device, with relevant subclasses under each major class, as shown in Figure 2 [Figure 2: see original paper].

Attributes enrich the description of classes, typically comprising three parts: attribute name, domain, and range. The attribute name is an abstract noun describing the attribute, while the domain and range refer to the class and data type the attribute points to, respectively. The attribute information for anti-Japanese war veterans' oral history archive resources includes 26 attributes. Specifically, it covers seven person-class attributes: language, former name, person name, ethnicity, nationality, political party, and gender; two unit-class attributes: former unit name and unit name; two military attributes: military position and subordinate department; one position-class attribute: position name; one place-class attribute: place name; one military establishment-class attribute: designation (military); one event-class attribute: event name; nine project-class attributes: abstract, recording duration, release duration, resource size, resource format, resource type, title, link, and page bottom source; one identity-class attribute: domain identity recognition; and one rights-class attribute: copyright ownership.

4.2 Data Layer of the Knowledge Graph

The data layer organization for the anti-Japanese war veterans' oral history archive resources knowledge graph includes four components: entity extraction, relation extraction, attribute extraction, and knowledge fusion.

4.2.1 Entity Extraction Entity extraction involves identifying contextual, lexical, and formal features of metadata to classify segmented word entities into predetermined categories. Based on the experimental data source, we extracted data entities conforming to the 11 major classes and their subclasses, including person, role, time, work experience, etc. The device class was not reflected in the data source, so this major class currently has no data. Event entities exhibited naming inconsistencies, where the same historical event had multiple names. For example, the "July 7th Incident" that opened the prelude to the national anti-Japanese war was expressed in the data source as "Lugou Bridge Incident," "Lugou Bridge Campaign," etc.; while the "August 13th Incident" had names such as "Defending Greater Shanghai," "Second Songhu War," and "Songhu Battle." Therefore, it is necessary to use encyclopedia websites and dictionaries as references to unify event entity names.

4.2.2 Relation Extraction Relation extraction involves extracting relation instances between the aforementioned class entities from the experimental data

source, including information about the two entities linked by the relation and the relation type. Taking the person class as an example, the extracted relation types are shown in Table 1 .

4.2.3 Attribute Extraction Attributes help make entities more substantial and three-dimensional. Since the same person, event, time, place, and other entities may exist in different oral history projects, and attribute information may also differ, attribute comparison and integration are required during attribute extraction. When extracting person attribute information, this paper only focused on the correspondence between person ID and attribute information, that is, after the initial extraction was completed, duplicate names were screened using person names for information comparison and attribute integration.

4.2.4 Knowledge Fusion After preliminary data screening and information extraction, a total of 103,170 data entries were obtained. Next, duplicate and conflicting information needed to be eliminated. The knowledge fusion issues in this experimental data source first manifested in data format, namely inconsistent time formats. For example, different date representations such as "2019 年 09 月 28 日," "1937/7/7," and "1940-2-13" were unified into the "year-month-day" format. Second, inconsistent data content appeared, such as different names for the same person. For instance, both "周恩来" and "周总理" refer to the person entity "周恩来." This phenomenon can be resolved through encyclopedia knowledge graph comparison and matching for name disambiguation. Finally, the data was batch-imported, resulting in 25 entity node files, 64 relation files, 30,975 nodes, and 64,037 relation pairs. This paper selected the Neo4j graph database to construct the anti-Japanese war veterans' oral history archive resources knowledge graph and used the Browser tool to achieve knowledge graph visualization. Due to interface display limitations, Figure 3 [Figure 3: see original paper] only shows a partial visualization effect of the knowledge graph data.

5 Multidimensional Knowledge Discovery of Oral History Archive Resources Based on Knowledge Graph

5.1 Knowledge Discovery Based on Project Overview

As oral history archive resources mostly rely on project practice, knowledge discovery based on project overview is indispensable. Taking projects as the main thread allows both macro-level "sketching" of overall-partial knowledge graphs and micro-level "glimpsing" of project-location knowledge graphs, providing pathways for deeply presenting project knowledge discovery processes.

To make semantic linking information between entities richer and more intuitive, we conducted a knowledge graph overview of 1,501 oral history projects. This revealed that project webpage release times concentrated in 2016 and 2017, while project collection time spans were larger, ranging from the 1960s-1970s to 2015, scattered across various years but concentrated in 2015, reflecting compre-

hensive rescue and release of numerous early scattered oral history archive materials from diverse sources. Meanwhile, after 2018, oral history projects basically achieved same-year collection and release, indicating improved overall project efficiency. Project collection locations covered 20 provincial-level administrative regions, encompassing North China (Beijing, Tianjin, Hebei), Northeast China (Liaoning), East China (Shanghai, Jiangsu, Zhejiang, Fujian, Jiangxi), Central China (Henan, Hubei, Hunan), South China (Guangdong, Guangxi Zhuang Autonomous Region), Southwest China (Sichuan, Guizhou, Yunnan), and Northwest China (Gansu, Shaanxi, Ningxia Hui Autonomous Region).

We separately retrieved the oral history project "Wu Dai: The Bond Between Soldiers and Civilians on Both Sides of the Mu River." Through knowledge graph "connections," the following information was expressed: copyright belongs to the Anti-Japanese War Research Center of Nanjing Normal University, project collection time was 1987, collection location was Beijing, and the location's region was Beijing-China. Further clicking on the entity, Figure 4 [Figure 4: see original paper] displays attribute information that primarily summarizes and extracts the external and formal features of the oral history archive resource project. For example, the project title is "Wu Dai: The Bond Between Soldiers and Civilians on Both Sides of the Mu River," ID number is "36175," URL link is http://lb.njnu.edu.cn/information/262/6477, page bottom source is "Eighth Route Army Taihang Memorial Hall," resource size is 21.22KB, resource format is text/html, and resource type is text and image.

Notably, users can click the URL link to jump from the knowledge graph visualization portal to the source webpage of the oral history project. This functionality undoubtedly brings great benefits to user query and retrieval. On one hand, citing project webpage data provides further verification of the authenticity and reliability of the knowledge graph data, allowing users to provide feedback, corrections, and updates to the knowledge graph by referring to the original text. On the other hand, users can trace more detailed first-person perspective oral information through URL links, providing powerful supplementary materials for knowledge chain presentation and complete depiction of multiple entities.

5.2 Knowledge Discovery Based on Event-Theme Relationships

Events, as important components of oral history archive resources, typically involve information about persons, time, and place, which can concisely summarize event outlines. Since the data source in this paper comprises anti-Japanese war veterans' oral history archive resources, knowledge discovery based on event-theme relationships mostly concerns anti-Japanese war historical events. To explore internal event connections and excavate hidden event information, we used event entities as the center to retrieve event-project relationships, event-time relationships, and event-place relationships.

5.2.1 Event-Project Relationship Extracting the event-project relationship knowledge graph and displaying relevant instances in fine granularity below

the graph, as shown in Figure 5 [Figure 5: see original paper]. From the peripheral region, the project-event knowledge graph mostly presents one-to-one relationships (one oral history project corresponds to one event, e.g., the project "Xu Shichang: The Centennial Ups and Downs of a Huangpu Anti-Japanese War Veteran" only mentions the "Zhejiang-Jiangxi Campaign" event) and one-to-many relationships (one oral history project corresponds to multiple events or one event exists in multiple oral history projects, e.g., the project "Chang Duanjian: The First Gun Snatched from the Enemy" includes the "Mayichuan Campaign," "Dongdi Campaign," "Xingshugang Campaign," and "Fanmazhai Campaign"; different oral history projects such as "Xu Xiangqian: Smashing the Japanese Army's Nine-Route Attack on Southeast Shanxi" and "Xu Shenji, Wu Fushan: Moving Earth, Army Songs Singing Triumph—Recalling the Xiangtangpu Ambush" both contain the "Shentouling Battle" event), without forming large-area interconnections, only scattered on the outer part of the graph.

From the central region, multi-line interweaving creates large-scale many-to-many relationships in the graph center, with extensive clustering. Some events repeatedly appear in different oral history projects, or certain oral history projects narrate multiple events, such as the July 7th Incident, Songhu War, Cheqiao Campaign, Taierzhuang Campaign, Pingxingguan Victory, Xiangxi Campaign, Hengyang Campaign, Jinan Campaign, Taiyuan Campaign, Hundred Regiments War, Liaoshen Campaign, Pingjin Campaign, and Xi'an Incident. This indirectly indicates the importance of these events for anti-Japanese war historical research, facilitating further verification or discovery of relevant historical events, more comprehensively grasping the situation at the time, and opening new problem domains and providing new clues for researchers' in-depth study [6].

5.2.2 Event-Time Relationship Retrieving the event-time relationship knowledge graph and using functions to obtain the event-time distribution graph (see Figure 6 [Figure 6: see original paper]), we found that historical events present a state of both dispersion and aggregation in time. A small number of years contain only one or two historical events, mostly years when the war had not yet begun or had already ended. Other years show event aggregation, concentrated between 1937-1949. This indicates that the data source primarily focuses on historical events from the full-scale anti-Japanese war period (i.e., historical events occurring between 1937-1945, such as the July 7th Incident, Shanghai Defense War, Linyi Battle, and First Changsha Battle), supplemented by historical events from the liberation war period (i.e., events between 1946-1949, such as the Pingjin Campaign, Huaihai Campaign, Shanghai Campaign, and Hangzhou Campaign).

5.2.3 Event-Place Relationship Figure 7 [Figure 7: see original paper] shows the event-place relationship knowledge graph. As demonstrated in the graph center, historical events involved in this data source concentrate in Shandong Province, Jiangxi Province, Zhejiang Province, Tianjin City, Beijing City,

and other regions. Using the count function, the top five provincial-level administrative regions by event quantity proportion are Jiangsu Province (14.4%), Shanghai City (13.8%), Hunan Province (10%), Hubei Province (8.8%), and Guangdong Province (8.8%). Additionally, the periphery of this knowledge graph involves events abroad mentioned by narrators, such as Korea, India, and Myanmar. Among them, the expedition to India and Myanmar occurred in both India and Myanmar. This indirectly confirms that the Burma battlefield was a strategic junction connecting China's two major anti-Japanese theaters—the China theater and the Pacific theater—to some extent demonstrating that knowledge graphs facilitate discovering event-place knowledge associations, helping to retrospect, examine, and deepen historical research, and promoting in-depth humanities research development.

5.3 Knowledge Discovery Based on Social Network Relationships

Social networks are relatively stable relationship systems formed through social activity interactions among individual members, based on points and edges. Knowledge discovery based on social network relationships establishes person-relationship mappings, forming knowledge graphs of character co-occurrence relationships to display and verify the social relationship status and importance of relevant figures in different historical periods. This paper includes two levels: power group analysis of superior-subordinate relationships and character 梳理 of kinship relationships.

5.3.1 Power Group Analysis of Superior-Subordinate Relationships

Superior-subordinate relationships constitute an important component of social networks in oral history projects. This paper uses centrality (betweenness centrality) to measure individuals' core roles in social networks (code shown in Figure 8 [Figure 8: see original paper]). When an entity has high betweenness centrality, it means the entity has strong control over other figures, further indicating its power, status, and identity are in core leadership positions. Sorting entity betweenness centrality values and extracting the top 50 figures, see Table 2 . Simultaneously, figures were divided into two major categories by political party: core figures of the Chinese Communist Party and core figures of the Kuomintang.

Among Chinese Communist Party core figures, some are members of the first-generation central leadership collective, such as Zhu De, Deng Xiaoping, Peng Dehuai, and Mao Zedong; others are core military figures, such as Nie Rongzhen, Liu Bocheng, Peng Dehuai, Su Yu, Li Xiannian, and Peng Xuefeng. Among Kuomintang core figures, represented by Chen Cheng, Wei Lihuang, Yan Xishan, and Zhang Fakui, as shown in Figure 9 [Figure 9: see original paper]. Since numerous factions existed within the Kuomintang, and factional development and struggle were results of political power transition, analysis of Kuomintang core figures during the anti-Japanese war period can begin with factions. This data source mainly involves two major factions: the Huangpu (Whampoa) fac-

tion and local factions. The Huangpu faction includes important figures such as Du Yuming, Luo Zhuoying, and Xue Yue. This graph displays the network relationships of figures associated with Du Yuming, including Qiu Qingquan, Chen Peng, Ding Zhanfu, and Luo Zhuoying. Local factions involve the Jin Army faction and Guangdong Army faction. The Jin Army faction centers on Yan Xishan, with core figures including Li Mengyuan, Zhao Chenghuan, Cai Rongshou, Rong Honglu, and Shang Zhen. The Guangdong Army faction, such as Zhang Fakui's faction, includes major figures like Huang Fumeng, Zuo Hongtao, Liang Zhiqi, and Li Leifu. Additionally, we found figures associated with Wang Jingwei include his subordinates Xian Huanliu, Chen Gongbo, Li Shiqun, Chen Bijun, Chai Shan, and Zhou Fohai, his nephew Wang Qi, and that Wang Jingwei and Chen Bijun had both spousal and superior-subordinate relationships. Abandoning monotonous textual narratives, the multi-order lines of knowledge graphs provide guidance for rapid capture and complete outline of character relationships—this is the charm of advanced knowledge discovery in knowledge graphs.

5.3.2 Core Character 梳理 of Kinship Relationships Kinship relationships constitute the basic unit of social life forms [16]. Social networks based on kinship relationships govern entire social structures and social order. To ensure accuracy and scientificity in kinship relationship classification and appellation, this paper selected the relatively complete "Dictionary of Kinship Terms" [17] as reference material for identifying, correcting, and integrating kinship relationships. Simultaneously, considering that kinship systems and terms vary across countries, particularly the detailed classification and numerous definitions in China's kinship relationships, this paper included not only ordinary kinship relationships but also relationships mentioned in the data source such as same-clan relationships, classmate relationships, colleague relationships, teacher-student relationships, and old acquaintances into the kinship relationship category for statistical analysis convenience, as detailed in Table 3 .

With clear kinship relationship types and corresponding names, this paper takes Ye Ting, one of the founders of the Chinese People's Liberation Army, as an example. The character kinship relationship graph is shown in Figure 10 [Figure 10: see original paper]. Overall, with Ye Ting as the core starting point, it includes four types of relationships between associated figures and their kinship relationships. Direct semantic information can be obtained, including his wife Li Xiuwen and children Ye Huaming, Ye Zhengming, and Ye Zhengda, as well as hidden character associations generated by multi-level relationships, such as Ren Guang and his wife Xu Ren, both being Ye Ting's subordinates. Taking Ye Ting's subordinate Li Zifang as an example, Li Zifang has a mother-daughter relationship with Shi Hetang and a father-son relationship with Li Feiluo. Thus, the knowledge graph "connects" figures like Ye Ting-Ren Guang-Xu Ren-Li Zifang, providing an optimized path for deep revelation and discovery of character relationships, where entity "connections" can directly target character relationships and achieve multi-order knowledge discovery.

5.4 Knowledge Discovery Based on Spatiotemporal Network Relationships

Spatiotemporal network relationships include both time and space dimensions. The combination of time and space outlines a multi-dimensional associated world, where planar entities such as characters and events achieve transformation from single-dimensional to multi-dimensional reality through time and space augmentation, providing solutions for exploring migration paths and patterns of entities like characters and events changing over time and space dimensions.

5.4.1 Analysis of Character Spatial Distribution Based on Social Relationships Spatial distribution characteristics of social relationship networks take characters as core entities. Exploring character spatial relationship distribution can, on one hand, focus on a specific region to obtain character aggregation areas, thereby indirectly reflecting the number of people from a region participating in the anti-Japanese war; on the other hand, it can focus on group characteristics to explore spatial distribution patterns of group social relationship networks. Mapping character native place data to a bar chart (see Figure 11 [Figure 11: see original paper]) enables character-space information "clustering." This reveals that character social relationship spatial distribution is relatively dispersed, mostly concentrated in East China, with Jiangsu Province and Zhejiang Province as two main aggregation areas, while also having small amounts in Central China, South China, and Southwest China.

5.4.2 Analysis of Character Spatiotemporal Migration Trajectories Based on Work Experience Changes in characters' work experience simultaneously generate spatiotemporal migration. Knowledge discovery concerning migration trajectories also reflects social change characteristics during the anti-Japanese war period and holds important reference value for humanities research. This paper's character work experience includes both unit and military components, involving job transfers and career changes, with these relationships accompanied by temporal and spatial features that constitute the character spatiotemporal work experience system.

Taking a single character's work experience as an example, we retrieved the work experience of founding general Lü Zhengcao to achieve associated information knowledge discovery through character spatiotemporal migration trajectories, as shown in Figure 12 [Figure 12: see original paper]. We can observe that Lü Zhengcao had two work experience segments (military and unit). He earliest served as training instructor team leader in the 16th Brigade of the Kuomintang Northeast Army in Liaoning Province, leaving in 1933 to become regiment commander of the 647th Regiment of the 116th Division of the Kuomintang Northeast Army, then regiment commander of the 691st Regiment. In 1937, Lü Zhengcao left the Kuomintang Northeast Army and joined the Communist Party in the same year, becoming commander of the People's Self-Defense Army, with

his work location migrating from Liaoning Province to Hebei Province. By 1938, Lü Zhengcao left the People's Self-Defense Army to become commander of the 3rd Column of the Eighth Route Army, subsequently serving as member of the Jin Central District Committee of the Chinese Communist Party and director of the Ji Zhong Administrative Office and Director of the Ji Zhong Administrative Office. In 1943, Lü Zhengcao was transferred to the Jin Sui Military Region as region commander, with his work location migrating to Shanxi Province.

6 Conclusion

This paper constructs a knowledge graph using anti-Japanese war veterans' oral history archive resources as experimental data, forming a high-dimensional knowledge representation space closely related to oral history archive data examples. It can assist researchers in directly "locating" required information, verifying the practical and application value of digital humanities technology methods represented by knowledge graphs in knowledge discovery research. The paper completes data-level and knowledge-level storage and knowledge association of oral history archive resources, with the resulting knowledge graph association network serving as important "raw material" for subsequent multidimensional knowledge discovery. The knowledge graph, with its powerful semantic processing capabilities and open organization capabilities, lays the foundation for knowledge-based organization and intelligent application. Multidimensional knowledge discovery research based on knowledge graphs provides new pathways for excavating hidden content in humanities resources, enriching and expanding the knowledge discovery research system. The authors hope to use points to drive surfaces and local areas to 联动 the overall situation, injecting new thinking, new perspectives, and new momentum into innovative breakthroughs in humanities resource knowledge discovery. Simultaneously, exploring cross-type, cross-channel multidimensional knowledge discovery system construction [18] will help library, information, and archive resources become more inclusive and interconnected.

Due to limited interface screenshot display, the completeness and operability of the knowledge graph in this paper are inadequate. Therefore, in future research, the authors will attempt to find integration points between knowledge graphs and web frontends to provide users with online retrieval operations and more interconnection function displays, broaden humanities research pathways, and promote layered progression of humanities resources in theoretical innovation, methodological transformation, and model exploration, assisting humanities research to advance toward new models with multi-development subject participation, personalized retrieval, deep excavation, and interactive integration.

References

[1] Niu Li, Gao Chenxiang, Zhang Yufeng, et al. Discovery, reconstruction, and storytelling: Paths and methods of archival research from a digital humanities

perspective [J]. Journal of Library Science in China, 2021, 47(1): 88-107.

[2] National Archives Administration. "14th Five-Year Plan" for National Archives Development [EB/OL]. [2021-11-08]. https://www.saac.gov.cn/daj/yaw/202106/899650c1b1e4c0e9ad

[3] FAYYAD U M. Advances in knowledge discovery and data mining [M]. Cambridge: MIT Press, 1996.

[4] Ma Li, Jiao Licheng. Research and design of a knowledge discovery system based on rough set theory [J]. Microelectronics & Computer, 2003(3): 8-12.

[5] Jin Xiao'en. Research on knowledge-based management of digital library user information based on knowledge discovery [J]. Library Science Research, 2013(17): 18-20, 34.

[6] Song Xueyan, Cui Haonan, Liang Ying, et al. Knowledge discovery research of celebrity diary resources from a digital humanities perspective—Taking Wang Shijie's diaries as an example [J]. Information Studies: Theory & Application, 2021, 44(6): 105-111.

[7] Sun Minglei, Fang Xiaoke, Chen Xin. Research on knowledge graph construction of celebrity archives from a digital humanities perspective—Taking writer archives as an example [J]. Shanxi Archives, 2020, (6): 79-88.

[8] Yang Xixi. Analysis of historical archive resource organization and development paths from a digital humanities perspective—Concurrently discussing historicism and logicalism in archival management [J]. Archives Science Bulletin, 2019(2): 17-22.

[9] Pan Yumin, Ye Zhengkai. Reasons why oral history archives are archives [J]. Beijing Archives, 2016(5): 14-17.

[10] Zhu Lingjun. Research on the path of archive knowledge discovery driven by data [J]. Archives and Construction, 2020(2): 30-34, 13.

[11] Gao Chenxiang. Research on a knowledge discovery model for regional government microblogs from an archival science perspective [D]. Xi'an: Northwest University, 2019.

[12] YU H F, CHEN Y M, TSENG L M. Archive knowledge discovery by proxy cache [J]. Internet research: electronic networking applications and policy, 2004, 14(1): 34-47.

[13] PATTUELLI M C, HWANG K, MILLER M. Accidental discovery, intentional inquiry: leveraging linked data to uncover the women of jazz [J]. Digital scholarship in the humanities, 2017, 32(4): 918-924.

[14] Pan Wei. Responses to historical geography informatization under the "digital humanities" background—Entering the 2.0 era of historical geography informatization [J]. Journal of Yunnan University: Social Sciences Edition, 2018, 17(6): 80-87.

[15] Gao Song, Wang Xiangnü. Research on development and utilization of oral history archive resources from a digital humanities perspective [J]. Shanxi Archives, 2021(3): 61-70.

[16] Xin Xin. From marriage law to civil code—Research on kinship system with Chinese characteristics [A]//Shanghai Law Society. Shanghai Law Research Collection (2020 Vol. 9, Total Vol. 33)—Civil Code Marriage Family Women's Rights Protection Collection [C]. Shanghai: Shanghai Law Society, 2020: 5.

[17] Bao Haitao, Wang Anjie. Dictionary of kinship terms [M]. Changchun: Jilin Education Press, 1988.

[18] Ruan Guangce, Xia Lei, Zhou Mengwei. Analysis of knowledge services from a cross-media intelligence perspective [J]. Information Studies: Theory & Application, 2021, 44(7): 79-85.

Author Contributions

Deng Jun: Paper revision.
Wang Ruan: Proposed overall research 思路 and framework; data collection and analysis; paper writing and revision.

Abstract: [**Purpose/Significance**] Knowledge discovery of oral history archive resources based on knowledge graphs represents a new attempt at knowledge discovery in the digital humanities field, providing new pathways for fine-grained resource association, semantic querying, and personalized exploration. [**Method/Process**] Taking the data from the Anti-Japanese War Veterans Oral Data Center of Nanjing Normal University as the data source, we constructed a knowledge graph of anti-Japanese war veterans' oral history archive resources. Based on the graph examples, we carried out multidimensional knowledge discovery research from the aspects of overall project overview, event-theme relationships, social network relationships, and spatiotemporal network relationships. [**Result/Conclusion**] Digital humanities technology methods, represented by knowledge graphs, provide powerful tool support for knowledge discovery research and inject new kinetic energy into the deep development of humanities resources.

Keywords: knowledge graph; oral history archives; knowledge discovery

Note: Figure translations are in progress. See original paper for figures.