Postprint: Research on the Construction of Semantic Associations for Archaeological Excavation Image Databases

Authors: Gao Jinsong, Han Muzhe

Date: 2023-04-01T16:02:50+00:00

Abstract

[Purpose/Significance] To address issues in the organization of primary sources, this paper proposes a method for the digitization and semantic association of archaeological excavation data, assisting archaeology practitioners in avoiding inefficient workflows. [Method/Process] First, through concrete examples, the characteristics of primary sources in humanities disciplines are analyzed to design the process and methodology for raw data digitization. Second, archaeological excavation data from the Chawuhu Cemetery in Hejing, Xinjiang, is selected as the empirical data source to construct a graph database of archaeological excavation data. Finally, using co-occurrence relationships among artifacts as an example, the construction of semantic associations for the graph database is implemented. [Results/Conclusion] The construction of the graph database of archaeological excavation data and its semantic associations provides new methods and approaches for the digitization of such data, demonstrating promotional value and practical significance in the field of digital humanities.

Full Text

Research on Semantic Association Construction of Graph Databases for Archaeological Excavation Resources

Gao Jinsong, Han Muzhe School of Information Management, Central China Normal University, Wuhan 430079

Abstract: [Purpose/Significance] Aiming at the current problems in the collation of primary sources, this paper proposes a method that can realize the datalization conversion and semantic association of archaeological excavation resources, helping archaeologists avoid inefficient workflows. [Method/Process]

Firstly, combining examples, the characteristics of primary sources in the humanities are analyzed, and the process and method for the datalization conversion of primary sources are designed. Secondly, archaeological excavation data from the Chawuhu Cemetery in Hejing, Xinjiang, is selected as the empirical data source to construct a graph database of archaeological excavation resources. Finally, taking the co-occurrence relationships among artifacts as an example, the semantic association construction of the archaeological excavation resource graph database is implemented. [Result/Conclusion] The construction of the archaeological excavation resource graph database and its semantic associations provides new methods and ideas for the datalization conversion of archaeological excavation resources, and has promotional value and practical significance in the field of digital humanities.

Keywords: archaeological excavation resources; graph database; semantic associations; Chawuhu Cemeteries
Classification Number: G255
DOI: 10.13266/j.issn.0252-3116.2021.09.012

Archaeology, as a subfield of history within the humanities, is one of the earliest disciplines to adopt digital humanities technologies as research tools. Among the six main directions of digital humanities practice, "GIS-based historical geography visualization in historical studies" and "image analysis, color restoration, and digital reconstruction in archaeology" are both digital humanities directions closely related to archaeology [1]. In a statistical analysis of British digital humanities projects based on DHCommons, among 542 projects with labeled research fields, archaeology was labeled 88 times, and it was concluded that "historical research (including archaeology), language and literature studies, and library/information and museum studies are the most important research fields for British digital humanities projects" [2].

However, during the archaeological excavation process, a continuous stream of fragmented and diverse types of excavation resources is produced, which typically does not meet the conditions for direct structured organization and data application. This makes it difficult for newly discovered materials to integrate into existing knowledge organization systems, thereby affecting the overall progress of digital humanities research in archaeology. Therefore, exploring methods for data organization and semantic association construction that are suitable for the direct input and collation of archaeological excavation resources has important practical significance.

Related Research

Currently, both domestic and international research has achieved practical results in the organization, storage, management, and application of archaeological information resources. Representative projects include the STAR project funded by the UK Archaeology Data Service (ADS) [4-5] and the IADB database at the University of Reading [6], the Archeo-Info system in Germany [7], the OCHRE system at the University of Chicago [8], and the E-Arch system jointly developed by the Institute of Archaeology of the Chinese Academy of Social Sciences and Tsinghua University [9]. The information resources contained in these projects are based on organized digital resources. However, for more primary archaeological excavation materials, paper archives remain the main preservation solution for many institutions both domestically and internationally. A large amount of archaeological excavation data is stored in different institutions in the form of physical objects and digital copies using traditional methods, which creates severe barriers to knowledge revelation and sharing. The rich knowledge contained in excavation materials is significantly compromised in terms of orderliness, openness, security, and research value. Among these, the efficiency and technical challenges faced during the recording, collation, and datalization of archaeological excavation resources cannot be ignored.

Compared to archaeology, the fields of cultural heritage and museums (Library, Archive, and Museum—LAM) have stronger academic sensitivity to digital humanities and humanities-oriented knowledge services, benefiting from the large-scale digital conversion of LAM institutional collection resources that has been basically completed in China over the past 20 years [3]. From a macro perspective, the domestic large-scale collection resource digital conversion phase has come to an end. However, in the humanities, new primary sources similar to archaeological excavation materials continue to emerge, and scholars in various humanities fields still widely use traditional, inefficient methods for data collection and collation. Current technical research in the digital humanities field does not provide adequate solutions for these sporadically produced, lightweight primary sources that are not suitable or cannot be directly incorporated into relational databases.

Comprehensive results from other digital humanities fields show that research on related technologies and methods mainly focuses on database construction, knowledge visualization, and the development of disciplinary service platforms based on these. Hot topics include linked data publishing [10], text mining [11], metadata organization [12], ontology modeling [13], and knowledge graph development [14] in the digital humanities domain. The data sources for empirical research typically come from structured and semi-structured digital resources already in the LAM collection system, which usually have relatively definite framework structures and referable metadata control schemes. This allows researchers to complete knowledge extraction and representation according to information needs, as well as further construction of relational databases, development of knowledge graphs, and subsequent knowledge services. For heterogeneous external databases, data interconnection can also be achieved through knowledge fusion methods.

From a process perspective, the starting point of the above research is often the knowledge organization stage, dealing with secondary or tertiary information that has undergone digital and textual organization. The research object of this paper—archaeological excavation resources—belongs to more fundamental primary information, corresponding to raw materials collected from various scenarios, which is also common in other humanities disciplines besides archaeology. In the early stages of digital humanities research, almost all projects undergo the digitalization conversion of LAM institutional collection resources. The general approach is to structurally organize resources with low structural levels and store them in relational databases. This stage requires substantial human involvement, particularly humanities scholars engaging in extensive low-level and complex data collection and collation work [15]. From a macro perspective, China has basically completed the first stage, is currently in the second stage with considerable effectiveness, and is gradually moving toward the third stage [16]. Correspondingly, humanities research also has rich accumulations.

However, for the macro research framework, the accumulation from traditional humanities research and large-scale digitalization stages is a prerequisite for subsequent steps. This makes it difficult for lightweight primary sources sporadically produced in recent years to follow the macro process in the short term. Therefore, the authors reconstruct the micro-framework for primary source collation. For newly discovered lightweight primary sources, humanities scholars can simultaneously perform two tasks: (1) digitize and externally structurally describe primary sources, corresponding to the digitalization process of the macro research framework, with results storable in relational databases mainly for archiving; (2) conduct data analysis and modeling to clarify various objects and their relationships and attributes in primary sources, transforming them into intermediate data that can assist humanities research and storing them in graph databases; (3) based on the above steps, further achieve semantic description and semantic association construction of primary sources using graph database-stored intermediate data. For intermediate data, existing metadata frameworks can be reused for standardization to achieve data integration with the macro research framework, or semantic knowledge graphs can be formed through semantic association construction to achieve integration with existing knowledge graphs.

Analysis and Collation of Primary Sources

Characteristics of Primary Sources

Primary sources refer to documents, artifacts, phenomena, and other objects containing original information. The original information directly obtained from them is often structurally inconsistent, and the distribution of materials is more fragmented. Whether a material carries unrevealed original information is the criterion for defining it as a primary source. Common LAM data resource examples can be summarized into three categories: "unstructured data," "semi-structured data," and "structured data." Unstructured data manifests as "original information carried by documents, artifacts, and objects themselves," which can all be considered primary sources [16]. In archaeological work, documents, physical objects, and data materials directly recorded, collected, aggregated, and statistically analyzed through field surveys and archaeological excavations all belong to primary sources. In other words, primary sources in archaeology can be subdivided into field survey materials and archaeological excavation materials.

Compared to natural and social sciences, knowledge elements in humanities have more obscure associative relationships and evolutionary patterns, making the difficulty of organizing their knowledge frameworks greater, and descriptions of objects and processes are typically subjective. Therefore, before research is completed, it is difficult to directly structurally describe and store large amounts of content and knowledge. Through humanities scholars' research, classification, content description, and other knowledge revelation are typically performed based on information extracted from primary sources and external structural descriptions. The knowledge products generated by effective knowledge revelation are the conditions for primary sources to possess further datalization conversion capabilities—that is, materials can be described, expressed, stored, and applied through corresponding data. Although LAM institutions serving the public contain a certain proportion of primary sources in their collection structures, most collection materials have research foundations. For distinction, the authors refer to materials with research foundations and conditions for datalization conversion as collection materials.

Primary sources in humanities differ from collection materials in various aspects, as detailed in Table 1 . At different research stages, the material entities faced by the two may overlap. Overall, as research deepens and the degree of material analysis improves, primary sources gradually transform into collection materials with improved datalization conditions.

The authors analyzed an inscribed bronze plate unearthed in 1993 from Jiunüdun Tomb No. 3 in Pizhou, Jiangsu, as an example. The analysis details are shown in Figure 1 [Figure 1: see original paper]. The "Primary Source" column contains archaeological excavation data from an excavation brief published in 2002 [17], while the "Collection Material" column supplements inscription interpretation information from 2019 [18]. Five bronze plates with identical decorative patterns and shapes were unearthed from the same tomb, and the example bronze plate was discovered during cleaning to have inscriptions on its bottom. Since the archaeological report did not distinguish them, the external structural description in the primary source column cannot provide accurate exclusive naming for this inscribed bronze plate. As research progresses and the inscription is interpreted, the structured description in the collection material column based on research content is sufficient to distinguish this inscribed bronze plate from the four co-unearthed coiled snake pattern plates, at which point it can be stored as structured data in a relational database for further knowledge management and service.

Combined characteristic comparison and example analysis show that before relevant research achieves certain results, objects in primary sources are difficult to define through descriptive naming and hierarchical classification. Existing structured description methods cannot identify and distinguish objects with differences in primary sources, making the structured information converted from these materials unhelpful for further disciplinary research and knowledge sharing even when incorporated into existing relational databases and knowledge graphs, and may even cause semantic ambiguity. However, the relationships of objects in primary sources within the knowledge network are relatively stable. By describing objects' known relationships in existing knowledge systems, primary sources can be converted into intermediate data with determined relationships, open naming, and undefined classification. Intermediate data directly converted from primary sources contains many unknown or pending attributes and is not suitable for direct public data sharing, but functionally can meet the needs of further humanities research, helping humanities scholars with data input, collation, and statistics analysis, and promoting the conversion of primary sources to collection materials. Additionally, intermediate data describes objects through relational structures and itself has corresponding knowledge revelation functions. If it can be semantically matched with the ontology structure of the target domain, it can be standardized and normalized and integrated into existing knowledge graphs as an external database, directly achieving datalization conversion of primary sources.

Methods for Primary Source Collation

Chen Tao et al. divide the digital humanities research process into three stages: resource digitalization conversion, digital resource text construction and research, and datafication and intelligent research of textual resources. This describes the macro process of current digital humanities research and proposes a macro digital humanities research framework [16]. Collection materials in the macro research framework will undergo digitalization conversion and structural description, be stored in relational databases, and converted into humanities data. China has basically completed the first stage over the past 20 years and is currently in the second stage with considerable effectiveness, gradually moving toward the third stage. From a micro perspective, for lightweight primary sources continuously produced from various humanities disciplines, humanities scholars find it difficult to actively perform digitalization conversion and structural description in a short time, preventing direct storage in relational databases. According to the macro digital humanities research framework, subsequent textification, datalization, and intellectualization are also difficult to advance.

Although relational databases and semantic knowledge graphs have strict metadata control and ontology structures, and perform better in terms of rigor, standardization, and long-term preservation and sharing, they are not the best choice for primary source collation. The authors believe it is necessary to reconstruct the processing methods and procedures for primary sources in humanities to enable them to synchronize with the macro digital humanities process with higher efficiency. A feasible approach is to perform external structural description on content in primary sources that can be easily structurally described, reuse existing relational database structures and metadata specifications for digitalization and textification, and describe other content that is difficult to standardize, unify, and standardize with existing relational databases through relationships between objects.

Graph databases excel at handling large amounts of complex, interconnected, and low-structured data, have stronger data compatibility, and express and process associative relationships more intuitively and efficiently. Their functions "focus more on knowledge mining and computation, discovering implicit knowledge and visualizing it, achieving functions such as question-based retrieval and spatiotemporal display, and promoting innovation in digital humanities research methods under artificial intelligence environments" [16]. Although graph databases are simple and easy to use, the lack of standardized controlled vocabularies makes it difficult for different graph databases to communicate with each other, and the data silo problem remains difficult to avoid. However, the authors' direct goal for primary source processing is to convert them into intermediate data. This approach is oriented toward the research process of humanities scholars rather than public knowledge sharing, making graph databases more applicable for the input, collation, and storage of primary sources. After comprehensive consideration of various factors, the authors decided to select Neo4j for the construction of the archaeological excavation resource graph database. The simple "N-E" (Nodes & Edges) and "K-V" (Keys & Values) structures of graph databases can accommodate large amounts of intermediate data, and their ease of use can also meet the data maintenance needs of non-professionals.

Primary Source Collation Model and Model Interaction

Due to differences in data acquisition and processing methods and target users, based on the reconstructed digital humanities research framework above, this paper proposes a collation model for humanities primary sources centered on graph databases (hereinafter referred to as the "collation model"). This model and the knowledge service model commonly used in current digital humanities platforms, which is centered on knowledge graphs (hereinafter referred to as the "service model"), are two independent models that can interact. Simultaneously, the graph database constructed in this paper should be distinguished from the graph database dedicated to data warehousing in the service model. The graph database constructed in this paper can be regarded as a data conversion module in the macro knowledge service system. Different from the knowledge service goal oriented to the public, its function domain is in the humanities research stage and is necessary preparation for digital humanities platforms to carry out continuous public knowledge services.

The interaction structure between the collation model and the service model is shown in Figure 3 [Figure 3: see original paper]. The left side is the collation model, and the right side is the service model. The two models have three interaction points:

First, external structural interaction. This refers to the initial discovery or production of primary sources, where collators use traditional methods to perform external structural description of materials and store them in the structured database underlying the service model. This description of directly perceived or observed content cannot deeply reveal resource connotations and serves a function similar to bibliographic cataloging, mainly for resource identification.

Second, intermediate data interaction. This interaction occurs after the "data modeling" step in the graph database construction. At this point, primary sources have been organized into intermediate data represented by nodes and relationships through analysis and modeling. By integrating intermediate data with service model data, some primary sources can be incorporated into existing knowledge bases, completing the initial parsing of some primary sources.

Third, semantic association interaction. This is one of the important application scenarios of graph databases. Through the deep query function of graph databases, it can assist users in association construction, thereby completing bottom-up ontology construction or correcting the ontology framework in the service model. This interaction mainly targets primary sources with high innovative research value. Its application scenario corresponds to the digital humanities research process. The full revelation of material connotations in these research efforts is the foundation for primary sources to integrate into existing knowledge graphs and provide knowledge services to general users through digital humanities platforms.

Among the three interactions, external structural interaction is completed through traditional methods and will not be elaborated further. The authors will combine practical examples to further explore the implementation methods of intermediate data interaction and semantic association interaction.

Construction of Archaeological Excavation Resource Graph Database

The construction of a graph database is a necessary condition for achieving intermediate data interaction between the collation model and the service model. This paper selects archaeological excavation resources, which have arduous collation tasks and strong foundational positions in the humanities discipline system, as data objects to construct a graph database. The empirical data is sourced from the excavation resources of tombs No. 1, No. 4, and No. 5 at the large clan cemetery group of Chawuhu in Hejing, Xinjiang.

Construction Process of Archaeological Excavation Resource Graph Database

The knowledge service platform corresponding to the collation model involves the design and development of the entire system platform. Under the premise of ensuring data security, a relatively simple B/S architecture (Browser/Server

Architecture) is sufficient to meet the input, storage, and application needs of intermediate data for small- to medium-scale limited user groups. This paper will focus on discussing the Neo4j graph database construction process and the implementation methods of some business logic based on Cypher, without elaborating on the user interaction layer and related database access connection mechanisms in the platform architecture.

The construction process of the archaeological excavation resource graph database mainly includes six steps: functional analysis, data preparation, data analysis, data modeling, graph generation, and knowledge application. Each step contains different contents, as specifically shown in Figure 4 [Figure 4: see original paper]. In the construction process, data analysis and data modeling are key steps. This section will focus on these two aspects combined with examples.

Data Analysis

The purpose of data analysis is to define the node and relationship types, properties, and property value types and domains that need to be presented in the graph database.

(1) Node Definition. Any node in a graph database must contain an exclusive node index (Index: ID), a label (Node: Label), and several properties (Property: Keys). Node numbering needs to consider the basic classification and hierarchical logic of the knowledge represented by various nodes in relevant disciplinary systems. The numbering itself can form a basic knowledge framework based on prior knowledge. Compared to simple sequential numbering, this numbering rule is hierarchical, comprehensive, and has certain scalability. Node labels are defined based on dataset subclass divisions, and nodes with the same label can be regarded as instances of that class entity (Individual). Taking the most common tomb 遗迹 in archaeological excavation resources as an example, the information contained in the "Tomb Registration Form" attached to archaeological reports varies in detail but always involves several categories: "site," "tomb shape," "burial form," "owner information," and "excavated artifacts." These can be temporarily designated as the main node classes in tomb excavation resources, and corresponding adjustments can be made to the main classes when encountering specific situations in practice.

Node properties are typically descriptive or limiting content of relevant nodes, supporting storage of text, numeric, vector, and other data types, as well as various formats of graphics, dynamic graphics data, and links. The allocation method of node properties is not fixed and requires case-by-case analysis in practice. Node and relationship properties and property values in graph databases can also be obtained through graph computation and undergo independent or batch addition and deletion operations.

The node definitions for the archaeological excavation primary source graph database are shown in Table 2 .

(2) Relationship Definition. The initial relationships in a graph database are all direct relationships without further statistics, reasoning, or processing. The authors identified five main types of representative relationships from the empirical dataset, as shown in Table 3 .

In Neo4j graph databases, relationships must have defined directions, but during graph traversal and other retrieval operations, relationships are default undirected or bidirectional. Among the above relationships, there are four inter-class relationships and one intra-class relationship. The intra-class relationship "earlier_{than}" is the only directional relationship, identifying the stratigraphic sequence between some tomb 遗迹. These relative positional relationships can be used to determine the relative chronology of related tombs in ordered archaeological strata.

Relationship class names typically characterize association types. Two nodes and the relationship class name between them can usually be regarded as a complete RDF triple. Relationship property allocation differs from nodes, with relationship properties more often corresponding to statistical attributes. Property values can be applied to graph computation and occasionally used to store other content.

Data Modeling

Graph database data models include minimal data models and complex data models. The former is a theoretical model of the network's initial structure, while the latter extracts real materials from example data sources for modeling.

(1) Minimal Data Model. The minimal data model of the archaeological excavation resource graph database is shown in Figure 5 [Figure 5: see original paper]. This model displays the five node classes and five initial relationship classes described above. In the figure, the box header represents the class name, the frame represents the data types of properties and property values, and the relationship label also shows the relationship class name, relationship properties, and data types of property values.

(2) Complex Data Model. The complex data model construction selects two tombs with superimposed relationships from the data source as examples. The relevant primary source materials are:

Material 1: Registration information for M052_{C4} and M233_{C4} at Chawuhu Cemetery from the original report's tomb registration form, as shown in Figure 6 [Figure 6: see original paper].

Material 2: Stratigraphic superimposition information for M052_{C4} and M233_{C4}, stated as "Superimposition relationship… M52→M233."

Effective information from the materials can organize 21 nodes and 20 initial relationships. The model expression is shown in Figure 7 [Figure 7: see original paper]. The example omits owner information and details such as tomb superimposition relationships, stone enclosures, and tomb scales, which can be input into the graph database as independent nodes or node properties in practical application. Artifacts with incomplete, unclear classification, or non-standard descriptions, such as "bean handle" and "sheep ribs," are screened out.

Through the minimal model diagram and complex model diagram, it can be seen that the proposed archaeological excavation resource graph database has a complete structure and clear logic, meeting the conditions for real data import. Importing real data into the graph database and removing isolated nodes generates an initial graph database of Xinjiang Chawuhu Cemetery archaeological excavation resources, containing 606 nodes and 2,739 relationships. These data resources stored in the graph database can be regarded as intermediate data converted from primary sources and can share data with other graph databases and relational databases through data migration tools.

Semantic Associations of Archaeological Excavation Resource Graph Database

The third interaction between the collation model and the service model is association interaction, aiming to achieve deep-level semantic interaction. By applying relevant functions of graph databases, it promotes knowledge discovery during archaeologists' research on excavation resources, which then acts on the knowledge graphs of existing archaeological knowledge service platforms through knowledge fusion. Its implementation mainly relies on the deep query function of graph databases, through which archaeologists can mine and construct some deep-level associations in primary sources.

Deep Query and Association Construction of Archaeological Excavation Resource Graph Database

The initial graph database records only 底层 knowledge nodes and initial associations, which are all first-degree associations in terms of depth. In various humanities disciplines, there exist deep associations with practical significance. In archaeology, for example, the relationships among various artifacts contained in each stratum or 遗迹 unit (such as a tomb, pit, or house foundation) are called co-occurrence relationships among artifacts [20]. These help researchers separate meaningful fixed artifact assemblages from the overall artifact collection, discover patterns, and thus conduct more detailed research on chronology segmentation and cultural type identification. This section demonstrates the deep query and association construction functions of the archaeological excavation resource graph database using the construction of artifact co-occurrence relationships as an example.

Co-occurrence relationships are a second-degree association within the "artifact" node class. In the initial graph database constructed from Chawuhu Cemetery archaeological excavation resources, the intermediate node for co-occurrence relationships is the "site" node. If two different types of artifacts are unearthed at the same site, they are considered co-occurrent, with frequency being the number of sites where both were unearthed. Taking "Liuhe Cup AII (Rid: 511220)" and "Spoon Cup AII (Rid: 511g20)" as examples: first, query all second-degree associations between them (see Figure 8 [Figure 8: see original paper] left), finding that 25 tomb sites contained both "Liuhe Cup AII" and "Spoon Cup AII." Count these "site" nodes as Sc. Second, establish a "coexistence_{with}" relationship between the "Liuhe Cup AII" and "Spoon Cup AII" nodes, writing the Sc value into the relationship's frequency property to complete the co-occurrence relationship creation (see Figure 8 right).

Traversing the entire graph database for deep queries discovered and created 3,804 types of co-occurrence relationships. The Cypher batch processing code for assigning co-occurrence relationship frequencies is shown in Table 4 .

The deep query and association construction functions of graph databases can extract and display meaningful knowledge associations hidden in flat data networks. Through layer-by-layer construction of second-degree, third-degree, and even deeper relationships between nodes, the knowledge network architecture can remain intact while separating instances, achieving more stable knowledge storage and meeting higher-level knowledge service needs. Furthermore, knowledge graphs widely used in knowledge services are a semantic network relying on ontology for knowledge organization. In primary sources, semantic associations that are not available or not concerned with in existing knowledge graphs may be discovered, which are often the main targets of knowledge discovery. Based on deep query and association construction, the 梳理 of semantic associations can be regarded as a bottom-up ontology construction process. This deep association creation based on graph databases can effectively increase the flexibility and applicability of domain ontologies, which is self-evident for digital humanities knowledge base construction with fragmented knowledge systems and strong foundational variability.

Semantic Association Examples of Archaeological Excavation Resource Graph Database

Based on the co-occurrence relationships among artifacts realized in Section 5.1, the authors take M089 and M156 from Chawuhu Cemetery No. 4 as examples to demonstrate the connection structure between the graph database as a data layer and the knowledge graph, reflecting the position of second-degree and deeper associations constructed based on the graph database in the knowledge graph structure, as shown in Figure 9 [Figure 9: see original paper].

The semantic associations constructed based on the archaeological excavation resource graph database can help efficiently, timely, and effectively integrate primary sources organized and researched by archaeologists into existing knowledge service platforms, enabling primary sources to achieve true datalization and accelerating the transformation of humanities research results into knowledge products that can serve the public.

Conclusion

First, this paper clarifies the practical problems existing in current archaeological excavation resource collation, reviews and summarizes relevant domestic and international practical projects and research results, combines actual cases to analyze the characteristics of humanities primary sources represented by archaeological excavation resources, and proposes the viewpoint of converting primary sources into intermediate data. Second, selecting the Neo4j graph database as a tool for primary source collation, the paper reconstructs the primary source collation process based on the macro digital humanities research framework, proposes an interactive graph database model for digital humanities primary source collation combined with the mainstream knowledge service model centered on knowledge graphs, and analyzes the interaction forms between the two models. Third, taking archaeological excavation resources from the Chawuhu Cemetery in Hejing, Xinjiang as an example, the paper discusses in detail the construction process of the archaeological excavation resource graph database, proposes corresponding data analysis and data modeling methods, and achieves intermediate data conversion of primary sources. Finally, using the co-occurrence relationships among artifacts as an example, the paper implements semantic association construction of the archaeological excavation resource graph database through traversal-based deep query and association construction technology, providing ideas for intermediate data datalization conversion and further knowledge fusion with knowledge graphs. Future work can combine natural language processing, context-aware similarity computation, and more graph algorithms to further develop the applied functionality of this method, enabling it to play a greater role in digital humanities construction and development.

References

[1] Wang Xiaoguang. The emergence, development, and frontiers of "Digital Humanities" [EB/OL]. [2021-04-02]. http://blog.sciencenet.cn/home.php?mod=space&uid=67855&do=blog&id=

[2] Lin Zefei. Analysis of research hotspots in British digital humanities projects—An empirical study based on the DHCommons project database [J]. Information and Documentation Services, 2018(1): 97-104.

[3] Liu Wei, Xie Rong, Zhang Lei, et al. National data infrastructure construction for humanities research [J]. Journal of Library Science in China, 2016, 42(5): 29-39.

[4] May K, Binding C, Tudhope D. A STAR is born: some emerging semantic technologies for archaeological resources [C]//Proceedings of the 36th conference on computer applications and quantitative methods in archaeology. Budapest: CAA, 2008: 402-407.

[5] Binding C, May K, Souza R, et al. Semantic technologies for archaeology resources: results from the star project [C]//Proceedings of the 38th conference on computer applications and quantitative methods in archaeology. Granada: CAA, 2010: 555-561.

[6] Integrated archaeological database [DB/OL]. [2021-04-02]. http://iadb.org.uk/.

[7] Battenfeld I, Beckmann N, Schultze J, et al. Unifying archaeological databases using triples [C]//4th International conference on cooperation and promotion of information resources in science and technology. Beijing: IEEE, 2009: 281-284.

[8] OCHRE Data Service [DB/OL]. [2021-04-02]. https://voices.uchicago.edu/ochre/.

[9] Zhang Shuangyu. Research on archaeological data mining and optimization of E-Arch archaeological information system [D]. Beijing: Tsinghua University, 2012.

[10] Zhu Fanfan, Gao Jinsong, Liang Yanqi. Creation and publication of linked data for museum collections—Taking China's top ten paintings as an example [J]. Library Theory and Practice, 2018(4): 96-101.

[11] Guo Jinlong, Xu Xin. Research on text mining in digital humanities [J]. University Library Science, 2019, 45(6): 34-49.

[12] Wang Lili, Zhu Xiaomei. Research on design and application of metadata cataloging standards for ancient book seals [J]. Library, 2020(1): 106-111.

[13] Zhou Yaolin, Zhao Yue, Sun Jingqiong. Research path for organization and retrieval of intangible cultural heritage information resources—Investigation and design based on ontology methods [J]. Journal of Intelligence, 2017, 36(8): 166-174.

[14] Zhou Lina, Hong Liang, Gao Ziyang. Construction of Tang poetry knowledge graph and design of its intelligent knowledge service [J]. Library and Information Service, 2019, 63(2): 24-33.

[15] Chen Tao, Liu Wei, Shan Rongrong, et al. Research on application of knowledge graphs in digital humanities [J]. Journal of Library Science in China, 2019, 45(6): 34-49.

[16] Zeng Lei, Wang Xiaoguang, Fan Wei. Smart data in the fields of libraries, archives, and museums and its role in digital humanities research [J]. Journal of Library Science in China, 2018, 44(1): 17-34.

[17] Kong Lingyuan, Chen Yongqing. Excavation of Jiunüdun Tomb No. 3 in Pizhou City, Jiangsu Province [J]. Archaeology, 2002(5): 19-30.

[18] Ma Yongqiang, Cheng Wei. Interpretation of inscriptions on bronze plates unearthed from Jiunüdun Tomb No. 3 in Pizhou, Jiangsu [J]. Cultural Relics, 2019(10): 79-81.

[19] Xinjiang Institute of Cultural Relics and Archaeology. Xinjiang Chawuhu—Excavation report of large clan cemetery [M]. Beijing: Oriental Press, 1999.

[20] Zhang Zhiheng. General theory of Chinese archaeology [M]. Nanjing: Nanjing University Press, 2009.

[21] Wang Hong, Zhang Qingqing, Cai Weiwei, et al. Research on domain ontology storage methods based on Neo4j [J]. Application Research of Computers, 2017, 34(8): 2404-2407.

Author Contributions: Gao Jinsong: Proposed research ideas, designed research plan, revised paper; Han Muzhe: Data collection and experiments, wrote and revised paper.

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