Research on the Construction of Oracle Bone Inscriptions Knowledge Graph and its Application in Historical Language Analysis

Dandan Song*
Tianjin Normal University
Tianjin, 300387, China

Changhao Su
Tianjin Normal University
Tianjin, 300387, China

Abstract:As a precious cultural heritage of Chinese civilization, the study of oracle bone inscriptions is of great significance to the fields of philology, history and archaeology. This paper focuses on the construction of oracle bone inscription knowledge graph and its application in historical language analysis, focusing on the text organization, grammatical structure analysis and semantic information mining of oracle bone inscriptions. By adopting natural language processing and machine learning methods, this paper realizes the automatic annotation and deep semantic relationship mining of oracle bone inscriptions. At the same time, the construction of cross-modal knowledge graphs combined with multimodal data provides systematic technical support and multi-dimensional research perspectives for oracle bone inscription research. The research results show that the oracle bone inscription knowledge graph plays a significant role in improving the efficiency of oracle bone interpretation, exploring the social relationship network of the Shang Dynasty and promoting the digital protection of traditional culture.

Keywords: Oracle bone inscriptions; knowledge graph; historical language analysis

1. INTRODUCTION

Oracle bone inscriptions are a treasure of ancient Chinese culture and an important symbol of the development of human civilization. As one of the earliest Chinese characters, oracle bone inscriptions, with their unique artistic form and historical value, occupy an irreplaceable position in the fields of philology, archaeology and art. From its discovery to its research, oracle bone inscriptions are not only regarded as a record of history, but also a cultural symbol that transcends time and space, showing the ancient people's profound understanding and unique expression of nature, society and the universe.

From the perspective of literary art, the shape structure of oracle bone inscriptions is diverse and full of wisdom, covering pictographic, ideographic, phono-semantic and loan forms. These forms not only carry the symbolic meaning of the original pictures, but also develop into a highly abstract symbol system. The pictographic features of the oracle bone script are particularly prominent, and its character shapes often directly imitate the forms of nature. For example, the character "日" (sun) looks like the round sun, and the character "月" (moon) looks like the crescent moon. This intuitive depiction of natural forms gives oracle bone inscriptions a strong pictorial flavor, making them a visual representation of the interaction between ancient peoples and nature. In terms of font structure, the square and round changes in oracle bone script show a unique beauty. Some glyphs have strong and powerful lines and obvious symmetry; while others are flexible and interesting. For example, the character "水" has smooth lines, which seem to show the shape of flowing water; while the character "木" uses an upright shape to show the growth characteristics of trees. This font design not only makes the text expressive, but also provides the source and inspiration for later calligraphy art.

Oracle bone inscriptions are not only a carrier of historical records, but also an important part of Chinese traditional culture. It has passed down the thoughts, beliefs and lifestyles of our ancestors in a symbolic way, becoming an important link between the past and the present. The artistic features displayed by oracle bone inscriptions not only provide rich aesthetic inspiration for future generations, but also inspire modern art and design. Many artists have drawn elements from oracle bone inscriptions and incorporated them into modern calligraphy, font design, and visual art, giving this ancient writing system new vitality.

In addition, the study of oracle bone inscriptions is also of great significance to the development of anthropology, linguistics and philology. It is key material for exploring the origin of Chinese characters and provides an indispensable basis for understanding the formation of the Chinese writing system. By studying oracle bone inscriptions, we can gain a deeper understanding of many aspects of ancient society, including politics, economy, religion, and culture.

Figure. 1 The Example of Oracle Bone Inscriptions

2. THE PROPOSED METHODOLOGY

2.1 The Oracle Knowledge Graph Construction

In recent years, with the rapid development of information technology and artificial intelligence, the research on oracle bone inscription information processing has gradually entered

a stage of in-depth and systematic development, which has not only promoted the digital protection and dissemination of oracle bone inscriptions, a precious cultural heritage, but also provided new research ideas and tool support for related fields. In the study of oracle bone inscriptions, researchers have gradually explored a variety of technical means, combining multidisciplinary methods such as the computer science, linguistics, information visualization and data mining, and have achieved remarkable results, especially in the storage, analysis and application of oracle bone inscription data.

Technical achievements in oracle bone information processing

Construction of oracle bone character database

The establishment of oracle bone character database is the basic work of oracle bone informatization research. By collecting and organizing oracle bone rubbings, literature and interpretation results, researchers have created a standardized digital character database. These character databases not only contain oracle bone characters, interpretations and phonetic information, but also combine contextual data such as time, geography, and sacrifices, providing a solid foundation for multi-dimensional research on oracle bone characters. For example, tools such as the "Oracle Bone Online Dictionary" facilitate researchers to quickly retrieve and analyze oracle bone characters.

Computer-assisted oracle bone splicing

Splicing is an important part of oracle bone research, which is to piece together and restore broken oracle bone pieces. Traditional splicing methods rely on the experience and intuition of experts, which is time-consuming and labor-intensive. Through image recognition and machine learning technology, researchers have developed a computer-assisted splicing tool that can analyze the texture, text position and carving direction of oracle bone pieces through algorithms to achieve automatic matching and splicing, greatly improving research efficiency.

Oracle bone script corpus annotation and editing

In terms of corpus annotation, researchers introduced natural language processing (NLP) technology to systematically annotate the oracle bone script corpus with syntactic structure, semantic role, and contextual information. These annotated data not only help the study of oracle bone script interpretation, but also provide high-quality training data for machine translation and knowledge graph construction.

Oracle bone script machine translation

Oracle bone script machine translation is one of the hot research directions in recent years. By training deep learning models, researchers try to achieve automatic translation from oracle bone script to modern Chinese. This technology is still in the exploratory stage, but has achieved initial results in specific fields (such as common sacrificial terms and royal records).

The construction of oracle bone knowledge graph still faces many challenges, especially in expressing deep semantic relationships.

Mining of deep semantic relationships

Oracle bone knowledge not only includes direct information about glyphs and interpretations, but also contains complex social, religious and political relationships. For example, the lineage of the Shang kings, the geographical locations of various local countries, and the division of responsibilities of different Zhenren. These deep semantic relationships cannot be fully expressed by a simple "entity-attribute-relationship" model. Researchers need to introduce more complex semantic network modeling methods, combining historical background and domain knowledge to achieve higher-level knowledge expression.

Technical support for cross-domain integration

The construction of oracle bone knowledge graph cannot be separated from the theoretical and methodological support of multiple disciplines. For example, metrological citation analysis and co-occurrence analysis can help reveal the vocabulary associations and text contexts in oracle bone inscriptions; information visualization technology can visualize complex knowledge networks. In particular, based on scientific knowledge graphs (such as the MKD method), researchers can more clearly show the core structure, development process and interdisciplinary frontiers of oracle bone research.

Handling of semantic uncertainty

Since some of the contents of oracle bone interpretation are still controversial, its semantic expression has a certain degree of uncertainty. How to deal with these uncertainties in the knowledge graphs, maintaining the scientific nature of the data while allowing for reasonable interpretation space, is one of the current research difficulties. Viable solutions include using probabilistic models or the fuzzy logic techniques to support multiple possible semantic interpretations.

2.2 The Cross-modal Knowledge Graph

Cross-modal knowledge graphs use computer technology to study oracle bone inscriptions, especially the in-depth analysis of grammar, syntax and semantic information of oracle bone inscriptions, which is an important task in the intersection of modern science and technology and traditional culture. In this process, the organization and digitization of oracle bone inscriptions is the starting point of the research, laying the foundation for the subsequent construction and analysis of language models.

The compilation of oracle bone scripts not only covers the interpreted texts of the original oracle bone rubbings, but also includes a large amount of extended materials surrounding oracle bone research, such as historical documents, archaeological records, research monographs, academic reviews, and related educational materials and digital resources. The completeness and accuracy of these data directly determine the depth and breadth of subsequent research.

Integration of interpretation and literature

The compilation of oracle bone inscriptions requires the standardization of the original character forms and interpretations, while conducting in-depth analysis based on the background information in the oracle bone documents. This not only includes the interpretation of individual words, but also involves the restoration of the sentence structure and text context of the oracle bone inscriptions.

Fusion of multimodal data

In addition to the characters themselves, oracle bone script research also requires the integration of multimodal information such as the shape of the oracle bones, carved textures, and excavation sites. These data are constructed

through image processing and database, and form a multi-dimensional research system together with text interpretation.

Digitalization and Visualization

By converting the text data of oracle bone inscriptions into digital format and displaying it through information visualization technology, we can more intuitively understand the relationship network, language structure and semantic features in oracle bone inscriptions. For example, the graphical knowledge network displays the sacrificial activities, geographical distribution and social structure of the Shang Dynasty, providing researchers with a new analytical perspective.

3. CONCLUSION

Based on the needs of oracle bone inscription research, this paper explores the methods and applications of constructing oracle bone inscription knowledge graphs. First, by systematically organizing and standardizing oracle bone inscriptions, a digital corpus with characters, interpretations and contextual information as the core is constructed, which provides a solid foundation for the construction of knowledge graphs. Secondly, this paper combines natural language processing technology with deep learning models to realize grammatical structure analysis, entity recognition and semantic relationship mining, and constructs a multi-level oracle bone inscription knowledge graph. At the same time, by introducing multi-modal data fusion technology, the morphology, texture and excavation information of oracle bone inscriptions are integrated to construct a cross-modal knowledge graph, providing a multi-dimensional analysis framework for oracle bone inscription research.

The study found that knowledge graph technology not only improves the efficiency of oracle bone inscription interpretation and splicing, but also can intuitively display the complex relationship network of Shang Dynasty society, including clan genealogy, geographical distribution and religious activities. In addition, the model based on uncertainty processing proposed in this paper also shows good adaptability in dealing with interpretation ambiguity and semantic ambiguity. In the future, with the further development of information technology, the construction and application of oracle bone knowledge graph will inject new vitality into the digital protection and academic research of traditional culture, and help the inheritance and development of traditional culture in the modern context.

4. REFERENCES

[1] Wang, Mengru, Yu Cai, Li Gao, Ruichen Feng, Qingju Jiao, Xiaolin Ma, and Yu Jia. "Study on the evolution of Chinese characters based on few-shot learning: From oracle bone inscriptions to regular script." Plos one 17, no. 8 (2022): e0272974.

[2] Liu, Kexin, Xiaohong Wu, Zhiyu Guo, Sixun Yuan, Xingfang Ding, Dongpo Fu, and Yan Pan. "Radiocarbon dating of oracle bones of late Shang period in ancient China." Radiocarbon 63, no. 1 (2021): 155-175.

[3] Fu, Xuanming, Zhengfeng Yang, Zhenbing Zeng, Yidan Zhang, and Qianting Zhou. "Improvement of oracle bone inscription recognition accuracy: A deep learning perspective." ISPRS International Journal of Geo-Information 11, no. 1 (2022): 45.

[4] Chang, Xiang, Fei Chao, Changjing Shang, and Qiang Shen. "Sundial-gan: A cascade generative adversarial networks framework for deciphering oracle bone inscriptions." In Proceedings of the 30th ACM International Conference on Multimedia, pp. 1195-1203. 2022.

[5] Li, Bang, Qianwen Dai, Feng Gao, Weiye Zhu, Qiang Li, and Yongge Liu. "HWOBC-a handwriting oracle bone character recognition database." In Journal of Physics: Conference Series, vol. 1651, no. 1, p. 012050. IOP Publishing, 2020.

[6] Hung, Shyh-Shiun, Hen-Hsen Huang, and Hsin-Hsi Chen. "A complete shift-reduce Chinese discourse parser with robust dynamic oracle." In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, pp. 133-138. 2020.

[7] Gao, Junheng, and Xun Liang. "Distinguishing oracle variants based on the isomorphism and symmetry invariances of oracle-bone inscriptions." IEEE Access 8 (2020): 152258-152275.

[8] Xuzheng, Lu, Cai Hengjin, and Lin Li. "Recognition of oracle radical based on the capsule network." CAAI transactions on intelligent systems 15, no. 2 (2020): 243-254.

[9] Li, Jing, Qiu-Feng Wang, Kaizhu Huang, Xi Yang, Rui Zhang, and John Y. Goulermas. "Towards better long-tailed oracle character recognition with adversarial data augmentation." Pattern Recognition 140 (2023): 109534.

[10] Wang, Mei, Weihong Deng, and Cheng-Lin Liu. "Unsupervised structure-texture separation network for oracle character recognition." IEEE Transactions on Image Processing 31 (2022): 3137-3150.

[11] Yue, Xuebin, Hengyi Li, Yoshiyuki Fujikawa, and Lin Meng. "Dynamic dataset augmentation for deep learning-based oracle bone inscriptions recognition." ACM Journal on Computing and Cultural Heritage 15, no. 4 (2022): 1-20.

[12] Han, Wenhui, Xinlin Ren, Hangyu Lin, Yanwei Fu, and Xiangyang Xue. "Self-supervised learning of Orc-Bert augmentator for recognizing few-shot oracle characters." In Proceedings of the Asian Conference on Computer Vision. 2020.

[13] Guo, Ziyi, Zihan Zhou, Bingshuai Liu, Longquan Li, Qingju Jiao, Chenxi Huang, and Jianwei Zhang. "An Improved Neural Network Model Based on Inception-v3 for Oracle Bone Inscription Character Recognition." Scientific Programming 2022, no. 1 (2022): 7490363.

[14] Liu, Mengting, Guoying Liu, Yongge Liu, and Qingju Jiao. "Oracle bone inscriptions recognition based on deep convolutional neural network." Journal of image and graphics 8, no. 4 (2020): 114-119.

[15] Li, Jing, Qiu-Feng Wang, Rui Zhang, and Kaizhu Huang. "Mix-up augmentation for oracle character recognition with imbalanced data distribution." In Document Analysis and Recognition–ICDAR 2021: 16th International Conference, Lausanne, Switzerland, September 5–10, 2021, Proceedings, Part I 16, pp. 237-251. Springer International Publishing, 2021.

[16] Fujikawa, Yoshiyuki, Hengyi Li, Xuebin Yue, C. V. Aravinda, G. Amar Prabhu, and Lin Meng. "Recognition of oracle bone inscriptions by using two deep learning models." International Journal of Digital Humanities 5, no. 2 (2023): 65-79.

[17] Han, Simon J., Piers Kelly, James Winters, and Charles Kemp. "Simplification is not dominant in the evolution of Chinese characters." Open Mind 6 (2022): 264-279.