Knowledge Representation and Semantic Association Construction for Unearthed Pottery Shapes in Archaeological Typology: Postprint

Authors: Han Muzhe, Gao Jinsong, Li Yu

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

Abstract

[Purpose/Significance] Aiming at archaeological typology, this paper proposes a knowledge representation model suitable for feature description of unearthed artifacts, and based on this, proposes corresponding semantic mapping and ontology extension methods to break through the semantic barriers caused by traditional typological methods and promote knowledge sharing and reuse. [Method/Process] First, it analyzes the semantic barrier problems caused by archaeological typological thought and its traditional methods, and proposes a targeted knowledge representation strategy for archaeological typology; second, combining the corpus characteristics of archaeology and typological logic, it decomposes the vessel shape description of unearthed pottery along two dimensions: genus-species relationship and whole-part relationship; subsequently, it proposes knowledge representation schemes and feature vector expressions for pottery parts and types layer by layer, thereby constructing a knowledge representation model for unearthed pottery vessel shapes; next, based on the knowledge representation model, it reveals an archaeological typology ontology extension method based on conditional equivalence mapping to realize the construction of archaeological typology semantic associations; finally, taking two pottery vessels from the Liujia site in Qinghai as examples, it demonstrates the form and process of applying the proposed method for typological knowledge representation and its ontology graph visualization effects. [Results/Conclusion] Taking pottery vessel shapes as an example, the knowledge representation and semantic association construction of unearthed artifacts from the perspective of archaeological typology is a new attempt at digital humanities research in the field of archaeology, which can provide technical support for the transformation of typology research from experience-based to data-driven.

Full Text

Knowledge Representation and Semantic Association Construction of Unearthed Pottery Shapes for Archaeological Typology

Han Muzhe[1], **Gao Jinsong**[1], **Li Yu**[2,3] [1]School of Information Management, Central China Normal University, Wuhan 430079 [2]School of Archaeology and Museology, Sichuan University, Chengdu 610225 [3]Academy of Art and Design, Sichuan University of Culture and Arts, Mianyang 621000

Abstract: [Purpose/Significance] From the perspective of archaeological typology, this paper proposes a knowledge representation model suitable for describing the characteristics of unearthed artifacts, and corresponding semantic mapping and ontology extension methods to break through the semantic barriers caused by traditional typological methods and promote knowledge sharing and reuse. [Method/Process] First, we analyze the semantic barriers created by archaeological typological thought and its traditional methods, and propose a targeted knowledge representation strategy for archaeological typology. Second, combining the linguistic characteristics of archaeological corpora with typological logic, we decompose the shape descriptions of unearthed pottery along two dimensions: genus-species relationships and whole-part relationships. Subsequently, we propose knowledge representation schemes and feature vector expressions for pottery parts and types layer by layer, thereby constructing a knowledge representation model for the shapes of unearthed pottery. Next, based on this model, we reveal an ontology extension method for archaeological typology grounded in conditional equivalence mapping to achieve semantic association construction. Finally, using two pottery vessels from Qinghai Liuyan as examples, we demonstrate the form and process of typological knowledge representation using our method and its ontology graph visualization effects. [Result/Conclusion] The knowledge representation and semantic association construction of unearthed artifacts from an archaeological typology perspective, using pottery shapes as an example, represents a new attempt at digital humanities research in archaeology that can provide technical support for transforming typological research from experience-based to data-driven approaches.

Keywords: Archaeological Typology; Unearthed Pottery; Knowledge Representation; Semantic Association Construction

Disciplinarity constitutes the core dimension of the three dimensions of digital humanities. How to consider the development path of digital humanities from traditional humanities research, and the possibilities for further integration and development of digital technology and humanities research, are questions that digital humanities scholars must deeply contemplate [1]. Archaeology is an important field of digital humanities research. Current digital humanities research in archaeology has achieved certain results in data compilation and registration [2], knowledge service platform construction [3], and GIS digital mapping and scene reconstruction [4]. However, the truly challenging aspect that can deeply penetrate archaeological research—employing data science to assist archaeologists in professional analysis—remains relatively weak. This indicates that the integration of archaeology and digital humanities has not yet reached a profound level.

Although there exists a rudimentary conventional terminology system for describing unearthed artifacts, the field of archaeology has not yet implemented controlled vocabularies, inevitably generating a series of subsequent semantic problems due to the randomness of linguistic expression. Sometimes, even within the same document, corresponding semantic confusion phenomena appear. For example, for the pottery vessel "dou" (a stemmed bowl), it is generally believed to consist of three parts from top to bottom: "pan" (plate), "bing" (stem/handle), and "zu" (foot). However, for the "stem" of dou, some literature habitually calls it "ba" (handle), causing descriptive confusion with the "ba" of other pottery types that have handles. Obviously, these two "ba" terms refer to morphologically and functionally different things in different contexts. This chaotic state creates widespread discrepancies in descriptions of unearthed artifacts across documents, preventing some archaeologists from referring to and reusing existing description norms during original data compilation. This affects the efficiency of archaeological knowledge reproduction and may further intensify the chaotic state. Yu Weichao pointed out the difficulty in typological practice of "not knowing what name would be accurate for many vessel shapes." He argued that in cross-cultural archaeological research, artifacts with the same name but different forms undoubtedly cause conceptual confusion, yet unifying vessel names across all archaeological cultures would be impossible for a long time [5]. To date, this situation remains largely unchanged.

This paper attempts to propose a knowledge representation model suitable for describing the characteristics of unearthed artifacts for archaeological typology research. By modeling the decomposition of physical features, it aims to break the semantic barriers caused by archaeological classification problems and achieve deep, fine-grained representation of archaeological knowledge.

2 Related Research

Currently, only one study addresses archaeological typology knowledge representation and semantic association construction. A. Bruno et al. constructed a corresponding task ontology model using the classification problem of lighting fixtures in Islamic archaeology as an example, representing a new attempt at semantic organization for archaeological typology [6]. However, the research topic, proposed vocabulary, and associative relationships cannot be reused in domestic archaeological knowledge representation practice, and the typological foundation of European Islamic archaeology differs significantly from Chinese archaeology.

Most metadata description specifications and ontology frameworks for cultural artifacts are initiated and proposed by collection institutions represented by museums and museum alliances. Commonly used specialized metadata specifications include CIDOC Conceptual Reference Model (CIDOC CRM) [7], Categories for the Description of Works of Art (CDWA) [8], and Visual Resources Association Core Categories (VRA Core) [9], aiming to provide standard norms and consistent information description frameworks for the collection, display, and sharing of cultural resources. The metadata compilation standard system drafted by China's State Administration of Cultural Heritage and Peking University covers metadata standards and cataloging rules for various types of cultural artifacts, with the draft standard publicly released online [10]. These knowledge representation schemes can provide reliable resource collection and organization frameworks for cultural heritage protection and collection database construction, and can provide terminology support in corresponding fields. However, due to certain differences between archaeology and cultural heritage protection in terms of description purposes, methods, granularity, and focus, metadata specifications and ontology frameworks developed from a cultural heritage protection perspective are generally too general and lack specificity. Their revealed knowledge granularity is difficult to meet the specific needs of archaeological typological analysis and cross-cultural research. Currently, few digital humanities studies combined with archaeology can truly penetrate the real difficulties in the research process, and knowledge representation of archaeological typology constitutes a key challenge that digital humanities must address to deeply engage with archaeological research.

3 Analysis of Current Status of Archaeological Classification Knowledge Representation

Archaeological typology is one of the fundamental methodologies of archaeology. The most basic problem in typology is classification. Since raw materials obtained from fieldwork are numerous and fragmented, archaeologists cannot describe each artifact individually, necessitating classification. Through reasonable organization of different categories at various levels, a classification system (hierarchy) can ultimately be constructed [17]. Classification thought has accompanied the birth of modern Chinese archaeology for a century. Although methods have changed several times, the underlying thought remains consistent. It is often combined with "queuing" (sequential arrangement) thought to reveal temporal development patterns. Classification methods are most commonly used to describe pottery, bronze, stone artifacts, and features like tombs and settlement patterns. Among these, the classification of pottery is almost the most universal and important issue in archaeological data analysis of prehistoric periods.

3.1 Analysis of Knowledge Reasoning Processes in Archaeological Classification Methods

From a scientific research methodology perspective, archaeological typology primarily uses inductive reasoning, generally defined as reasoning from the particular to the universal, representing a form of probabilistic reasoning [18]. In archaeological typology, individual artifacts themselves or other features that can be directly observed and recorded constitute the "particular" in the definition of inductive reasoning. The role of induction is to identify commonalities and differences from these "particular" artifact entities, grouping entities with more similar features into classes to form "universal" representations that refer to a category of artifacts. This universal symbolic representation constitutes a type. Since archaeological excavation itself can be viewed as a sampling-based information collection activity, the statistics and induction of pottery shapes from a site belong to incomplete inductive reasoning activities, producing knowledge types that constitute incomplete empirical knowledge.

Using pottery cauldrons from the Chawuhu No. 4 cemetery in Xinjiang as an example, Figure 1 [Figure 1: see original paper] illustrates the knowledge reasoning rules of archaeological classification methods. As shown in Figure 1, three pottery vessels from Chawuhu No. 4 cemetery belong to raw data before organization. After archaeologists observe and describe them, the recorded knowledge corresponds one-to-one with individual artifacts. However, at No. 4 cemetery alone, 64 pottery cauldrons were unearthed. Due to limited space in archaeological reports, it is impossible to describe each pottery vessel individually when they exist in batches. Therefore, inductive reasoning methods must be used to classify similar artifacts, summarizing classes, types, and forms layer by layer. From the classification results, the detail knowledge units in artifact classification descriptions decrease sequentially from form, type, to class, with the thinking process filtering out specificity layer by layer while retaining commonality. However, archaeologists' induction of class, type, and form may not proceed sequentially but can flexibly employ both upward induction and downward subdivision based on experience. The underlying thought remains incomplete inductive reasoning.

Although this classification thought provides a structural foundation for knowledge representation, in archaeological typology practice, the standards actually used for classification are selected very flexibly, even arbitrarily. For example, in the same-level knowledge induction illustrated in Figure 1, the selected characteristic knowledge units are not unified—some mention overall shape, some refer to local features and shapes, and some involve comparisons of size or weight. Additionally, beyond vessel shape, decorative pattern types, positions, designs, and material characteristics are frequently used in archaeological typological descriptions without fixed sequences or expression levels, and with varying degrees of detail. This makes it difficult to find a unified knowledge structure framework when conducting archaeological typological knowledge organization.

3.2 Analysis of Existing Problems in Archaeological Classification Methods

When the above classic classification thought is expanded using traditional methods based on archaeologists' empirical knowledge, it often presents or creates the following problems:

3.2.1 Arbitrary Classification Standards

Classification standards correspond to knowledge representation dimensions, which are prerequisites for achieving unified knowledge representation. However, in archaeological typology practice, the selection of classification standards is highly flexible and even arbitrary. As shown in Figure 1, the characteristic knowledge units selected for knowledge induction at the same level are not unified. This flexibility, while accommodating the complexity of real-world artifacts, creates significant barriers to systematic knowledge organization and computational analysis.

3.2.2 Chaotic Terminology System

Although CIDOC CRM ontology family concepts and other related metadata models can be reused to represent and organize archaeological domain knowledge, no existing terminology vocabulary can properly handle concepts involved in archaeological typology. Terminology system chaos frequently occurs in archaeological practice, as exemplified earlier. Another important issue in typological description is the extensive use of fuzzy expressions—relative concepts such as large/small, high/low, thick/thin are widely used, sometimes with comparative terms at different levels. This is characteristic of incomplete empirical knowledge widely present in humanities and social sciences, making knowledge organization and representation of such descriptions difficult.

3.2.3 Lack of Sharing and Integration Solutions for Typological Symbolic Markers

Current typological symbolic marker methods used in Chinese archaeology have certain norms, typically presented as "category name (Chinese character) + type alias (uppercase Latin letter) + form alias (uppercase Roman numeral)." Archaeologists sometimes reduce or expand levels according to needs and do not necessarily pursue normative symbolic marker systems. Since archaeological reports are usually organized by site, archaeologists maintain consistency of typological symbolic markers within the same report for convenience in further statistics and analysis. For example, in the burial registration form [19] of Chawuhu No. 4 cemetery's M229, artifacts can be directly marked as "Cauldron AI." However, this symbolic marker system is difficult to interoperate across different sites or even across different excavation or organization batches of the same site. For instance, Chawuhu site's No. 4 and No. 1 cemeteries were both excavated by the Xinjiang Institute of Archaeology between 1985-1989, with complete excavation reports published together as monographs. Yet even the pottery typological markers between these two cemeteries have minor inconsistencies: "Lugged Cup Type A" and "Lugged Cup Type B" in No. 4 cemetery correspond to "Lugged Cup Form II" and "Lugged Cup Form I" in No. 1 cemetery, respectively. This prevents unification of burial goods grade information in the burial registration forms of the two cemeteries, creating obstacles for subsequent researchers in data usage and statistics.

Archaeological classification methods are typically applied vertically within a site or culture as a unit. In cross-site or cross-cultural research, original typological symbols are often directly abandoned due to lack of integration solutions, becoming "disposable" symbolic systems. In new research processes, archaeologists often must re-induce and classify types based on their own knowledge structures using a small number of comprehensively described original materials, severely affecting work efficiency.

In summary, traditional archaeological classification methods have obvious defects in knowledge organization structure, terminology description systems, and symbolic representation systems. Together, these create semantic barriers that hinder knowledge sharing and reuse, and are unfavorable for depicting and restoring physical objects and factual phenomena. For example, for "Cauldron CI" registered in M95 of Chawuhu No. 4 cemetery, since it was not separately expanded in the report, information about this artifact can only be obtained through deductive reasoning from descriptions of this pottery type. Except for clearly knowing it has double flat lugs, details such as base shape, belly shape, and rim shape are all uncertain.

Moreover, the current status of artifact description based on traditional classification methods is very unfavorable for subsequent introduction of digital humanities methods. Existing typological symbolic marker systems mask detailed knowledge as feature items, significantly reducing the effectiveness of knowledge applications. Typological symbolic markers contain almost no semantic information. To illustrate, examining the artifact information in Table 1, their typological markers are all different, and the artifact class names alone cannot provide comparable semantic information. Thus, even with corresponding semantic distance measurement methods, reliable results cannot be obtained.

However, after investigation, the author extracted the classification information for "lugged cups" from No. 4 and No. 1 cemeteries as shown in Figure 2 [Figure 2: see original paper]. From the descriptions and schematic diagrams, "Lugged Cup B" from No. 4 cemetery and "Lugged Cup I" from No. 1 cemetery should belong to the same category. Based on this understanding, comparing the two burials in Table 1 allows for a more objective assessment of their similarity in unearthed artifacts. Therefore, how to break through the semantic barriers caused by current archaeological classification methods is a prerequisite for further knowledge organization and application.

3.3 Strategies for Archaeological Typology Knowledge Representation

The above analysis shows that the main problems in current archaeological classification methods lie in the inconsistency of knowledge structure, terminology systems, and symbolic systems, which significantly reduce the comprehensibility of their semantic information and limit their effective application scope. To solve these problems, unified knowledge representation methods must be established for archaeological typology, with corresponding formalization schemes preset to enable integration with domain ontologies through mapping methods, thereby achieving fusion of typological knowledge and archaeological domain knowledge.

The first step in knowledge representation schemes for archaeological typology is to find universal language from original corpora to describe corresponding objects. Then, using knowledge units expressed in universal language as basic units, knowledge representation structural models are established based on the inherent logical relationships of objects. Universal language mainly divides into two types: one consists of general concepts understandable through common sense without any professional background, such as "bowl," "plate," "mouth," "base," "large," "small," "round," "flat," etc. These concepts and their designations in a language are conventional and rarely cause semantic understanding obstacles. The other type consists of domain-specific universal language within the target field, which may require basic professional knowledge, such as "gui," "xu," "ring foot," "spout," "flared," "contracted," etc. These concepts are universal within archaeology and museology and require no further explanation or limitation.

Through examples, we can see that universal language describes very single features: "bowl," "plate," "gui," "xu" are all category names for pottery; "mouth," "base," "ring foot," "spout" are names for different parts of pottery. These noun-attribute universal languages are often substitutes for objects or feature subjects, corresponding to entities in semantic webs. "Large," "small," "round," "flat," "flared," "contracted" belong to adjectival modifiers that modify objects or feature subjects, corresponding to properties in semantic webs. Throughout typological descriptions, most pottery types are represented by noun phrases with the structure "adjective + head noun" (adj. + n.), showing certain structural stability. This indicates that archaeological typology knowledge possesses a foundation for knowledge representation. Generally, this knowledge structure can be transformed and represented using feature vectors (eigenvector).

Whether knowledge representation for archaeological typology can break the semantic barriers analyzed in the previous section and promote knowledge sharing and reuse constitutes the significance of this research. From the "lugged cup" case, traditional typological symbolic marker systems cannot directly represent feature details of related vessel shapes, making them contain minimal effective semantic information. However, if we decompose the universal language phrases in their descriptions into independent words and characters, reorganize them according to their original descriptive logic, and establish a knowledge representation model, we can structurally store the knowledge units and their relationships associated with a certain artifact through the knowledge representation model, laying a foundation for subsequent knowledge application and introduction of other digital humanities methods.

4 Data Collection and Organization Scheme

Since the categories of unearthed artifacts are complex, this paper selects pottery, the most widely used category in archaeological typology, as the research object. The sample scope is limited to complete pottery containers and their necessary components like lids. Pottery sherds, pottery tools like spindle whorls and pottery balls, and pottery figurines, supports, and other uncertain containers are excluded. This research primarily targets archaeology researchers and their disciplinary research processes. Therefore, in the knowledge acquisition stage, we still use authoritative literature in the disciplinary field as the data source for organizing corpora, extracting and specifying terminology using specific natural language processing methods, and defining attributes and relationships based on real samples. Archaeological typology is mainly applied to prehistoric period research, and the Neolithic period is also the most important stage for pottery application in archaeological typology research. After comprehensive consideration of data source authority and representativeness, this paper selects *Chinese Archaeology: Neolithic Volume* compiled by the Institute of Archaeology, Chinese Academy of Social Sciences [20] as the literature data source for corpus collection, with some line drawings and case materials supplemented by archaeological reports from Qinghai [21] and Xinjiang [19].

4.1 Organization of Original Terminology Vocabulary for Unearthed Pottery

Pottery unearthed from Neolithic sites is generally the primary identifier of cultural characteristics. Descriptions of unearthed artifacts in literature data sources are mainly elaborated in "cultural characteristics." The corpus collection and processing strategy in this paper divides into three steps: manually traverse the entire text content of the literature, extract the parts describing pottery in the main text by sentence units, and store them in serial number form; split the collected natural sentences, retaining only vocabulary and phrases related to pottery description, which requires restructuring or semantic completion of some sentence structures to preserve association information of related vocabulary; further split the vocabulary and phrases obtained after splitting, organize, clean, and deduplicate them to form a terminology vocabulary, retaining derivable semantic relationships in thesaurus form as a foundation for subsequent semantic association construction. A specific process example is shown in Figure 3 [Figure 3: see original paper].

Archaeological descriptions of a pottery artifact typically involve vessel class, shape, texture, manufacturing technique, slip, painted patterns, and other aspects. During corpus collection and original terminology extraction, terms are classified according to their described content combined with context, with genus-species and whole-part relationships organized in thesaurus form. Due to limited sample size in single cases, some terms not easily classifiable in the early organization stage will be reclassified as sample size increases. The six sentences in original corpus 215.2.1 shown in Figure 3 already cover 61 terms across five aspects. The complete terminology vocabulary involves even more aspects and levels, making hierarchical division crucial for clear knowledge representation. Theoretically, expression of pottery artifact instances can be simply divided into four description dimensions: shape, texture, technique, and decoration. If archaeological culture-level associations are considered, dimensions like culture, chronology, and associated objects can also be integrated, thereby embedding pottery expression itself into existing semantic knowledge networks for reuse.

Since typology most commonly analyzes "shape," this paper selects "shape" as the main object for development. Organizing all original corpora from the data source according to the above method and extracting the "shape" related corpus portion yields the material library for the next step of feature extraction. Based on the content of the "shape" material library, the split words can be divided into five types: shape terminology, part terminology, relative modification, descriptive modification, and quantitative modification. Among these five concept types, shape terminology and part terminology describe the overall and local features of target artifacts, while the other three modifier types are attributes that modify the above concepts.

4.2 Dimensions of Unearthed Pottery Shape Description

Traditional metadata applied in museology typically does not delve into the part level. Even when component descriptions are involved, they are mixed within shape descriptions. However, from an archaeological typology perspective, descriptions of pottery parts sometimes become crucial turning points for distinguishing pottery's temporal evolution. Moreover, beyond descriptive modifications of pottery morphology, some comparative relative features and quantitative features are equally important in analysis. Therefore, this paper treats genus-species relationships and whole-part relationships as the two primary dimensions for pottery shape description. The genus-species dimension treats individual pottery as a whole, using its overall shape's practical function as the basis for pottery category division. The whole-part dimension structurally decomposes the pottery whole into several functional components, and further arranges and combines possible qualitative and quantitative modification attributes for each part to achieve precise description.

4.2.1 Genus-Species Relationships of Pottery Shapes Current archaeological typology is a complex field highly dependent on experience, with inconsistent feature points and widely varying knowledge granularity demands. The main work of typology is defining the genus-species relationships of pottery shapes, with division logic from top to bottom being: variety, type, and form, from abstract to concrete, ultimately associating with specific artifact instances.

Generally, pottery variety (variety) is predetermined based on prior knowledge, mainly according to the practical function of certain pottery, such as cooking vessels, drinking vessels, eating vessels, storage vessels, and water-drawing vessels. Specific vessel shapes usually have conventional category names. Most people can identify different vessel categories in daily life such as "bowl," "plate," "bottle," and "cup" through common sense. The differentiation of unearthed pottery categories also mainly relies on experience, with a few ancient vessel types no longer in practical use referring to appellations and definitions in literature and classics, such as "ding" (tripod), "li" (tripod with hollow legs), "dou" (stemmed bowl), and "yan" (steamer). The identification features of pottery categories can be defined as "class name subjects," which mainly rely on manual annotation for extraction, with manual extraction results manifested as top-level shape terminology.

After determining categories, the division of pottery types within the same category mainly relies on some characteristic partial elements. For example, for "hu" (jar) in archaeology, typing can divide it into "belly-lugged hu," "pierced-lug hu," "spouted hu," etc., based on local features. Among these, characteristic part terminology such as "belly lug," "pierced lug," and "spout" can be called "type features."

If pottery samples belonging to the same type but with different forms are discovered at a site or within a cultural area, further form division is required. Form division has no unified standard and sometimes considers factors beyond shape such as painted patterns, decorative styles, and clay texture and color. If only shape is considered, the main reference is shape terminology or part terminology with various modification constraints. These features can be collectively called "form features." An example of typological description for unearthed pottery is shown in Figure 4 [Figure 4: see original paper].

4.2.2 Whole-Part Relationships of Pottery Shapes Archaeological typology rarely specifically discusses whole-part relationships of pottery, but in typological analysis, pottery part features are major reference elements. For example, when dividing forms of a single vessel class "belly-lugged hu" from the same site, its unified type feature "belly lug" and "flat base" should be treated as fixed parameters. The main feature points are the shapes of mouth, neck, shoulder, and belly, with comparison knowledge mainly coming from the "relative modification" level of "part terminology." When summarizing features of different types within a site's different chronological phases or an archaeological culture, the investigation may expand from vessel classes to cross-vessel-class part features. For instance, Xinjiang's Chawuhu culture is typified by spouted vessels, where the investigation of "spout" in "part terminology" would encompass spouted cups, spouted jars, and spouted hu, crossing vessel classes.

Currently, no classification research or terminology standardization work for pottery parts exists in archaeology. This paper can only rely on information induction from selected literature data sources. To contain and store contents, pottery containers typically have a mouth-above-base-below structure. Therefore, this paper divides pottery parts according to their structural positions into upper parts (PH), middle parts (PM), lower parts (PB), and attachment parts (PA). The division methods and examples for each part are shown in Figure 5 [Figure 5: see original paper].

Extending from the upper-middle-lower region division method in Figure 5, expansion marking can also be performed according to the vertical structure of parts. The numbering rule is that the higher the corresponding part position, the smaller the serial number. For example, the shoulder of a sample "Hu (Dahecun T1 :5)" is typically above the belly, so its "broad shoulder" and "belly · UD" can be marked as PM1 and PM2 respectively.

Each part terminology can also separate many subclasses with modification constraints. Statistics from terminology separated from literature data sources show that modifiers mainly divide into three types: descriptive modification, relative modification, and quantitative modification. Descriptive modification provides qualitative description of part features, such as "flat base," "round base," "belly lug," and "bag-shaped foot." Relative modification comprises comparative or fuzzy quantitative modifiers, mostly pairs of words modifying morphology, such as "large/small, concave/convex, wide/narrow, flared/contracted." Quantitative modification refers to quantitative attributes of related parts, such as the number of "lugs" being 2 for double-lugged vessels and the number of "feet" being 3 for tripods. Beyond representing quantity, dimensions, proportions, angles, relative position coordinates, and other meaningful statistical values may all be options for quantitative modification, with data types potentially more flexible and diverse. The subdivision of attributes belongs to archaeologists' research scope and will not be further discussed in this paper.

5 Knowledge Representation Model for Unearthed Pottery Shapes

After clarifying the two description dimensions of genus-species relationships and whole-part relationships for unearthed pottery shapes, we must further represent pottery shape knowledge. Here, we must still follow archaeological typology logic for layer-by-layer representation, gradually refining to ultimately propose an overall pottery shape knowledge representation model.

5.1 Knowledge Representation of Pottery Parts

As shown above, pottery parts are important reference factors for pottery type division. Therefore, in formal representation, part concepts can be treated as object properties of pottery types. Since divided pottery parts need to include structural position information and three modification attributes for that part, the formula expression is:

$$P_{mn} = \{R_L, R_{P_m}, F_D, F_R, F_C\}. \quad L \in (H, M, B, A)$$

Where P_{mn} represents a specific part; R_L represents the structural position of that specific part, with values distinguishing whether the part belongs to upper (H), middle (M), lower (B), or attachment (A); R_{P_m} represents the default part to which the specific part belongs; F_D, F_R, F_C are three attributes of the specific part, with no limit on attribute quantity.

Parts can be further subdivided according to combinations of the three modifier terms, so the three attributive modifiers should be represented as data properties of each part. After division, pottery part knowledge representation needs to include the part's structural position and three modification attributes for that pottery part. Different combinations of modification attributes can all serve as fine classification criteria for pottery parts. Pottery part description can only set three levels: the top level is the part's structural position, marking whether the part is located at the pottery's upper, middle, lower side, or as an attachment; the middle level is the default part (P_m) without any modification attributes specified, such as general "mouth," "belly," "base," "lug," etc. The parts marked with "UD" in Figure 4 have not specified modification attributes; parts with any subdivision by adding modification attributes below this are specific parts (P_{mn}), with bottom-level part descriptions needing to include specific values.

In actual data sources, all three modification attributes may be unspecified. Unspecified attributes should be assigned default values, with specific assignment methods requiring case-by-case analysis.

5.2 Knowledge Representation of Pottery Types

The inherent logic of archaeological typology is the main basis for the following pottery type division and knowledge expression. Considering only the genus-species hierarchy and pottery part representation, the logic of unearthed pottery type division can be simplified as shown in Figure 6 [Figure 6: see original paper].

Figure 6 shows the basic logic but does not cover association information or include relevant attributes for reasoning and computation. Next, we will expand pottery shape descriptions layer by layer with rule subdivision to propose a specific knowledge representation model that supports reasoning and computation.

5.2.1 Knowledge Representation of Pottery Varieties

For variety description, the four aspects mentioned above cover relatively general information, with association information written using upward association methods. The related information written for varieties is the practical function of that pottery class. There are many ways to describe outer contour. Image feature annotation methods based on two-dimensional or three-dimensional coordinate systems have mature applications in image retrieval, but modeling costs are high and there is no immediate necessity for introduction. Moreover, pottery of the same category may differ significantly in form and size, which likely directly reflects functional differences in antiquity and should naturally not be merged at the variety level. In summary, pottery shape expression can include four information aspects: relations (R), parts (P), outline (O), and size (S):

$$V = \{R, P, O, S\}$$

Some common and fixed varieties have specific overall shapes, which can be decomposed into necessary or excluded parts. For example, the pottery class "ding" (tripod) must have a "base," and the base must have divided "solid feet." The concept set of "solid feet" has mutually exclusive relationships with "stem," "ring foot," "hollow foot," etc., creating another necessary relationship between "ding" and the above parts—"ding" cannot contain the above parts. Conversely, pottery categories containing the above parts cannot be "ding." These axioms can be used in knowledge graph applications to classify some pottery categories through semantic reasoning. However, the possibility of determining pottery categories through semantic reasoning is small, as many samples can contain similar parts and even have similar functions and overall shapes. Under controllable data scales, manual annotation remains the most effective category division method. Nearly a century ago, scholars like Liang Siyong [22] and Li Ji [23] already attempted manual annotation of pottery shapes using combined qualitative and quantitative methods. Similar thinking can still be applied to future semi-automatic or automatic feature extraction and semantic annotation processes. Additionally, specifications and dimensions are also references for type division. Some samples may have similar shapes but vastly different sizes, likely reflecting functional differences in antiquity and should not be merged at the variety level. In summary, pottery shape expression can include four information aspects: relations (R), parts (P), outline (O), and size (S).

The expression for pottery variety can be written as:

$$V_x = \{R_{Usage}, P_{V_x}^{must}, P_{V_x}^{cannot}\}. \quad P_{V_x}^{must} \cap P_{V_x}^{cannot} = \emptyset$$

Where V_x represents a pottery variety; R_{Usage} represents the possible practical functions of that pottery class; $P_{V_x}^{must}$ represents parts that must be included in complete state of that pottery class; $P_{V_x}^{cannot}$ represents parts that cannot be included in that pottery class. The values of $P_{V_x}^{must}$ and $P_{V_x}^{cannot}$ are usually default parts. Using "ding" as an example, its variety representation model is shown in Figure 7 [Figure 7: see original paper].

5.2.2 Knowledge Representation of Pottery Types and Forms

In the following description of type (V_{xy}), association information can link to the URI of its belonging variety information and inherit defined attributes. For components, the focus is writing the type's 标志性 type-feature parts. The value range of type-feature parts is the absolute complement of the "impossible parts" set of its belonging variety, denoted as $P_{V_x}^{cannot}$. Similarly, the pottery type level is still insufficient for morphology and size statistics, so no values are assigned in the type expression. The pottery type expression can be written as:

$$V_{xy} = \{R_{V_x}, P_{V_{mn}}^{type}\}. \quad P_{V_{mn}}^{type} \subset P_{V_x}^{cannot}$$

Where V_{xy} represents a pottery type; R_{V_x} represents the direct upper class of this pottery type, i.e., its direct variety; $P_{V_{mn}}^{type}$ represents the type-feature part of this pottery type.

The form (V_{xyi}) expression covers the URI of its belonging type information and inherits attributes. For components, all specific characteristic parts possessed by this form of pottery need to be written. Pottery of the same form usually has similar morphology, so outer contour must be described in pottery form expression. Related algorithms in image retrieval can extract and annotate sample overall two-dimensional contours and key part features, then complete pottery category matching and recommendation through similarity calculation of annotated image features [24-25]. However, under controllable data volumes, manual annotation and statistics-based descriptions are more practical in traditional humanities. There are multiple means to describe pottery basic outer contours using statistical methods. Since literature images are mostly line drawings at eye level and many lack original size annotation, outer contour depiction requires proportional description, considering only the main parts of the shape and ignoring attachments. Two methods are feasible: 1) Tangent point method—align pottery base with X-axis, place the midpoint of base line at the origin of a right-angle coordinate system, make internal and external tangents to the pottery outer contour, record intersection coordinates of each pair of tangents, and restore and compare the pottery basic outer contour through coordinate positions. 2) Cross-section proportion method—applicable to pottery with symmetrical main shapes, select the maximum and minimum diameter parts from upper, middle, and lower sides, record diameters of each part, calculate and write the ratio of diameters of different parts to roughly determine the basic shape trend. The outer contour descriptions obtained by the two methods are coordinate sets and arrays respectively, requiring different methods for similarity calculation. Pottery of the same form often has similar specifications, and the mean size of representative samples or multiple samples can be written as reference. The representation of pottery instances obtained directly through describing specific artifacts is basically the same as forms. The expression for pottery form and instance can be written as:

$$V_{xyi} = \{R_{V_{xy}}, P_{V_{xyi}}, O_{V_{xyi}}, S_{V_{xyi}}\}. \quad P_{V_{xyi}} \subseteq P_{mn}$$

Where V_{xyi} represents pottery form or instance; $R_{V_{xy}}$ represents the direct upper class of this pottery piece, i.e., its direct pottery type (or variety if not typed or uncertain); $P_{V_{xyi}}$ represents all specific characteristic parts of this form or instance, with value range being specific parts; $O_{V_{xyi}}$ represents shape contour information obtained through contour description methods; $S_{V_{xyi}}$ represents specification and size information of this form or pottery.

5.3 Knowledge Representation Model for Unearthed Pottery Shapes

Integrating existing materials, we can further construct a knowledge representation model for unearthed pottery shapes combined with examples, as shown in Figure 8 [Figure 8: see original paper]. This model can represent pottery shape information extracted layer by layer from archaeological literature descriptions. Besides integrating structured data from existing databases, its given description methods also facilitate users' knowledge extraction from literature descriptions or direct observation of physical objects. In the model's structural layer, terminology related to pottery types and part features is organized and induced from archaeological corpora, maximizing reuse of disciplinary terminology and restoring archaeological expression methods. In the future, terminology collected in the instance layer, after semantic disambiguation and deduplication, can be provided to users as normative vocabulary in thesaurus form, with associated terms visually displayed through photos, line drawings, etc., greatly facilitating archaeologists' data compilation processes and achieving maximum terminology reuse and consistent expression.

The structural layer demonstrates value examples for the three modification attributes of parts. Different value combinations of these three attributes can achieve precise depiction of any pottery part. Moreover, according to actual needs, through weight assignment and control, customized knowledge representation meeting different contexts can be achieved. The model examples simplify the settings for the three attributes. Among them, F_D and F_R, two qualitative modification attributes, usually require semantic alignment before determination. F_R values sometimes need reference ranges given, and qualitative attributes may have multiple simultaneous values in application. The example uses only the "quantity" parameter for F_C, but actual value methods may be much richer.

Representing unearthed pottery shape knowledge in this way forms the basis for achieving formal representation. The classification problem in archaeology is addressed through this structured approach.

6 Semantic Model Construction for Unearthed Pottery Shape Knowledge

The knowledge representation model for unearthed pottery shapes proposed above completes the conceptualization of target domain terminology and representation of terminology relationships, mainly solving the process of "describability of objects." To truly achieve knowledge organization and representation at the application level, further formal representation based on ontology is needed to complete the "machine realizability" function, enabling the archaeological typology-oriented knowledge representation method provided in this paper to fuse with domain knowledge graphs and achieve deeper knowledge services.

6.1 Ontology Extension Method for Archaeological Typology Based on Semantic Mapping

Different from general ontology construction schemes, this paper does not intend to construct an independent domain ontology model but attempts to propose an extension scheme for unearthed pottery shape representation that can integrate with existing relevant domain ontologies. The domain ontology model used for integration with the pottery shape knowledge ontology in subsequent sections is an archaeology domain ontology model constructed by the authors based on the CIDOC CRM ontology family. Since few entity classes in the pottery shape knowledge ontology can be reused from existing domain ontology models, this extension ontology can not only interact with specific domain ontologies but also possesses certain generalizability. The key to achieving ontology integration is selecting interactive entity classes in the domain ontology and appropriate semantic mapping methods. Semantic mapping between ontologies is a bridge between heterogeneous ontologies [26], commonly used in ontology integration, ontology reconstruction and reuse, and knowledge exchange between ontology models [27].

The basic logic of typology is induction—selecting parts containing corresponding features from numerous facts to merge into classes. The most basic semantic mapping rule that can be used to express this knowledge logic is conditional equivalence mapping, expressed in OWL using the EquivalentTo statement, which represents equivalence relationships between two entity classes and can be constrained through object property or data property relationships. The basic rules of conditional equivalence mapping can be schematically shown in Figure 9 [Figure 9: see original paper].

In Figure 9, there is entity class A containing instances a1, a2, a3; and entity class B containing instance b. Object property P has domain class A and range class B. Data property D has domain class A with values m, n.

If we want to extract and induce all instances in class A associated with instance b through object property P into class A', the mapping statement is:

$$A' \equiv A \sqcap (\exists P.\{b\})$$

Through ontology reasoning functionality, entity class A' will be inferred as a subclass of class A, containing instances a2 and a3.

If we want to extract and induce all instances in class A with data property D value m into class A", the mapping statement is:

$$A'' \equiv A \sqcap (\exists D.\{m\})$$

Through ontology reasoning functionality, entity class A" will be inferred as a subclass of class A, containing instances a1 and a2.

The above lists the most basic forms of conditional equivalence mapping. More complex mapping conditions can be defined for multi-level nested expressions in applications, but to ensure rule stability, semantic mapping constraint conditions should not exceed three.

Additionally, in the pottery type representation process, we set certain parts that a pottery type cannot contain ($P_{V_x}^{cannot}$). This setting itself is not merely for knowledge representation but helps avoid erroneous ontology reasoning results, improve ontology quality, and play a role in subsequent knowledge computation when establishing ontology models for pottery classes by setting exclusion conditions through semantic association rules. According to Formula 3's rules, the method to constrain $P_{V_x}^{cannot}$ in the typology ontology model is by adding NegativeObjectPropertyAssertions entries. Using the pottery class "ding" shown in Figure 7, which cannot contain the default part "hollow foot," as an example, it can be represented as:

$$\text{not(ding } P9\_consists\_of \text{ hollow foot)}$$

6.2 Research on Archaeological Typology Ontology Extension

Based on the knowledge representation model for unearthed pottery shapes and the semantic mapping rules for typology ontology extension, the ontology extension task at this stage is to formally convert each part of the above expressions, integrate them with the domain ontology according to semantic association rules, and thereby achieve integration of typology knowledge into the domain knowledge graph. After analysis, the formalization strategy for the unearthed pottery shape expression is shown in Table 2 .

As shown in Table 2, knowledge elements in the unearthed pottery shape expression are defined one by one. The extension model adds two new entity classes: $arc : HE92\_Object\_Component\_Part$ representing shape parts and $arc : HE93\_Object\_Component\_Position$ representing part structural positions. Others are entity classes from the original domain ontology, constrained by conditional equivalence mapping rules, used for hierarchical representation of pottery shape parts and pottery types. This achieves formal semantic representation of archaeological typology, enabling it to fuse with knowledge graphs in archaeology or cultural heritage fields. Due to data integrity and space limitations, Table 2 does not expand the form expression but treats it directly as instance processing. Its formal representation method is similar to that of types.

The interaction position between the typology extension ontology and the most general CIDOC CRM model is between $HE92\_Object\_Component\_Part$ and $E19\_Physical\_objects$, with their direct association represented by $P9\_forms\_part\_of$ $(consists\_of)$. Under $HE93\_Object\_Component\_Position$, four subclasses are defined to represent upper, middle, lower, and attachment structural positions: $HE931$, $HE932$, $HE933$, and $HE934$. In general applications, these four subclasses do not directly participate in knowledge representation, but they can serve as framework nodes in knowledge computation processes to determine branch conditions of knowledge structure trees. Additionally, in practical applications, the structural positions of pottery parts can help represent knowledge about incomplete parts of some unearthed pottery.

7 Empirical Study of Unearthed Pottery Shape Knowledge Representation

This paper still uses the two pottery vessels from Qinghai Liuyan shown in Figure 8 as examples to demonstrate the effect of pottery shape classification knowledge representation. Both painted pottery belly-lugged hu examples are from the Liuyan cemetery, with artifact excavation numbers Liuyan M528:1 and Liuyan M578:4, both belonging to the Majiayao culture, with cultural phases being early Banshan type and early Machang type respectively. For convenient representation, this paper simplifies the original report's type division, with the simplified classification structure shown in Figure 10 [Figure 10: see original paper].

In Figure 10, the pottery variety is "hu" (jar) (V_6). "Hu" in pottery is a very common storage vessel, whose must-include parts are "mouth," "neck," "belly," and "base," with generally no limited impossible parts. Therefore, the knowledge representation for the variety "hu" of the two artifact examples is:

$$V_6 = \{R_{Usage} = \text{storage vessel}, P_{V_6}^{must} = \{\text{mouth, neck, belly, base}\}, P_{V_6}^{cannot} = \emptyset\}$$

The type is "belly-lugged hu" (V_{64}), whose associated upper class is "hu" (V_6), with type feature being "double belly lugs" (PA_{11}), represented as:

$$V_{64} = \{R_{V_6} = \text{hu}, P_{V_{mn}}^{type} = \{\text{double belly lugs}\}\}$$

The forms are "belly-lugged hu A" (V_{64A}) and "belly-lugged hu B" (V_{64B}). Omitting contour and specification information, their expressions are respectively:

$$V_{64A} = \{R_{V_{64}} = \text{belly-lugged hu}, P_{V_{xyi}} = \{\text{straight mouth, straight neck, round shoulder, flat drum belly, doub}$$

$V_{64B} = \{R_{V_{64}} = \text{belly-lugged hu}, P_{V_{xyi}} = \{\text{flared mouth, short neck, broad shoulder, deep belly, double belly l}$

For the specific parts appearing in V_{64A} and V_{64B}, taking "short neck" (PH_{21}), "flat drum belly" (PM_{221}), and "double belly lugs" (PA_{11}) as examples, the addition of modification attributes yields:

$$PH_{21} = \{R_L = H, R_{P_m} = \text{neck}, F_R = \text{short}, F_C = 1\}$$

$$PM_{221} = \{R_L = M, R_{P_m} = \text{belly}, F_D = \text{flat}, F_D = \text{drum}, F_C = 1\}$$

$$PA_{11} = \{R_L = A, R_{P_m} = \text{lug}, F_D = \text{belly lug}, F_C = 2\}$$

Transforming the hierarchical feature vectors into semantic information according to the association rules shown in Table 2, the formal representation of the two pottery examples is shown in Table 3 .

Inputting the above data into the ontology editing tool Protégé 5.5.0, the ontology graph interface generated using the OntoGraf tool is shown in Figure 11 [Figure 11: see original paper]. The figure identifies entity classes according to the CIDOC CRM ontology concept hierarchy. If pottery shapes are expanded to the "type" ($HE91164\_V_{64}$, belly-lugged hu) level, the hierarchy of the extended domain ontology model can be extended to the instance layer.

Conclusion

The thinking of precise description and quantitative comparison of unearthed artifacts from a typological perspective was proposed at the inception of modern archaeology, yet effective practical methods remain lacking today. Guided by the inherent logic of archaeological typology, this paper attempts to describe unearthed pottery shapes from two dimensions—genus-species relationships and whole-part relationships—and uses a feature vector model to represent knowledge of different parts and type hierarchies of pottery, thereby constructing a typology-oriented knowledge representation model for unearthed pottery shapes. On this basis, we propose an archaeological typology ontology extension method that integrates extended ontologies with existing domain ontologies through conditional equivalence mapping, promoting the fusion and sharing of typological knowledge and domain knowledge.

Using pottery shapes as an example for knowledge representation and semantic association construction of unearthed artifacts from an archaeological typology perspective represents a new attempt at digital humanities research in archaeology. It is expected to provide archaeologists with consistent structured terminology vocabularies and formalized typological ontology models in applications.

Related methods help efficiently solve deep, fine-grained representation problems of unearthed artifact types. Knowledge applications and services based on this foundation are expected to promote knowledge sharing and improve reuse effects in archaeology. Digital humanities applications in archaeological research are still in their infancy, and current research has many directions for further exploration. Due to the lack of large-scale publicly available digital corpora, this study only targeted and organized a small amount of data for experiments. Subsequent knowledge organization efforts should, on the basis of mature datasets, attempt to solve large-scale automatic semantic annotation and knowledge extraction problems for archaeological typology corpora. In terms of knowledge services, we need to explore digital implementation of archaeological typology applications based on semantic similarity computation.

References

[6] ALMEIDA B, ROCHE C, COSTA R. Archaeological classification and terminology: the case of Islamic archaeology of the al-Andalus [C]// TOT h2 2017. Terminologie & Ontologie: Théories et Applications, Chambéry, France. 2017: 221-236. [7] CIDOC-CRM [DB/OL]. [2021-09-14]. http://www.cidoc-crm.org/. [8] CDWA [DB/OL]. [2021-09-14]. http://www.getty.edu/research/publications/electronic_{publications}/cdwa/. [9] VRA Core [DB/OL]. [2021-09-14]. http://www.loc.gov/standards/vracore/. [10] Draft Release of Metadata Standards for Digital Protection of Cultural Artifacts [EB/OL]. [2021-09-14]. https://www.lib.pku.edu.cn/portal/cn/news/0000001494. [11] ZHANG Kaixing, WANG Li, LIU Xianxi, et al. Design of virtual assembly process knowledge base for rice harvester chassis based on ontology [J]. Journal of Agricultural Mechanization Research, 2015(11): 121-124. [12] WANG Ying, DAI Yiru, WANG Jian. Design and implementation of steel production process knowledge base system based on three-dimensional framework ontology modeling [J]. Computer Knowledge and Technology, 2017(2): 242-244. [13] LI Changjie, MING Xinguo, QIU Kunhua, et al. Knowledge representation method for aircraft tooling design based on ontology [J]. China Mechanical Engineering, 2014, 25(19): 2614-2619. [14] REN Minshan, CAI Hongxia, LI Jun. Construction of civil aircraft assembly process knowledge base system based on ontology [J]. Industrial Control Computer, 2018, 31(4): 122-123. [15] QIU Hongjun, TAO Hua, GAO Xiaobing, et al. Aircraft assembly process design knowledge ontology [J]. Journal of Northwestern Polytechnical University, 2007, 25(1): 32-36. [16] ZHANG Faping, LI Li. Research on business process knowledge push method based on multi-dimensional context model [J]. Journal of Computer-Aided Design & Computer Graphics, 2017, 29(4): 751-758. [17] ZHANG Guangzhi. Six Lectures on Archaeology [M]. Beijing: SDX Joint Publishing Company, 2013: 57. [18] CUI Shuai. Analysis of Knowledge Representation from the Perspective of Non-Classical Logic [D]. Taiyuan: Shanxi University, 2018. [19] Xinjiang Institute of Cultural Relics and Archaeology. Chawuhu in Xinjiang—Excavation Report of a Large Clan Cemetery [M]. Beijing: Oriental Press, 1999. [20] Institute of Archaeology,

Chinese Academy of Social Sciences. Chinese Archaeology: Neolithic Volume [M]. Beijing: China Social Sciences Press, 2010. [21] LIANG Siyong. Neolithic pottery from the Xiyin Village prehistoric site in Shanxi [J]. Sanjin Archaeology, 1996(00): 300-338, 351-354. [22] LI Ji. Prehistoric Remains from Xiyin Village [M]. Beiping: Tsinghua University Research Institute, 1927. [23] Qinghai Provincial Cultural Relics Management Office. Qinghai Liuyan [M]. Beijing: Cultural Relics Publishing House, 1984. [24] ZHOU Mingquan, FAN Yachun, GENG Guohua. A shape description method for 3D models based on spatial symmetry transformation [J]. Acta Electronica Sinica, 2010, 38(4): 853-859. [25] QIAN Jianjun. Research on semantic-based Thangka image annotation and retrieval [D]. Lanzhou: Northwest Minzu University, 2010. [26] WANG Haofen, QI Guilin, CHEN Huajun. Knowledge Graph: Methodology, Practice and Application [M]. Beijing: Publishing House of Electronics Industry, 2019: 193-195. [27] ZHANG Lingyu, CHEN Shuxin, ZHANG Xin. A context-based semantic mapping method [J]. Application Research of Computers, 2014, 31(10): 2990-2993, 2997.

Author Contributions: Han Muzhe: Proposed research ideas, designed research plan, conducted experiments and analysis, wrote paper; Gao Jinsong: Designed research plan, provided guidance, revised paper; Li Yu: Collected and organized data, participated in experiments, participated in paper writing and revision.

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