Proceeding of ICIHCS 2025 Symposium: Integration & Boundaries: Humanities/Arts, Technology and Communication
DOI: 10.54254/2753-7048/2025.KM25974

Research on Oracle Bone Inscription Document Processing under Digital Technology

Zhaoyi Wang

College of Liberal Arts, Yangzhou University, Yangzhou, China
222308312@stu.yzu.edu.cn

Abstract. The study of oracle bone inscriptions has always been a significant subject for promoting the research of Chinese character culture in China. China has conducted research on the informatization processing of oracle bone inscriptions for over three decades and established several representative databases. This paper specifically lists five major existing domestic oracle bone inscription databases and their data collection categories. Through an in-depth analysis of the current status of digital technology's intervention in oracle bone inscription research, it is concluded that the construction of oracle bone inscription databases using artificial intelligence exhibits characteristics of automation, intelligence, and high-speed processing. This is of profound positive significance for integrating existing oracle bone inscription research achievements, promoting the disciplinary development of oracle bone inscriptions, and facilitating the dissemination of Chinese character culture. Therefore, strengthening the in-depth exploration of oracle bone inscription digitization and applying emerging technologies to the construction of oracle bone inscription databases have become the future development trends of oracle bone inscription research.

Keywords: Oracle bone inscription database, Types and quantities of oracle bone inscriptions, Digital processing

1. Introduction

As the earliest systematic writing system discovered in China, oracle bone inscriptions encapsulate abundant historical and cultural information of the Shang Dynasty, playing an indispensable role in the exploration of the origin of Chinese civilization. Nevertheless, due to the scarcity and dispersion of oracle bone materials, as well as the intricacy of oracle bone inscription decipherment, traditional oracle bone inscription research confronts numerous challenges. With the rapid advancement of digital technology, researchers can leverage digital network technology to digitize research manuscripts, documents, and other materials related to oracle bone inscriptions. This enables convenient access, comparison, and collation of original oracle bone inscription research works, regardless of time and space constraints. Moreover, document digitization can mitigate the risk of loss or damage that may occur during the manual examination of oracle bone inscription research manuscripts and other original materials, thereby further expanding and enhancing the application scope and utilization rate of original documents [1]. This study aims to discuss the application status of digital technology in oracle bone inscription document processing by reviewing relevant articles

DOI: 10.54254/2753-7048/2025.KM25974

on platforms such as CNKI, Wanfang Database, and Baidu Scholar. The research includes statistical analysis of the types and quantities of oracle bone inscriptions in existing databases, as well as an assessment of the extent of digital technology's intervention in oracle bone inscription research. Additionally, suggestions for further promoting the digitization process of oracle bone inscriptions are put forward.

2. Existing domestic oracle bone inscription databases and their collection types

2.1. Database classification

Currently, numerous oracle bone inscription databases have been restored and constructed through digitization technology both at home and abroad. The author has conducted a statistical survey of multiple oracle bone inscription databases and found that as of June 23, 2025, the Yin Qi Wen Yuan database has the largest collection of oracle bone inscriptions and related documents. In addition, existing domestic oracle bone inscription databases also include eight major databases, namely the Hong Kong Handa Ancient Books Database Retrieval System-Oracle Bone Inscription Database, the National Library of China Oracle Bone Inscription Database, the National Cheng Kung University (Taiwan) Oracle Bone Inscription Full -Text Image Database, the Knowledge and Network Application of Rubbing Collection-Oracle Bone Inscription Rubbing Database, the Hantang Collection Database, the Huayuanzhuang East Oracle Bone Inscription Retrieval System, the Oracle Bone Inscription Material Database, and the Oracle Bone Inscription Image Database. The specific collection details are presented in Table 1.

Table 1. The inclusion status of oracle bone inscriptions in five domestic databases (data as of June 23, 2025)

Database Name	Type of Collected Data	Product Type	Number of Collections	Collection Content	Update Situation
Yin Qi Wen Yuan [2]	Document (PDF format), oracle bone inscription rubbings	Document library + description library + character shape library	35,585 oracle bone inscription treatises; 154 descriptions; 239,902 oracle bone inscription images	Oracle bone inscription character shape library, oracle bone inscription description library, oracle bone inscription document library	From 2016 to the present

DOI: 10.54254/2753-7048/2025.KM25974

Hong Kong Handa Ancient Books Database Retrieval System-Oracle Bone Inscription Database [3]	Oracle bone inscription description books	Database (image + annotation)	67,683 pieces of divination inscriptions	Oracle bone inscription database, bronze inscription database, bamboo and silk manuscript database, Pre-Qin and Han Dynasties database, Wei, Jin, Southern and Northern Dynasties database, Chinese traditional classified book database, Chinese ancient vocabulary database; nine oracle bone inscription description books, including Collection of Oracle Bone Inscriptions, Supplement to the Collection of Oracle Bone Inscriptions, Oracle Bone Inscriptions in the Southern Area of Xiaotun, Oracle Bone Inscriptions Collected in Su, De, Mei, and Ri, Collection of Oracle Bone Inscriptions Collected by White, Oracle Bone Inscriptions in the East of Huayuanzhuang, Yin Ruins, Oracle Bone Inscriptions Collected by the Institute of Oriental Culture, University of Tokyo, Collection of Oracle Bone Inscriptions Collected in the UK, Oracle Bone Inscriptions Collected by the Reference Library Affiliated to Tianli University	From 1996 to the present
National Library of China [4]	Oracle bone inscription objects, oracle bone inscription rubbings	Database (image + annotation)	35,651 pieces, including 6,575 oracle bone inscription objects; 13,776 oracle bone inscription rubbings	Large-scale collection with a wide range	From 1909 to the present
National Cheng Kung University (Taiwan) Oracle Bone Inscription Full-Text Image Database [5]	Oracle bone inscription bibliography and image data	Database (image + annotation)	41,956 pieces of oracle bone inscription bibliography and image data	Collection of Oracle Bone Inscriptions (thirteen volumes), Compilation of Inscriptions on Oracle Bone Inscriptions in Yin Ruins (upper, middle, and lower volumes), and General Compilation of Transcriptions of Inscriptions on Oracle Bone Inscriptions in Yin Ruins (upper and lower volumes)	From 1995 to the present
Knowledge and Network Application of Rubbing Collection-Oracle Bone Inscription Rubbing Database [6]	Oracle bone inscription rubbings	Database (image + annotation)	More than 40,000 pieces	Shang Dynasty oracle bone inscriptions, pottery inscriptions, brick inscriptions, Pre-Qin bronze inscriptions, seals, mud seals, Han Dynasty stone carvings, epitaphs of various dynasties, and Buddhist stone carving rubbings; a large number of oracle bone inscriptions and bamboo slips excavated around the 1930s	From 2004 to the present

DOI: 10.54254/2753-7048/2025.KM25974

Furthermore, some oracle bone inscription databases are continuously updating their collections. The Hantang Collection Database has amassed 14 types of documents related to oracle bone inscriptions. The Huayuanzhuang East Oracle Bone Inscription Retrieval System, developed by the Center for Chinese Character Research and Application at East China Normal University. It contains Oracle Bone Inscriptions in the East of Huayuanzhuang, Yin Ruins (2003 edition) compiled by the Institute of Archaeology, Chinese Academy of Social Sciences. However, the exact number of collected items is not disclosed. The Oracle Bone Inscription Material Database, jointly developed by Professor Huang Tianshu of Capital Normal University and Professor Wang Yunzhi of Henan University. It has gathered 32 types of domestic and foreign oracle bone inscription descriptions and joining materials. The system has inputted nearly 80,000 original oracle bone inscription pieces, over 157,300 inscriptions, with a total of approximately 3.4 million words in the original text and annotations; the Oracle Bone Inscription Image Database, developed by the research group of Professor Han Jiangsu at Anyang Normal University, consists of three sub-databases: the oracle bone inscription picture library, the annotation library, and the original character library, and has collected 9 types of oracle bone inscription descriptions, such as Collection of Oracle Bone Inscriptions, British Collection, and Supplement, totaling 72,264 oracle bone inscription pieces [7].

Despite these databases' substantial collections of oracle bone inscription materials, a significant number of oracle bone inscriptions remain un-digitized. Liu Yongge, Director of the Key Laboratory of Oracle Bone Inscription Information Processing at Anyang Normal University, Ministry of Education, has indicated that the total number of oracle bone inscriptions unearthed from the Yin Ruins is approximately 160,000 pieces, which are currently housed in 181 museums, libraries, research institutions, and universities both at home and abroad. Among them, about 120,000 pieces are located within China. Currently, tens of thousands of oracle bone inscriptions are in the hands of private collectors. Due to illegal excavation and trafficking, over 30,000 pieces of oracle bone inscriptions have been lost overseas, scattered in institutions such as the Carnegie Museum, the Royal Museum of Scotland, the British Museum, and Princeton University, and many of them have not been made public in the academic community, resulting in incalculable losses to China's cultural heritage [8].

2.2. Database classification

The oracle bone inscriptions collected in the above-mentioned databases mainly include oracle bone inscriptions (original inscriptions), oracle bone inscription facsimiles, oracle bone inscription rubbings, and other categories. The specific comparison is shown in Table 2.

Table 2. A comparison of the three types of oracle bone inscriptions

Type	Meaning	Advantages
Oracle bone inscriptions	Text records directly engraved on oracle bones	As primary source documents, they possess the highest research value, recording information on divination, sacrifice, warfare, agriculture, astronomy, and other aspects
Oracle bone inscription facsimiles	Replicas directly copied from the original oracle bone inscriptions	Offer an intuitive understanding of the shape and structure of the originals, facilitating scholars' comprehension of the writing habits and artistic styles of oracle bone inscriptions
Oracle bone inscription rubbings	Text image materials obtained by rubbing the surface of oracle bone inscriptions	Provide high-definition oracle bone inscription images, enabling scholars to conduct detailed research and analysis

DOI: 10.54254/2753-7048/2025.KM25974

3. Current situation of oracle bone inscription digitization processing

Currently, the primary applications of digital technology in oracle bone inscription research involve digitization processing, including: (1) digitization of oracle bone inscription materials, reference books, and research documents; (2) denoising processing of oracle bone inscription rubbings or images; (3) conversion of oracle bone inscription dot-matrix fonts into scalable vector fonts recognizable and processable by computers; (4) realization of coding input, handwriting input, and recognition of oracle bone inscriptions, enabling their display on computer screens; (5) decipherment and joining of oracle bone inscriptions; (6) restoration of oracle bone inscription rubbing fonts through computer technology [7]. Through digitization, the original information of oracle bone inscriptions can be preserved for an extended period, facilitating access and study by scholars. For instance, the Yin Qi Wen Yuan platform employs big data and cloud computing technologies to digitize oracle bone inscriptions, achieving the integration and digital storage of oracle bone inscription descriptions, images, character shapes, and academic treatises. This overcomes the challenges of sharing and promotion of oracle bone inscription descriptions and literature resources, which were previously hindered by the difficulty of oracle bone inscription input and the tedium of information annotation. In the process of oracle bone inscription research, obtaining a complete oracle bone inscription, fully extracting the text information on it, and intuitively presenting the appearance of oracle bone inscriptions to the public to promote their dissemination are crucial research stages. Therefore, this paper will focus on the analysis of oracle bone inscription joining, oracle bone inscription recognition and detection, and 3D modeling and collection of oracle bone inscriptions.

3.1. Oracle bone inscription joining

Oracle bone inscription joining refers to the splicing operation of identifying the materials of discovered oracle bone inscription fragments and then joining fragments of the same material based on the shape of the oracle bones and the inscriptions carved on them, to accurately restoring the positions of the fragments. Currently, oracle bone inscription joining primarily relies on manual splicing. Researchers need to utilize information such as inscriptions, semantics, context, annotations, materials, and contour consistency to search for potential associated information among a vast number of oracle bone inscription fragments, resulting in a complex and process. The intervention of digital technology provides new approaches for oracle bone inscription joining. In the field of computer science, the oracle bone inscription joining problem can be divided into three phases: acquisition of geometric, texture, and other features; pairwise matching; and global search [9]. Computer technology can rapidly match the features of oracle bone inscription fragments by establishing a digital data model and combining computer vision analysis with intelligent algorithms. This improves the accuracy and efficiency of joining, making the spliced oracle bone inscriptions valuable new materials for historical and archaeological research.

3.2. Recognition and detection of oracle bone inscriptions

The recognition and detection of oracle bone inscriptions are key links in digital technology's intervention in oracle bone inscription research. Traditional oracle bone inscription recognition methods generally consist of feature extraction and feature classification. The objective of feature extraction is to obtain the unique features of oracle bone inscription images, while feature classification determines which oracle bone inscription character a feature belongs to base on the

DOI: 10.54254/2753-7048/2025.KM25974

extracted features. Common feature extraction methods include Feature Transform (SIFT), Histogram of Oriented Gradients (HOG), and Local Binary Pattern (LBP). Common feature classifiers often employ Support Vector Machines (SVMs). However, this method is highly dependent on algorithm designers, and the selection of different features and recognizers can lead to significant variations in recognition results [10]. In recent years, recognition technologies based on deep neural networks have made remarkable progress. For example, the hierarchical-representation-based oracle bone inscription character recognition method proposed by Wang Changhu from Microsoft Research Asia, and the methods proposed by research teams from several domestic universities to improve recognition rates for different convolutional neural networks, such as using facsimile character shapes to assist in rubbing character shape recognition, optimizing models to enhance recognition accuracy, and proposing data augmentation strategies, have all improved the accuracy and efficiency of oracle bone inscription recognition to a certain extent. In 2022, a professional research team from East China Normal University developed a standardized dataset called OBI-100, which based on three typical convolutional network. After conducting numerous experiments and applying diverse optimization techniques, the features of oracle bone inscription enabled it to achieve an accuracy rate of 99.5% in the 100-class OBI recognition task in a practical and efficient manner [11].

3.3. 3D modeling and collection of oracle bone inscriptions

The 3D modeling and collection of oracle bone inscriptions aim to preserve oracle bone inscription information more comprehensively. The Key Laboratory of Oracle Bone Inscription Information Processing at Anyang Normal University, Ministry of Education, launched the "Global Digital Revitalization of Oracle Bone Inscriptions" initiative. Relying on artificial intelligence and other technologies, it has achieved technical breakthroughs in "micro-trace analysis" and "3D modeling of oracle bone inscriptions" and proposed the "oracle bone inscription full-information data model," which includes 3D models, micro-trace enhancement maps, micro-trace grayscale maps, digital facsimiles, digital rubbings, scientific image layers, etc. This has enabled restoration and protection of oracle bone inscription objects in the digital space, providing richer and more accurate information for oracle bone inscription research [12].

4. Prospect of oracle bone inscription digitization construction

Digital technology has achieved remarkable results in oracle bone inscription document processing and research. Technologies such as digital processing, text recognition and detection, oracle bone inscription joining, 3D modeling, and collection have been applied to oracle bone inscription document research. Existing oracle bone inscription databases cover a wide range of oracle bone inscription types and quantities, providing a solid data foundation for research. However, the intervention of digital technology in oracle bone inscription research still faces challenges such as low data quality, technical difficulties, and a shortage of professional talents. With the deepening of digital technology in oracle bone inscription document research, in terms of data, it is expected to achieve more comprehensive and accurate digital collection and integration of global oracle bone inscription resources and establish a more complete and standardized oracle bone inscription database. In terms of technology application, artificial intelligence algorithms will continue to be optimized to improve the accuracy and efficiency of oracle bone inscription recognition, joining, and decipherment, and realize in-depth human-computer collaborative interaction. In terms of communication, leveraging more diversified digital media platforms and innovative display methods

DOI: 10.54254/2753-7048/2025.KM25974

will enable oracle bone inscriptions to better penetrate public life, promoting the inheritance and development of oracle bone inscription culture. Meanwhile, interdisciplinary research will also become a trend. The of computer science, history, archaeology, linguistics, and other disciplines will bring new perspectives and methods to oracle bone inscription research, facilitating the achievement of more innovative research results.

5. Conclusion

Examining the current situation of domestic oracle bone inscription databases, significant differences exist among different databases in terms of collection scope, types, and quantities, making rapid access unfeasible. Digital technology has brought great convenience to oracle bone inscription document research, but there is still ample room for development to truly achieve database intelligence. This study proposes integrating the resources of major databases and relying on relevant artificial intelligence technologies to construct an oracle bone inscription database that is academically reliable. This will play a more favorable role in the scientific research of oracle bone inscription scholars and the development and dissemination of oracle bone inscription culture.

References

[1] Zheng, Y.X.: Ancient Books Digitization and the Future of Classical Literature Research. Literary Heritage (5), (2005)
[2] Yin Qi Wen Yuan, http: //jgw.aynu.edu.cn/ajaxpage/home2.0/index.html, last accessed 2025/06/23
[3] Handa Corpus, http: //www.chant.org/Default.aspx, last accessed 2025/06/23
[4] National Library of China, https: //www.nlc.cn/web/index.shtml, last accessed 2025/06/30
[5] Library of National Cheng Kung University, Taiwan, https: //www.baidu.com/link?url=jIxpTTJjiq7K7v0QEzf-L-0xbAg-7jmhhIuR_LSYeYX2aKjB14_5UXbDCHBM-uIT4-mC9TH-41KR0QQ-rp7p1a& wd=& eqid=aa96582500, last accessed 2025/06/23
[6] Topography Collection Knowledge and Network Application: Oracle Topography Database, http: //rub.ihp.sinica.edu.tw, last accessed 2020/06/23
[7] Zhang, F., Zhao, N.: Comparative study of domestic oracle database. Journal of Luohe Vocational and Technical College 21(2), 12–19 (2022)
[8] Sun, Y., Zhao, Y.S., Teng, J.P.: Digital technology makes the millennium oracle "reborn". Science and Technology Daily (005) (2024-09-26)
[9] Gao, W.Z.: Research on the classification and aggregation method of oracle topography based on deep learning. Southwest University, (2022)
[10] Gao, F., Song, Z.H.: Digital Resources Revitalization Oracle Research. China Social Science News (2023-07-22)
[11] Fu, X., Yang, Z., Zeng, Z., et al.: Improvement of oracle bone inscription recognition accuracy: A deep learning perspective. ISPRS International Journal of Geo-Information 11(1), 45 (2022)
[12] Zhang, A.S., Sun, X.F.: Four "helps" to promote the "life" of oracle writing – Anyang Normal University has comprehensively promoted the protection and development of oracle writing to achieve fruitful results. Henan Education (Higher Education) (11), 6 (2023)