OBSKP: Oracle Bone Studies Knowledge Pyramid Model With Applications

Jing Xiong*, Tao Guo**, Guoying Liu*** and Yining Chen****

* School of Computer Science, Qufu Normal University, Rizhao, China 276826;

** **** School of Computer and Information Engineering, Anyang Normal University, Anyang, China 455000;

* ** *** **** Key Laboratory of OBI Information Processing, Ministry of Education,
Anyang Normal University, Anyang, China 455000;

* *** Engineering Research Center of Oracle Bone Culture Intelligent Industry, Anyang, China 455000

Jing Xiong received a Ph.D. degree in Computer Application Technology from Ocean University of China in 2010. His research interests include Oracle Bone Inscriptions, knowledge graphs and NLP. He is currently a professor at Qufu Normal University and is engaged in research in the Key Laboratory of Oracle Bone Inscriptions Information Processing, Ministry of Education of China. He is a senior member of CCF.

Tao Guo received a Ph.D. degree in the School of Computer Science from Xidian University in 2009. His research interests include artificial intelligence, automation and fuzzy control. He is currently a professor at Anyang Normal University and is engaged in research in the Key Laboratory of Oracle Bone Inscriptions Information Processing, Ministry of Education of China.

Guoying Liu received a Ph.D. degree in geoscience and remote sensing from the State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University in 2009. His research interests include machine learning and computer vision. He is currently a professor at Anyang Normal University and is engaged in research in the Key Laboratory of Oracle Bone Inscriptions Information Processing, Ministry of Education of China. He is the head of the Engineering Research Center of Oracle Bone Culture Intelligent Industry.

Yining Chen is an undergraduate of the School of Computer and Information Engineering, Anyang Normal University. Her research interests include Oracle Bone Inscriptions and Chinese information processing. She is now a research intern at the Key Laboratory of Oracle Bone Inscriptions Information Processing, Ministry of Education of China.

Xiong, Jing,Tao Guo, Guoying Liu and Yining Chen. 2022. "OBSKP: Oracle Bone Studies Knowledge Pyramid Model with Applications." *Knowledge Organization* 49(7): 483-495. 24 references. DOI:10.5771/0943-7444-2022-7-483.

Abstract: Oracle Bone Studies (OBS) benefit a lot from the introduction of computer science and artificial intelligence (AI), which changes the traditional way of organization of OBS knowledge, and results in a rapid accumulation of important research data. However, these data are not fully explored, and the processing faces challenges in the storage, representation, sharing and reuse of OBS big data, due to the lack of knowledge organization and management tools and knowledge application models for OBS. To address the two outstanding problems faced by OBS research bottleneck: high dependence on expert knowledge but low degree of knowledge sharing and reuse; poor computability and interpretability due to qualitative analysis. A new knowledge organization model carved from multiple dimensions such as refinement, fragmentation, conceptualization, networking, vectorization, and multimodal fusion is proposed. The OBS Knowledge application Pyramid (OBSKP) model is proposed from the perspective of Computational Oracle Bone Studies (COBS). The OBSKP shows the direction for the OBS research in the era of AI and provides the storage and representation standards and specifications for the subsequent OBS research data. Moreover, several important applications based on the OBSKP are presented.

Received: 20 July 2022; Revised 7 September 2022; Accepted 30 September 2022.

Keywords: Oracle Bone Inscriptions, knowledge organization, knowledge graph, multi-modal data

1.0 Introduction

Since their discovery in 1899, Oracle bone inscriptions (OBI) have been studied for more than 120 years, and the research on OBI has gradually formed a renowned international science, Oracle bone studies (OBS). Over these years, an increasing number of participations from institutions and individuals in OBS research obtained fruitful results: a large number of photos, rubbings, and facsimiles sorted out, and tens of thousands of papers and descriptions published. In particular, the introduction of computer science and technology generates new data forms such as images, 3D models, audio, video, animation and so on. These basic data and literature of OBS bear the characteristic of big data and multimodality (Xiong et al. 2019; Xiong et al. 2021). Currently, Xiong et al. collected and sorted out large-scale basic data and built a knowledge-sharing platform (Xiong et al. 2019) for OBI big data. The platform includes 152 descriptions, 234662 pictures, 33369 literature, and 4697 single OBI characters, with open access at http://jgw.aynu.edu.cn/.

Although this platform greatly facilitates the OBS research, it still fails to break through the bottleneck of the more than 4500 OBI characters discovered so far, about 3000 of which have not been recognized. The contradiction between the difficulty of OBS research and the small number of experts has become increasingly prominent. It takes 10-20 years to train an expert, so it is particularly important to find a method of knowledge organization that is conducive to knowledge sharing and reuse. However, there are two outstanding problems in the traditional methods of organizing OBS knowledge: 1) the research is highly dependent on experts' empirical knowledge, but the sharing and reuse of such empirical knowledge is extremely low; 2) based on qualitative analysis, the results are poorly interpretable, and the knowledge organization is not conducive to computer understanding and processing.

These long-standing challenges have created an urgent need for new ways of knowledge organization in traditional OBS research, the main needs of which are:

1. OBI experts badly need a strong knowledge network to help them find relevant and useful information from a large number of documents scattered in diversified sources.
2. The study of OBS needs a lot of domain knowledge and knowledge reasoning since computer technologies help little with cognitive intelligence in OBS research. It is necessary to organize OBS knowledge at an abstract conceptual level.
3. The study of OBS cannot stand apart from the decision-making of experts, but these analyses are mostly qualitative and easily lead to inconsistency. Thus, current research methods call for convincing quantitative analysis.
4. The study of OBS requires full consideration of multiple modal data, because the current study treats images and texts separately by image processing or text mining, thus their relations are lost in processing.

Taking the detection and recognition of OBI as an example, the difficulty of automatic OBI detection and recognition based on single modality lies in the high cost of data annotation, but the low detection/recognition accuracy (Xing et al. 2019). Furthermore, the processing of Oracle bone/shell slices suffers from high noise conditions (Shi et al. 2014), and incompleteness of OBI characters on the edge of the fragments of the Oracle bone/shell slices. The strokes of these OBI characters are easily mistaken for the natural texture of the slices (Xing et al. 2019). In addition, when the image background contains a large amount of flake noise or the color of the OBI characters is close to the background, it is difficult to obtain high-quality image segmentation (Ying et al., 2018). Traditionally OBI experts seek some auxiliary resources for supplementation and verification. These resources usually appear in different forms, such as interpretations and facsimiles. The combination of these modalities will aid the verification, and hopefully help to break the bottleneck.

With the deepening study of OBS, more and more topics have emerged, the solutions to which call for the integration of more disciplines, technologies, and methods. The acceleration of intersection and integration of the disciplines have inspired a new stage of OBS research. Figure 1 shows the schematic diagram of the development of OBS promoted by interdisciplinary integration.

It is worth noting that there are three milestones in this development process:

1. In 2000, the concept of Oracle Bone Inscriptions Information Processing (OBIIP) was put forward, which marks the beginning of the real and systematic use of computer science in OBS. Before that, sporadic computer-aided studies existed (Chou and Opstad 1973; Tong et al. 1977), but the main function of the computer was to store OBI data (e.g., using Excel and Word) and display them (e.g., browsing pictures and using PDF). Later on, a large number of OBS studies were carried out with computers. OBIIP becomes a typical cross-field between social science and natural science.
2. In 2015, the concept of Computational Oracle Bone Studies (COBS) (Jing et al. 2015; Xiong et al. 2015) was

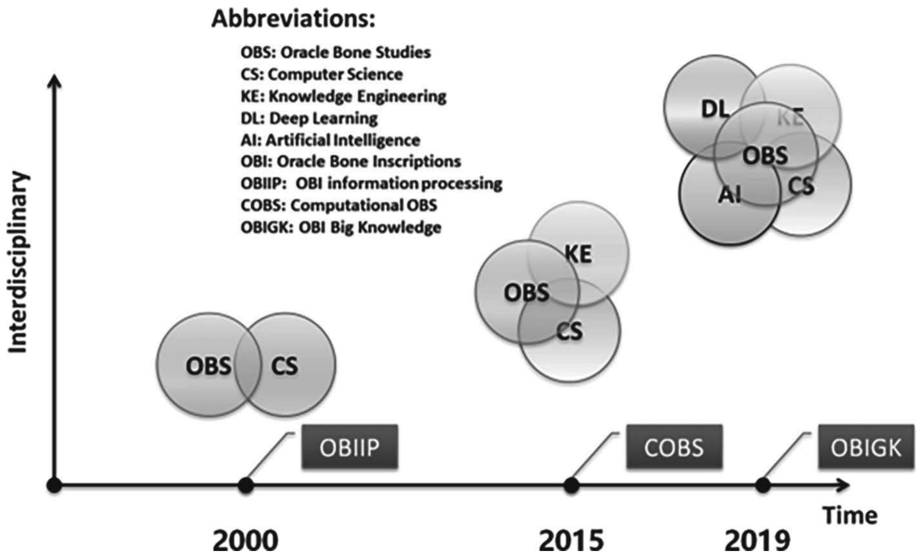

Figure 1. Interdisciplinarity is closer and closer in OBS research; in particular the proportion of computer science is getting heavier.

put forward. The management and reuse of expert knowledge is an important issue in OBS. Due to the difficulty in OBI study, only a few experts can interpret them at present. Therefore, knowledge engineering is integrated into OBIIP to realize knowledge representation and manage the corpus. With the help of experts, we annotated the OBI corpus, constructed the OBI language models and ontologies, and extended the application of computational linguistics to the field of OBS since its corpus characteristics are different from that of modern Chinese.

3. The concept of Oracle Bone Inscriptions big Knowledge (ObigK) emerged in 2019. Thanks to big data, deep learning, AI and other technologies, the OBI big data platform Xiong et al. built (Xiong et al. 2019) has played an important role. The traditional OBS research has gradually formed a paradigm of data-driven and knowledge-driven, and the research focus shifted from qualitative methods to quantitative methods. Compared with previous empirical methods, more research conclusions are drawn based on computer algorithms. The rapidly accumulated multi-modal data of OBS constitutes a large-scale knowledge graph (Xiong et al. 2021), which provides knowledge services for OBS research, and the OBI big data platform is transformed into a big knowledge platform.

As can be seen from Figure 1, OBS involves more disciplines, and the relationship between them is becoming closer, which increasingly reflects the engineering characteristics, and there is an urgent need for systematic and scientific knowledge organization and management methods. This is also the motivation of this study.

The main contributions of this paper are as follows:

1. Based on OBI big data platform, an Oracle Bone Studies Knowledge Pyramid (OBSKP) model is proposed. It describes the basis and application of OBIIP from the data layer (digitalization), presentation layer (knowledgization), and application layer (intellectualization).
2. Transforms the OBS study from traditional qualitative analysis to a quantitative analysis-based approach supplemented by qualitative decision-making. It increases the automation of information processing and reduces the burden on the OBI experts to focus on advanced decision-making activities. It also improves the interpretability of the research conclusions.
3. From the perspective of Knowledge Engineering (KE), the traditional study of OBS is improved, by emphasizing the refinement, fragmentation, conceptualization, networking, vectorization, and multimodal fusion of knowledge. The new approach to knowledge organization facilitates the sharing and reuse of OBS knowledge.

The rest of this paper is organized as follows: Section 2 introduces the relevant research. Section 3 introduces the OBI big data platform and OBSKP model. Section 4 introduces two typical application examples of the pyramid model. Finally, section 5 introduces our conclusions and points out the next research work.

2.0 Related work

Knowledge Organization (KO) is about activities such as document description, indexing and classification performed in libraries, databases, archives etc (Hjørland 2008). There are various methods of knowledge organization including classification systems, facet-analysis, information retrieval tradition, user-oriented views, bibliometric, domain analytic and others (e.g. ontological approach (Herre 2013), and genre and activity-based approach (Andersen 2015)).

Since the OBI was unearthed, experts and scholars began to study the knowledge organization of OBS. The early knowledge organization of OBS was based on the classification system. The important event was the publishing of *Chia-ku-wen tuan-tai yen-chiu li* by Tung Tso-pin (Tso-pin 1933), a famous OBI expert, in which he put forward ten dating criteria, namely, genealogy, titles, diviners, pit location, foreign statelets, persons, divination topic, grammar, epigraphic form, and calligraphy. According to these standards, the OBI was divided into 5 periods, covering the reigning periods of twelve hereditary kings of Shang Dynasty, each period including between two and four kings. Later, Li Xueqin and Yushang (1996) wrote *Periodization research of oracle bones & tortoise shells in Yin Ruins* and proposed the introduction of archaeological stratigraphic evidence into the periodization of OBI. It is also a classification method. This way of knowledge maintenance is full volume update, which generates a lot of redundancy and has a long update cycle.

With the involvement of computer science and information technology, OBS knowledge was organized to support information retrieval, such as OBI database (Mao 2010) and OBI semantic dictionary (Feng et al. 2014), but they could not solve the problem that the retrieved objects involved OBI images.

Since literature is an important data source for OBS research, the bibliometrics based mapping of knowledge domains (MKD) approach is used for OBS knowledge representation and organization, and plays an important role in constructing the OBS knowledge graph. (Xiong et al. 2019; Xiong et al. 2020). However, MKD cannot organize the multimodal knowledge of OBS.

Ontology can provide formal specifications and harmonize the definitions of concepts used to represent the knowledge of specific domains (Herre 2013). It is therefore widely used in knowledge organization (Marcondes 2013; Machado 2021). To fully define the concepts and express the relations between OBS knowledge entities, Xiong et al. (2011) studied the OBI ontology and built the OBS knowledge graph based on this ontology and OBS multimodal data (Xiong et al. 2020; Xiong et al. 2021). Although these methods can organize and manage OBS knowledge to some extent, they play little role in expert knowledge reuse, facilitating reuse by computer algorithms, or assisting expert decision-making. Moreover, in the environment of proliferating data and increasing interdisciplinarity, there is a lack of a systematic and comprehensive model for the integrated management of data, knowledge and applications of OBS. In knowledge organization domain the existing pyramid model data-information-knowledge-wisdom (DIKW) (Bates 2005; Baskarada and Koronios 2013) was widely used, but nowadays it is generally considered unsatisfactory (Frické 2019). There is still a need for development of knowledge services and applications. In addition, OBS in the new era involves unprecedented engineering characteristics such as big data, big knowledge, and multimodality, so it is necessary to build a suitable knowledge organization model according to the characteristics of domain knowledge and application needs.

3.0 OBI big data platform and knowledge pyramid model

After more than 20 years of research efforts and nearly 5 years of development, we built an open OBI big data platform entitled Yinqi Wenyuan (http://jgw.aynu.edu.cn/). It is a knowledge-sharing platform integrating the OBI font library, OBI description library, and OBI literature library. This is so far the largest and most comprehensive OBI basic data service platform. With the continuous increase of the data and improvement of the platform functions, more and more applications begin to serve the research of OBIIP. Based on this platform, we propose the OBSKP model for applications, as shown in Figure 2.

As can be seen from Figure 2, the model is composed of three layers: digitalization, knowledgization and intellectualization. The bottom is the data layer, which is derived from the three foundations of Yinqi Wenyuan: font library, description library, and literature library. The middle is the presentation layer, which organizes, stores, analyzes, and represents the OBI digital resources. This part forms the ObigK. The top is the application layer that is based on the ObigK. The application layer provides users with intelligent knowledge services, including, but not limited to, OBI full-text retrieval, knowledge recommendation, intelligent Question Answering system (QA), etc.

In OBSKP we implement the following functions that work as the foundations of this study.

1. We deploy the OBI font library, description library, and literature library on a unified platform to meet the different needs of OBS research. The three modules are self-inclusive but are integrated with each other, which greatly facilitates researchers' access to literature and descriptions.

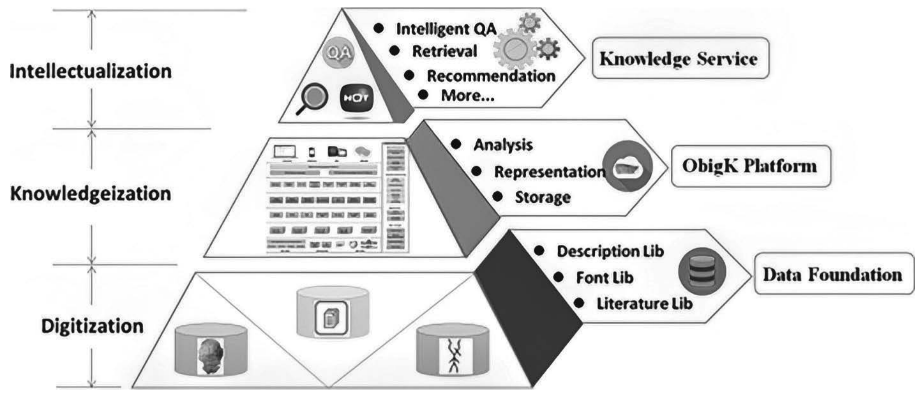

Figure 2. OBSKP model.

2. We use fragmented annotation to label the OBI image-character in the literature, so as to avoid the failure of retrieval of the current literature based on the OBI image-character.
3. According to the OBI font library, we replace the clerical simulated characters and unrecognized OBI characters in the OBI interpretation texts and realize the full-text literature retrieval based on OBI characters.
4. We build OBSKG as foundation of intelligent knowledge service, which establishes the relationship between knowledge entities and provide the function of knowledge reasoning. This function can aid the experts in literature study and OBI interpretations, and also provide OBI enthusiasts with intelligent QA services.

In the OBSKP model, the three layers respectively provide basic data, big knowledge platform and intelligent knowledge services for OBIIP. The rest of this section describes in detail the various layers of OBSKP.

3.1 Digitization layer

The task of the digitization layer is to collect and organize the basic OBI data, for establishing databases including OBI literature library, font library and description library. At present, the Yinqi Wenyuan big data sharing platform we built manifests the significant progress toward OBI digitalization, and the OBI basic data it holds are shared by users all over the world.

The literature library aims to collect and categorize the published literature since the discovery of OBI as texts or images. At present, its capacity has reached 33369. As shown in Figure 3.

The character glyphs of the font library come from the handwritings of OBI experts, and are coded by the sixth plane of Unicode system. At present, its capacity has reached 4697 OBI characters, as shown in Figure 4.

The descriptions library aims to collect all the published books and related pictures since the discovery of OBI. The main storage format is high definition (HD) scanned image. Its current scale has reached 152 books and 234,662 images, as shown in Figure 5.

3.2 Knowledgeization layer

The knowledgization layer is the core of knowledge organization since it is responsible for extracting knowledge factors from the underlying data and making them into a form that can be easily understood and shared by computers through knowledge representation methods. It adopts the organization approach based on knowledge association, extracting OBS knowledge factors from the digitization layer, analyzing and synthesizing them, and acquiring knowledge associations to generate higher-level compound knowledge. This layer focuses on the refinement, fragmentation, conceptualization, networking, vectorization, and multimodal fusion of OBS knowledge. Its purpose is to build a large-scale computable multi-dimensional knowledge network, to realize the sharing and reuse of OBS knowledge.

Refinement refers to the splitting and processing of basic data to extract fine-grained knowledge factors. That is, the OBI image is segmented into units of OBI character images are extracted (see Figure 6 (a)); Split the OBI characters in the font according to the radical and extract semantic components (see Figure 6 (b)); Separate and sort the images, interpretation texts and modern Chinese Texts in the literature (see Figure 6 (c)).

Fragmentation is mainly aimed at documents in the OBI literature, because they are usually published in the form of a mixed arrangement of image and text. The purpose of fragmentation is to organize the OBI image-characters in the documents, as shown in Figure 6(c). Here the OBI image-character refers to images formerly embedded in the published document as a way to input the character. Other occasions using OBI image-character include clipping char-

Figure 3. The literature library in the digitization layer.

Figure 4. The font library in the digitization layer.

acters from the rubbings, or storing and displaying handwritten OBI characters.

The implementation of fragmentation is as follows: separate image and text in the mixed document, extract the images of OBI characters in the document to form an image database, name each image-character with a unique identifier, and record its position in the original document. Here, the position recorded in the original document refers to the texts in the range around the position of the recorded image-character. For example, if the range is 5, there are 5 characters (including punctuation) before and after the image-character. If the OBI character corresponding to the image-character is a recognized one, its modern Chinese character will be recorded at the same time, and finally all the information will be stored as an XML file. An example fragment is as follows:

```xml
<paper>
    <title>谈两版圆体类龟腹甲的释读</title>
    <author>王子杨</author>
    <keyword>甲骨文;龟腹甲;文字释读</keyword>
    <publisher>出土文献</publisher>
    <publishtime>2011.03</publishtime>
<charImgs>
    <charImg>
        <hanzi>戠</hanzi>
```

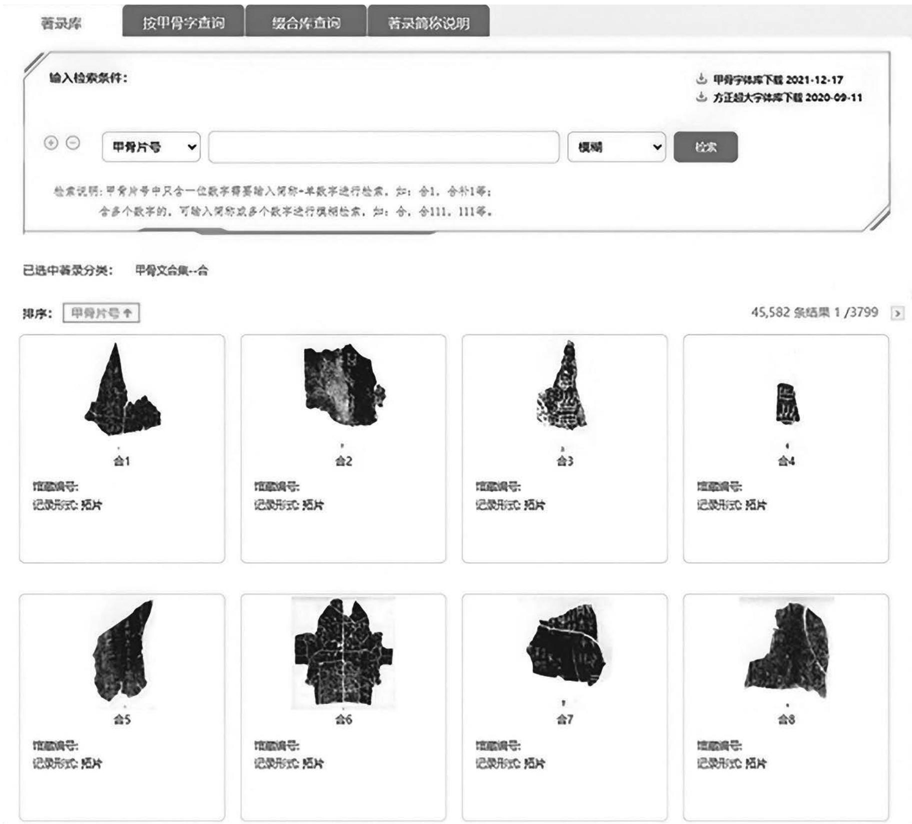

Figure 5. The description library in the digitization layer.

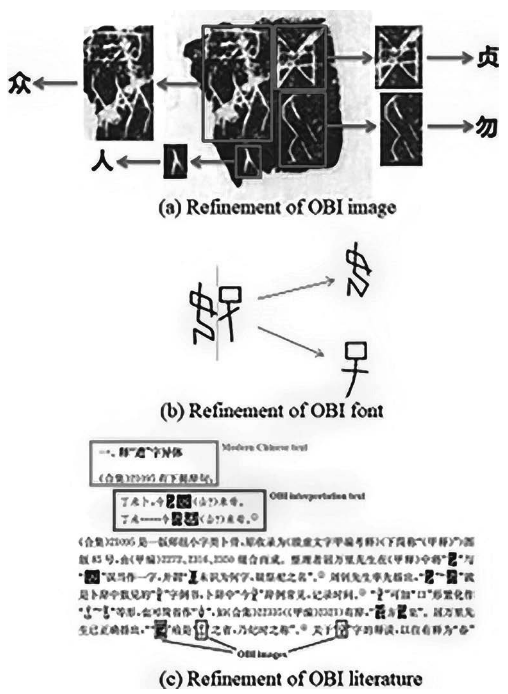

Figure 6. Refinement of OBS knowledge.

```
    &lt;winsize&gt;5&lt;/winsize&gt;
    &lt;type&gt;手写&lt;/type&gt;
    &lt;img src="imgs/谈两版圆体类龟腹甲的释读_095.jpg"&gt;谈两版圆体类龟腹甲的释读_095</>
    &lt;comefrom&gt;《合集》21885&lt;/comefrom&gt;
    &lt;before&gt;，形体作“&lt;/before&gt;
    &lt;after&gt;”，右部明&lt;/after&gt;
  &lt;/charImg&gt;
  ...
  &lt;charImg&gt;
    &lt;hanzi&gt;unknown&lt;/hanzi&gt;
    &lt;winsize&gt;5&lt;/winsize&gt;
    &lt;type&gt;拓片&lt;/type&gt;
    &lt;img src="imgs/谈两版圆体类龟腹甲的释读_107.jpg"&gt;谈两版圆体类龟腹甲的释读_107</>
    &lt;comefrom&gt;《合集》22414&lt;/comefrom&gt;
    &lt;before&gt;414 有“&lt;/before&gt;
    &lt;after&gt;”字，刻写&lt;/after&gt;
  &lt;/charImg&gt;
  ...
&lt;charImgs&gt;
&lt;/paper&gt;
```

Conceptualization refers to abstracting the things involved in the field of OBS to obtain some semantic models independent of the environment. We use ontology to realize conceptualization, because ontology as a formal, explicit specification of a shared conceptualization makes it possible to share and reuse knowledge. We have constructed the common knowledge ontology, description ontology, literature ontology and interpretation ontology of OBI.

Networking refers to building a large-scale knowledge network based on the semantic association between the knowledge factors after the refinement stage. We achieved networking by constructing an OBS knowledge graph and followed the semantic models of the conceptualization stage during the construction process.

Vectorization refers to the transformation of large-scale OBI data into vector form, including image vector, word vector, and graph vector, intending to facilitate calculation by computer algorithms, especially deep learning models. Vectorization is an important embodiment of transforming the study of OBS from qualitative method to quantitative method.

Multimodal fusion refers to the comprehensive consideration of the features of multi-source and heterogeneous OBI data, to remedy the deficiency that current OBI information processing separates image modality and text modality. It is achieved by constructing a multimodal knowledge graph (Xiong et al. 2021) and vector hybrid embedding. A screenshot of a multimodal fusion fragment is shown in Figure 7.

The tasks of the knowledgeization layer also include storage and semantic analysis. In terms of knowledge storage, we adopt different methods for different types of data. Specifically, the image data (including photos, rubbings, and facsimiles) are normalized and uniformly encoded and annotated with their attributes. Then they are stored in the relational database and displayed on the web page.

Metadata is extracted from the OBI literature, and the relational database is constructed. Some OBI characters in the literature have no corresponding modern Chinese characters and cannot be input into texts, therefore the images of these OBI characters obtained from the image database is embedded in the literature.

For the OBI font library, the OBI characters are encoded in Unicode and important information is stored, such as simplified Chinese characters, traditional Chinese characters, and Pinyin.

The networked knowledge factors are stored in the graph database such as Neo4j.

The task of semantic analysis is to conduct deeper level relevant analysis and knowledge reasoning based on the results of knowledge representation, such as word frequency analysis, semantic scene prediction, ontological reasoning, OBI fragment conjugation verification through interpretation text, and citation analysis, etc.

3.3 Intellectualization layer

The intellectualization layer is built on the ObigK to provide services for OBS applications. By using AI technologies such as big data, deep learning and knowledge graph, we aim to provide intelligent knowledge services for OBS experts, scholars and enthusiasts. These services include full-text retrieval, intelligent QA, knowledge recommendation and more. These are also the innovations of OBS research in the new era because they extend the traditional OBS study to the field of application. The rest of this section will deal with these intelligent applications at greater length.

OBI full-text retrieval includes OBI full-text description retrieval and OBI full-text literature retrieval.

The full-text retrieval of OBI description includes Oracle bone/shell retrieval, OBI interpretation retrieval, and mutual retrieval. Specifically, Oracle bone/shell retrieval and OBI interpretation retrieval can be realized according to the input keywords such as Oracle bone/shell slice number, OBI interpretation, and OBI character. Oracle bone/shell retrieval returns the target Oracle bones or shells and the full text of the interpretation corresponding to the bone/shell; OBI full-text interpretation retrieval returns the target sentences and gives the links of Oracle bone/shell slices containing these sentences. When using OBI characters as input keywords, there are several input methods: 1) input the modern Chinese characters corresponding to the OBI characters (if the OBI characters are recognized); 2) handwriting input method; 3) select keywords from the OBI font library; 4) upload the image of OBI character.

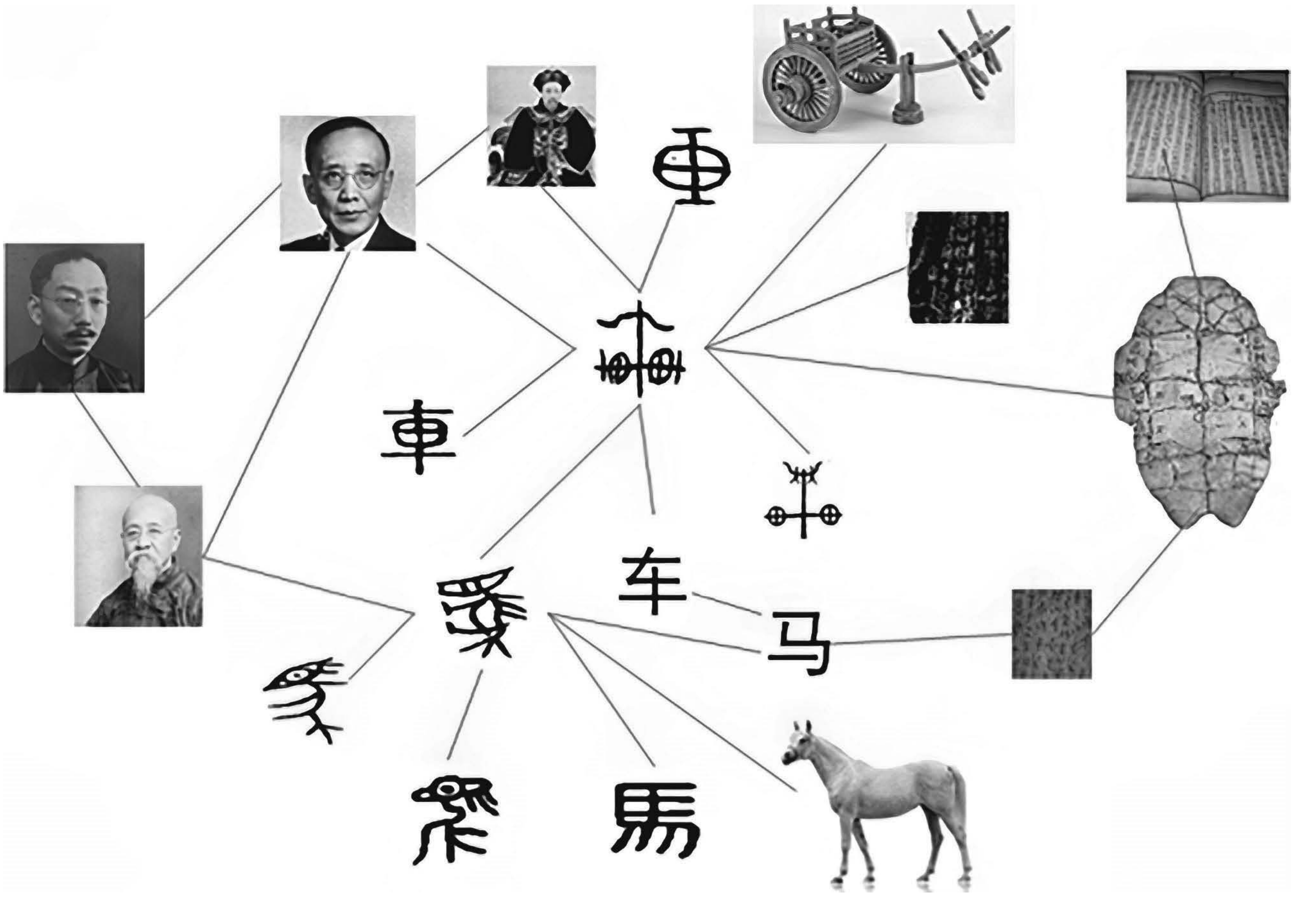

Figure 7. An example of multimodal fusion.

The full-text retrieval based on OBI characters is an important contribution to the field of OBS. When users input OBI characters or their corresponding image-characters to search, text retrieval or image retrieval technology will be called respectively. By using the annotation information in the XML file described in section 3.2, the search results will be returned, and the matched OBI characters will be located in the literature.

The intelligent QA system is implemented based on the knowledge graph, which is processed as follows: the natural language processing (NLP) technology is used to analyze the input item, extract the named entities and relation in the searching, and obtain the triple <h, r, t> (where h represents the head entity, t represents the tail entity, and r represents the relation between h and t). The extracted entity (h or t) is matched with the entity in the knowledge graph, and the extracted relation r is matched with the relation in the knowledge graph. If the matching is successful, it returns the answers; If there is no matching result, it acquires candidate answers through reasoning by the link prediction of knowledge graph. The specific operation is to use two elements of <h, r, t> to predict the third one, such as entity prediction t using h and r (see Figure 8 (a)), relation prediction r using h and t (see Figure 8 (b)), etc.

Content-based recommendation is used in our OBI knowledge platform, which is different from former recommendation systems based on metadata (authors, affiliations, and references). The specific implementation adopts the graph embedding methods, that is, to input the elements h, r, and t in the triple <h, r, t> into the graphical neural network, and to design the triple scoring function in combination with the emotional analysis of the academic viewpoint of OBI (that is, the difference in opinions to or against some research idea is an important new research hotspot in COBS field).

4.0 OBSKP application examples

There are many applications based on OBSKP, including OBI detection and recognition, automatic conjugation, semantic prediction, intelligent QA, etc. In this section, we only list two examples for explanation.

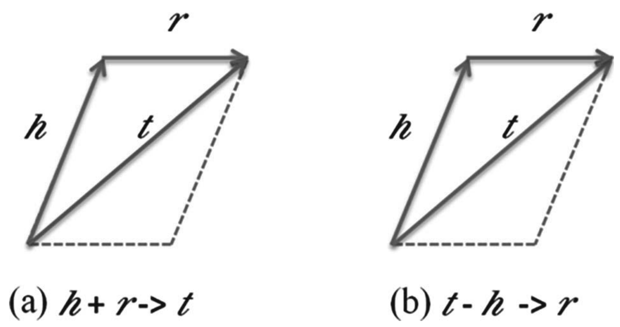

Figure 8. Examples of entity prediction and relation prediction.

4.1 Computer-aided oracle-bones conjugation

Oracle bone/shell conjugation is to splice the broken pieces of bones/shells together. During thousands of years of burial, a large number of Oracle bones/shells were broken into several pieces due to materials aging, the pressure of the strata, infiltration of water, and archaeological excavation. This brings many difficulties to OBS research. The OBI obtained through the conjugation of Oracle bones/shells is of high historical value and importance.

After the fragments of Oracle bones/shells are conjugated, the relations between the OBI characters can be found. Once the original Oracle bones/shells are restored, it becomes an important historical material for understanding the society of the Shang Dynasty. In the study of Oracle bone/shell conjugation, we need to comprehensively consider multiple data sources and features, such as photos, rubbings, facsimiles, conjugation pieces, conjugation methods, conjugation time, descriptions, academic papers, etc. Other necessary information to consider includes the shape of the Oracle bone/shell pieces, the interpretation texts, the OBI characters, and the stage divination.

Engaging in Oracle bone/shell conjugation research requires the experts' long-term scientific accumulation and keen insight into OBI literature. More often, there are many conjugation clues hidden in the vast amount of OBI basic data and literature. In addition, in the process of conjugation, we should combine various data for mutual verification and supplement. In the case that some rubbings are missing in conjugation, they can usually be replaced by photos, copies, or 3D images. In other cases, if the shapes can be combined correctly for candidate fragments, then the interpretation text is usually used for verification. There are usually implicit connections between these clues. Once the key points are found, the efficiency and accuracy of conjugation can be greatly improved. After these clues are stored in the OBSKG, valuable clues can be found through the node path search, to help OBI experts conjugate Oracle bone/shell pieces.

It is also important to exclude incorrect candidate conjugation results of automatic conjugation. On the one hand, the conjugated fragments are helpful to complete the OBI interpretation texts; on the other hand, the OBI interpretation can also verify the correctness of candidate conjugations. Purely relying on the OBI experts for interpretation both has many difficulties and is quite costly. The knowledge graph has great advantages in abnormal detection, carried out by algorithms such as retrieval, reasoning, consistency checking, outlier detection, and community discovery. Similarly, this advantage can be introduced into the study of Oracle bone/shell conjugation, and the detection of abnormal conjugation can provide clues for OBI experts. An example of Oracle bone/shell conjugation is shown in Figure 9.

The example shown in Figure 10 is a manual conjugation, in which the rubbings and facsimiles manifest the comprehensive use of the multiple modality data in OBS. From the shape and outline of the Oracle bone/shell pieces, they coincide well, which seems to be a correct conjugation result. However, by combining the interpretation texts of these two pieces, it can be found that there are errors in syntax and semantics, so it is judged as incorrect. This conclusion not only shows the importance of comprehensive utilization of multimodal OBI data, but also verifies the effectiveness of OBSKG in the study of Oracle bone/shell conjugation.

4.2 OBI knowledge QA assistant

The purpose of building OBSKG is to establish links between OBS knowledge entities and form a large-scale knowledge network. It can summarize the knowledge chain from the complicated mass of knowledge points according to certain logic, so as to realize knowledge retrieval and intelligent services. Intelligent QA can be realized by finding paths in the knowledge graph. On the basis of OBSKG, we developed an intelligent QA assistant named Wen Xiaoyuan. At present, it is a lightly complex OBS knowledge QA, as is shown in Figure 10.

The implementation method of intelligent QA assistant is described in section 3.3. The entities and relations are parsed from the user's questioning sentences for matching in the knowledge graph. The knowledge graph query state-

Figure 9. An example of oracle bone/shell conjugation.

Figure 10. OBS QA assistant Wen Xiaoyuan.

ments are generated and executed, and finally the possible answers are returned. We use the Neo4j cypher to realize knowledge graph query. At the same time, according to the characteristics of the field of OBS, we also adopt the following methods:

1. The commonly used word segmentation tools fail to achieve high accuracy in OBS field, so we have established an OBI dictionary to manage OBI terms and common collocations.
2. Usually, the nouns in the question entered by the user are the candidate elements for the entities, and the verbs are the candidate elements for the relations. In OBSKG, however, relations are sometimes expressed by nouns or polysyllabic words. To improve the quality of QA, we maintain a list of relation vocabulary that is formed by

extracting terms from the relation set of OBSKG, with their synonyms considered.

5.0 Conclusion

To break the bottleneck of OBS research, we reconsider the knowledge organization of OBS from the perspective of computer science, whose main goal is to address the sharing and reuse of expert knowledge, as well as the problems of computability and interpretability. A knowledge representation and organization method suitable for OBS is proposed in several dimensions, including refinement, fragmentation, conceptualization, networking, vectorization, and multimodal fusion. A comparison of this method with existing ones is shown in Table 1.

Based on the OBI big data-sharing platform, we propose an OBSKP model to serve OBIIP research from the perspective of knowledge engineering. We introduce the tasks of each layer of the pyramid model in detail. This model not only considers the basic research of OBS, but also pays attention to its application research, which has not been involved in previous studies. At the same time, we explain the feasibility and effectiveness of the model in the application research in the new era of OBS through two specific cases. However, at present, the model lacks effective quantitative indicators to evaluate its maturity. In future research, we will design and verify some quantitative indicators to improve the performance of the pyramid model.

References

Andersen, J. 2015. "Re-Describing Knowledge Organization - A Genre and Activity-Based View." In *Genre Theory in Information Studies,* edited by Jack Andersen. Bingley: Emerald, 13-42.

Baskarada, S. and A. Koronios. 2013. "Data, Information, Knowledge, Wisdom (DIKW): A Semiotic Theoretical and Empirical Exploration of the Hierarchy and its Quality Dimension." *Australasian Journal of Information Systems* 18 (1). https://doi.org/10.3127/ajis.v18i1.748

Bates, M. J. 2005. "Information and Knowledge: An Evolutionary Framework for Information Science." *Information Research* 10, no. 4. www.informationr.net/ir/10-4/paper239.html

Chou, H. and D. G. Opstad. 1973. "Computer Matching of Oracle Bone Fragments: A Preliminary Report on a New Research Method." *Archaeology* 26: 176-81.

Feng, G., T. Xiping et al. 2014. "Research on Semantic Dictionary Construction of Oracle Bone Inscriptions." *Journal of Anyang Normal University* 5: 43-7.

Frické, M. 2019. "The Knowledge Pyramid: The DIKW Hierarchy." *Knowledge Organization* 46: 33-46.

Herre, H. 2013. "Formal Ontology and the Foundation of Knowledge Organization." *Knowledge Organization* 40: 332-9.

Hjorland, B. 2008. "What is Knowledge Organization (KO)?" *Knowledge Organization* 35: 86-101.

Jing, X., G. Feng et al. 2015. "Research on Semantic Mining for Large-Scale Oracle Bone Inscriptions Foundation Data." *Data Analysis and Knowledge Discovery* 31(2): 7-14.

Machado, L. M. O. 2021. "Ontologies in Knowledge Organization." *Encyclopedia* 1 (1): 144-151. https://doi.org/10.3390/encyclopedia1010015

Mao, J. 2010. "Proposal and Construction of Full-Text Database of Oracle-Bone Inscriptions." *Research in Library Science* 12: 37-38.

Marcondes, C. H. 2013. "Knowledge Organization and Representation in Digital Environments: Relations Between

	traditional method	ours
research means	manual-based	automation-based, manual confirmation
knowledge structure	planar structure, linear association	multi-dimensional structure, network association
research difficulty	long cycle and great difficulty	short cycle and less difficulty
knowledge maintenance	full volume update	incremental update
data analysis	qualitative	quantitative analysis, qualitative decision making
modality	unimodal	multimodal
computability	poor computability and interpretability	strong computability and good interpretability
research type	fundamental research	fundamental and applied research
level of sharing and reuse	low	high
technologies and methods	human intelligence	artificial intelligence

Table 1: Comparison of methods

Ontology And Knowledge Organization." *Knowledge Organization* 40: 115-22.

Shi, X. S., Y. J. Huang et al. 2014. "Location Method of Rubbing Text of Oracle Bone Inscriptions Based on Threshold Segmentation And Morphology." *Journal of Beijing Information Science & Technology University* 29(6): 7-10.

Tong Enzheng, Zhang Shengkai and Chen Jingchun. 1977. "Guanyu yong dianzijisuanji zhuihe Shang dai jiagu suipian de chubu baogao." *Kaogu* 3:205-9

Tso-pin, T. 1933. "Chia Ku Wên Tuan Tai Yen Chiu Li." *Ch'ing Chu Ts' ai Yüan-p'ei Hsien Shêng Liu Shih Wu Sui Lun Wên Chi*. Peiping: Academia Sinica, 323-424.

Xing, J., G. Liu et al. 2019. "Oracle Bone Inscription Detection: A Survey of Oracle Bone Inscription Detection Based on Deep Learning Algorithm." In *AIIPCC '19: Proceedings of the International Conference on Artificial Intelligence, Information Processing and Cloud Computing*. New York: ACM, 1-8. doi.org/10.1145/3371425.3371434

Xiong, J., G. Liu et al. 2021. "Oracle Bone Inscriptions Information Processing Based on Multi-Modal Knowledge Graph." *Computers & Electrical Engineering* 92: 107173. DOI:10.1145/3371425.3371434

Xiong, J., L. Zhong et al. 2011. "Ontology Construction Method and its Application in Oracle Bone Inscriptions." *Journal of Wuhan University of Technology (Information & Management Engineering)* 33: 953-7.

Xiong, J., Q. Jiao et al. 2019. "Entity Discovery and Relation Extraction of Oracle Bone Studies Based on Mapping Knowledge Domains." *Information Technology and Informatization* 12: 182-6.

Xiong, J., Q. Jiao et al. 2019. "Oracle Bone Inscriptions Big Knowledge Management and Service Platform." In *Data Science. ICPCSEE 2019*, edited by X. Cheng, W. Jing, X. Song and Z. Lu. Communications in Computer and Information Science, vol 1058. Singapore: Springer. https://doi.org/10.1007/978-981-15-0118-0_10

Xiong, J., Q. Jiao et al. 2020. "Oracle Bone Studies Knowledge Graph Construction Based on Multi-Source Heterogeneous Data." *Journal of Zhejiang University (Science Edition)* 47: 131-41.

Xiong, J., X. Niu et al. 2015. "Word Segmentation for Computational Oracle Bone Inscriptions." *Metallurgical and Mining Industry* 7: 224-9.

Xueqin, L. and P. Yushang 1996. *Periodization Research of Oracle Bones & Tortoise Shells*. Shanghai: Classics Publishing House.

Ying, H. E., H. E. Xiaoju et al. 2018. "Image Segmentation of JiaGuWen Image Based on Wavelet Transform and FCM." *Journal of Tianjin University of Science & Technology* 6: 62-6. https:// caod.oriprobe.com/issues/1874179/toc.htm