

海岱地区甲骨文全信息数字档案资源建设研究

翟雪¹, 梁建国², 韩胜伟³, 仇利萍³, 熊晶^{2,3*}

¹曲阜师范大学工学院 山东日照

²曲阜师范大学计算机学院 山东日照

³甲骨文信息处理教育部重点实验室 河南安阳

【摘要】海岱地区甲骨文的发现证明了商文化在殷商都城之外的延伸和分布,证实了商王朝对东方的扩张和经营,具有重要的研究价值。然而,海岱地区甲骨文的知识体系较为庞大且极为复杂,目前缺乏一个系统化的档案记录方式。为解决这一问题,通过整合博物馆馆藏资源、学术文献及甲骨文数据库等多源数据,构建了海岱地区甲骨文知识图谱,并在此基础上实现多模态信息对齐与全信息数字档案建设。甲骨文全信息数字档案不仅可以全面展示海岱地区甲骨文的知识联系,并对其知识体系进行细粒度的组织,还为证明海岱地区的商文化是与本地文化的一种融合提供了有力依据。该方法可为其他领域的古籍数字档案资源建设提供有益的借鉴。

【关键词】甲骨文;数字档案;知识图谱;全信息

【基金项目】国家自然科学基金(U1504612):面向甲骨学知识图谱的实体发现及语义关系挖掘研究;山东省自然科学基金面上项目(ZR2024MF146):基于虚拟知识图谱的多源异构海洋数据安全共享研究

【收稿日期】2026年4月19日

【出刊日期】2026年5月7日

【DOI】10.12208/j.ssr.20260163

Construction of comprehensive digital archive for Oracle Bone Inscriptions in the Haidai region

Xue Zhai¹, Jianguo Liang², Shengwei Han³, Liping Qiu³, Jing Xiong^{2,3*}

¹College of Engineering, Qufu Normal University, Rizhao, Shandong

²School of Computer Science, Qufu Normal University, Rizhao, Shandong

³Key Laboratory of Oracle Bone Inscriptions Information Processing, Ministry of Education of China, Anyang, Henan

【Abstract】The discovery of Oracle Bone Inscriptions (OBI) in the Haida region demonstrates the extension and distribution of Shang culture beyond the Shang dynasty's capital, confirming the Shang dynasty's expansion and governance in the eastern territories. This holds significant historical value and research importance. However, the knowledge system of Haida OBI is vast and highly complex, and currently lacks a systematic approach to archive organization. To address this issue, a knowledge graph of OBI from the Haidai region is constructed by integrating multi-source data, including museum collection records, academic literature, and specialized OBI databases. Building upon this foundation, multi-modal information alignment is further implemented, and a comprehensive digital archive is established. This approach not only comprehensively showcases the knowledge connections within Haida OBI but also provides strong evidence supporting the fusion of Shang culture with local culture in the Haida region. It holds reference value and significance for emulation in promoting digital research on ancient texts in other fields.

【Keywords】Oracle Bone Inscriptions; Digital archives; Knowledge graph; Comprehensive information

1 引言

甲骨文是中华民族珍贵的文化遗产,是汉字的源头和中华优秀传统文化的根脉。作为我国目前发现的最古老成熟文字,甲骨文既是研究汉字发展和中国早期历史的宝贵材料,也是中华五千年文明历程中不可

或缺的一环。进入 21 世纪以来,甲骨文研究已成为国家与中华优秀文化体系构建相关的战略需求,与中华文明影响力和国家文化软实力息息相关。

2017 年 10 月甲骨文成功入选“世界记忆名录”,标志着世界对甲骨文重要文献价值的高度认可,其意

第一作者简介:翟雪(1977-)女,山东章丘人,硕士,经济师。

*通讯作者:熊晶

义十分重大。越来越多的学者开始从事甲骨文研究工作, 积累了大量甲骨文基础研究数据。2019年10月, 甲骨文大数据平台“殷契文渊(<https://jgw.aynu.edu.cn/>)”面向全球公开发布, 为甲骨文研究提供大数据支持, 标志着甲骨文研究进入智能化时代。

殷墟甲骨是中国商朝后期都城遗址——河南安阳殷墟出土的王室占卜记事用的龟甲兽骨。而入选“全国十大考古新发现”和“山东百年百项重要考古发现”的济南大辛庄商代遗址则是海岱地区商文化的典型代表之一。有学者称“大辛庄是山东境内最早的、最大的、最丰富的、最重要的、特征最鲜明的、出土甲骨文最早的高代遗址。”自1899年殷墟甲骨文发现以来, 大辛庄甲骨文是首次在商代都城以外地区发现的甲骨文^[1], 被称为甲骨学史上具有界标意义的重大发现, 为商代历史研究提供了宝贵的第一手资料。这些甲骨文与殷墟甲骨文同属一个系统, 同时还具有地方特色, 显示了商文化在海岱地区的延伸和分布, 以及商王朝对东方

的扩张和经营。因此, 海岱地区甲骨文具有重要的史料价值和研究意义。

但是, 目前针对海岱地区甲骨文的数字档案建设研究相对较少。海岱地区位于我国东部, 该地区出土的甲骨文数量较多, 且具有独特的地域特色, 有助于推动夏商周三代文化的研究。然而, 一方面, 甲骨文涉及到多层级的知识组织, 如甲骨文载体包括照片、拓片、摹本和释文等多种模态, 且每个甲骨文字及其位置与释文一一对应, 如图1所示; 另一方面, 海岱地区甲骨文的数字档案研究起步较晚, 且涉及地域广阔、文化背景复杂, 导致其系统化研究相对滞后。因此, 亟需一个有效的数字档案建设方法, 为传承和弘扬中华优秀传统文化, 促进文化产业发展发挥重要作用^[2]。

基于此背景, 我们采用知识图谱对海岱地区甲骨文知识进行高效管理, 通过构建甲骨文全信息数字档案资源, 以期实现“每一片甲骨身世清晰, 每一个已识字释义明白”。

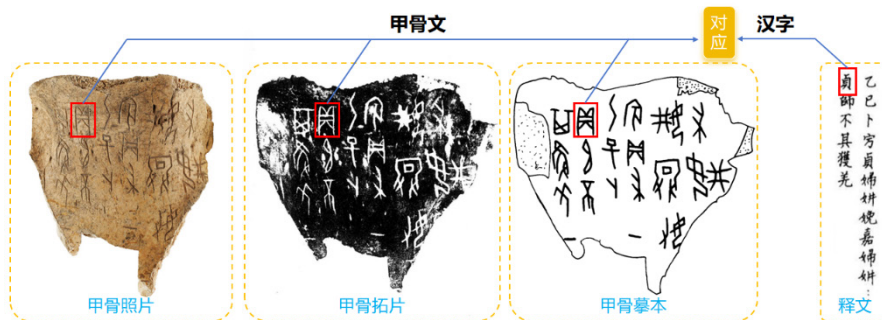


图1 甲骨文数字资源

2 国内外研究现状

知识图谱以其强大的语义表达能力、存储能力和推理能力, 为知识组织和智能应用提供了有效的解决方案^[3]。近年来, 知识图谱在档案信息组织方面的研究与应用成为热点^[4,5]。刘丹丹^[6]认为知识图谱是实现古籍特藏知识重组的关键技术。张文亮等^[7]基于知识图谱构建了满文古籍知识库, 实现了对满文古籍的知识检索、数据分析、智能推荐和AI问答等知识服务。郝春秋^[8]指出构建红旗渠档案知识图谱, 能对红旗渠档案形成过程中存在的人、地、时、事、物等隐性知识进行挖掘, 有助于打破文档网络的界限。

知识图谱作为一种新兴的数据表示和知识组织技术, 在甲骨文研究领域也得到了广泛应用。通过知识图谱技术, 甲骨文专家学者和爱好者可以发现甲骨文知识之间的关联性, 揭示甲骨文考释成果以及了解商代

社会历史和文化等^[9]。熊晶等^[10]构建了殷墟甲骨文知识图谱用以满足用户对甲骨文基础知识的自动问答需求。Xiong等^[11]提出了一个甲骨文研究应用金字塔, 并指出知识图谱是该金字塔智能应用层的基础。Song等^[12]认为甲骨文知识图谱有助于提高甲骨文释读效率, 探索商代社会关系及促进传统文化数字化保护。李沿增^[13]提出了一个结合目标检测和知识图谱的网络结构, 用来提高甲骨文自动识别的数量。

虽然知识图谱在甲骨文研究领域取得了以上成果, 但是仍面临以下挑战: 1) 甲骨文研究数据主要是照片、拓片、释文, 但质量参差不齐, 需要进一步清洗和标准化, 它们之间的对应关系也需要进一步理清; 2) 如何准确、全面地表示甲骨片信息及甲骨卜辞记录的知识, 是一个亟待解决的问题; 3) 知识图谱技术需要与其他学科(如人工智能、古文字学、历史学等)进行交叉融合,

以实现全方位多维度的甲骨文知识关联; 4) 已有的甲骨文知识图谱主要以殷墟甲骨文为对象, 缺乏针对海岱地区甲骨文的知识图谱。而且, 海岱地区甲骨文与殷墟甲骨文在考古信息、历史渊源、数据整理、知识关联、活化利用等方面也亟需一套可行的档案信息组织框架。

因此, 构建海岱地区甲骨文知识图谱具有重要意义且已成为甲骨文研究者和爱好者的迫切需求。

3 海岱地区甲骨文知识图谱

构建知识图谱的关键是实体及其属性的抽取, 以及建立实体之间的关系^[14,15]。海岱地区甲骨文知识图谱构建步骤如下:

(1) 数据收集和预处理。首先, 从海岱地区博物馆的甲骨文馆藏信息、甲骨文专业书籍、学术论文、“殷契文渊”甲骨文大数据平台等多种来源收集甲骨文图像、文字记录、出土背景及研究解读等相关数据。然后,

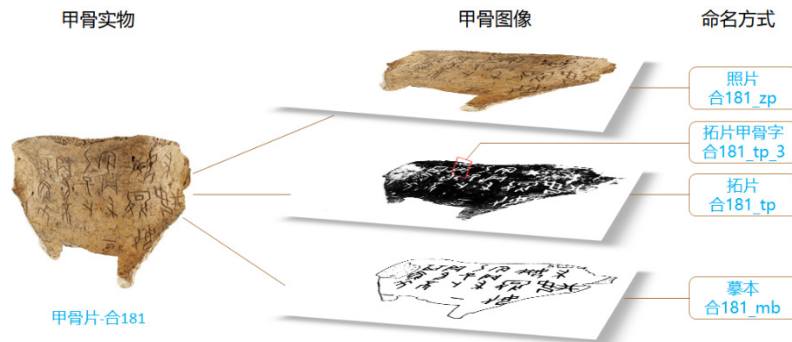


图2 甲骨知识组织方式

表1 本体中的实体类别

实体类别	说明	示例
著录	记录甲骨片信息的著录	甲骨文合集
甲骨片	采用著录中的命名	合 181
甲骨照片	某片甲骨的照片形式	合 181_zp.jpg
甲骨拓片	某片甲骨的拓片形式	合 181_tp.jpg
甲骨摹本	某片甲骨的摹本形式	合 181_mb.jpg
甲骨释文	某片有字甲骨对应的释文	合 181_sw.jpg
甲骨文字符	从某片甲骨中裁剪下来的单个甲骨文	合 181_tp_3.jpg
博物馆	收藏某片甲骨的博物馆	山东博物馆

表2 本体中的关系类别

关系类别	说明	三元组示例
照片	甲骨片与其照片形式的对应关系	<合 181, 照片, 合 181_zp.jpg>
拓片	甲骨片与其拓片形式的对应关系	<合 181, 拓片, 合 181_tp.jpg>
摹本	甲骨片与其摹本形式的对应关系	<合 181, 摹本, 合 181_mb.jpg>
释文	甲骨片与其释文的对应关系	<合 181, 释文, 合 181_sw.jpg>
包含	甲骨片与其上甲骨文字符的对应关系	<合 181_tp.jpg, 包含, 合 181_tp_3.jpg>
馆藏地	甲骨片实物的收藏位置	<合 181, 馆藏地, 山东博物馆>
收录	著录与甲骨片的记录关系	<甲骨文合集, 收录, 合 181>
重片	甲骨片的重复收录关系	<重博 3, 重片, 合 2837>

对数据进行预处理, 包括去除重复、清洗噪声、统一格式、分词、词性标注等工作, 为后续实体及属性抽取奠定基础。

知识图谱需充分考虑甲骨片实物及其对应的照片、拓片和摹本的对齐信息, 采用统一的知识组织方式以及命名规则。以《甲骨文合集》中收录的甲骨片“合 181”为例, 其知识组织如图 2 所示。

(2) 本体建模。本体提供甲骨文领域内的共享概念模型, 它定义了领域内的概念、实体及其关系。该模型对甲骨文领域的核心要素进行了形式化定义, 并综合了其历史、文化和考古学背景, 形成了一个全面而系统的知识体系。通过本体建模, 实现了对甲骨文知识的有效组织与管理, 为知识图谱的构建提供了概念框架, 本体模型中的实体类别如表 1 所示。

本体模型中的关系如表 2 所示。

可以在右侧显示(其中,未识别的甲骨文因无对应的现代汉字,仍然以原始字形显示)。通过甲骨文自动检测方法^[6],甲骨文片上的每个甲骨文均会形成一个矩形框,以该矩形框的范围设置锚点,当鼠标悬停在锚点上时,

将动态显示该甲骨文对应的现代汉字(如图5左侧的“贞”字悬浮显示在甲骨文上)。同时,图5右侧的信息框将同步显示鼠标悬停位置所对应的释文汉字,并提供其详细解释。



图4 甲骨文全信息展示



图5 甲骨文动态悬停显示

5 总结与展望

本文针对海岱地区的甲骨文馆藏情况, 构建了一个甲骨文知识图谱, 并基于该图谱从著录级、馆藏级、片级、字符级等多个级别, 以及照片、拓片、摹本、释文、字符等多个粒度, 实现了甲骨文的全信息数字档案资源建设。该方法可以全方位展示甲骨文的图片和文本信息, 极大地满足了甲骨文研究者及爱好者的需求。

但是, 由于海岱地区甲骨文数字档案资源建设是一个庞大的工程, 目前还有很多工作尚在进行中, 如山东博物馆收藏甲骨 1.05 万余片, 但现有的知识图谱只包含其中的一小部分。下一步研究计划中, 我们拟采用大语言模型技术加快甲骨文知识的分级整理工作, 并随着甲骨文的出土及信息公开(如济南大辛庄遗址博物馆)情况不断扩充海岱地区甲骨文知识图谱的规模。

参考文献

- [1] 房道国. 济南大辛庄商代甲骨文发现纪实[J]. 大众考古, 2017(09):65-69.
- [2] 贝文玥. 文化数字化战略背景下博物馆文物数字档案应用路径[J]. 兰台世界, 2025(06):141-144.
- [3] 王萌, 王昊奋, 李博涵, 等. 新一代知识图谱关键技术综述[J]. 计算机研究与发展, 2022,59(09):1947-1965.
- [4] 关莉红. 数字人文视角下历史档案遗产的知识组织方法[J]. 山西档案, 2025(05):83-85.
- [5] 王玲, 吴俊贤. 基于知识图谱的高校档案管理系统[J]. 兰台世界, 2025(04):114-119.
- [6] 刘丹丹. 数字人文视域下古籍特藏知识重组模式研究[J]. 山西档案, 2025(05):112-114.
- [7] 张文亮, 丁淼, 魏来. 满文古籍知识库研究与构建[J]. 图书情报工作, 2025,69(05):126-139.
- [8] 郝春秋. 本体构建在红旗渠档案资源开发中的应用研究[J]. 兰台世界, 2025(03):55-58.
- [9] 熊晶, 钟珞, 王爱民. 甲骨文知识图谱构建中的实体关系发现研究[J]. 计算机工程与科学, 2015,37(11):2188-2194.
- [10] 熊晶, 刘永革, 韩胜伟, 等. 基于知识图谱的甲骨文智能问答系统[J]. 信息技术与信息化, 2023(05):212-216.
- [11] Xiong J, Guo T, Liu G, et al. OBSKP: oracle bone studies knowledge pyramid model with applications[J]. KO KNOWLEDGE ORGANIZATION, 2023,49(7):483-495.
- [12] Song D, Su C. Research on the Construction of Oracle Bone Inscriptions Knowledge Graph and its Application in Historical Language Analysis[J]. International Journal of Science and Engineering Applications, 2025,14(1):61-64.
- [13] 李沿增. 基于目标检测和知识图谱的古文字识别研究[D]. 长春: 吉林大学, 2023.
- [14] 熊晶, 焦清局, 刘运通. 基于多源异构数据的甲骨学知识图谱构建方法研究[J]. 浙江大学学报(理学版), 2020, 47(02):131-141.
- [15] Xiong J, Liu G, Liu Y, et al. Oracle Bone Inscriptions information processing based on multi-modal knowledge graph[J]. Computers & Electrical Engineering, 2021,92(6):107173.
- [16] Xing J, Liu G, Xiong J. Oracle bone inscription detection: a survey of Oracle bone inscription detection based on deep learning algorithm[C]// Proceedings of the International Conference on Artificial Intelligence, Information Processing and Cloud Computing, Sanya, China, ACM, 2019:1-8.

版权声明: ©2026 作者与开放获取期刊研究中心(OAJRC)所有。本文章按照知识共享署名许可条款发表。

<http://creativecommons.org/licenses/by/4.0/>



OPEN ACCESS