

数智时代古籍数字化再造的逻辑与进路

雷珏莹 侯西龙 王晓光

摘 要 随着古籍保护计划和古籍数字化出版工作的持续开展,我国已经积累了海量的古籍数字资源。数智化时代背景下,如何应用新技术和新理念来实现古籍蕴涵的优秀文化基因的提炼、展示与活化利用,逐渐成为新时代古籍数字化工作的主要目标和核心任务。文章在古籍数字化基础上提出了“古籍数字化再造”的概念,辨析古籍数字化与古籍数字化再造工作的区别,提炼古籍数字化再造的理论基础,从载体记录层、内容组织层和表现应用层构建了古籍数字化再造的堆栈模型,并从理论、技术、用户、机制、模式五个维度提出古籍数字化再造的实现进路,为古籍资源的深度开发与活化利用提供参考。古籍数字化再造是未来古籍活化工作的重要组成部分,可为新时代古籍工作提供新的指导理念。

关键词 古籍数字化再造; 智慧数据; 数字人文; 知识组织

分类号 G203; G256

作者简介 雷珏莹,武汉大学信息管理学院博士研究生; 侯西龙,武汉大学信息管理学院助理研究员、博士后; 王晓光(通讯作者),武汉大学信息管理学院教授,武汉大学文化遗产智能计算实验室主任,Email:wxguang@whu.edu.cn。

0 引言

古籍是中华文明传承与发展的重要载体,不仅具有重要的文物价值和学术价值,还蕴含着中华民族宝贵的历史记忆和文化基因。自 20 世纪 70 年代开始,我国古籍数字化工作起步并取得了长足的进步。^[1]2019 年,国家图书馆(国家古籍保护中心)组织的第四次古籍数字资源调查统计数据显示,全国古籍数字资源发布总数已超过 7.2 万部。上海图书馆公益性发布馆藏家谱 8000 余部、珍贵古籍 470 余部;中华书局利用已

基金项目:本文系国家社会科学基金重大项目“文化遗产智慧数据资源建设与服务研究”(21&ZD334)、湖北省自然科学基金创新研究群体项目“大数据资源语义组织与管理应用研究”(2019CFA025)研究成果之一。

出版资源开发了“中华经典古籍库”，先后完成九期共 5039 种古籍成果的数据加工，资源总量达到 20 亿字；爱如生数字技术公司的“中国基本古籍库”收书达 1 万种；阿里巴巴联合四川大学等多家机构开展的“汉典重光”项目首批 20 万页古籍已完成数字化。这些丰富的古籍数字资源为古籍整理研究与出版发布工作提供了重要支撑。

十八大以来，习近平总书记多次提出“让书写在古籍里的文字活起来”的要求。2022 年 4 月，中共中央办公厅、国务院办公厅联合印发的《关于推进新时代古籍工作的意见》对新时代古籍工作的开展进行了统筹部署。如何利用大数据、人工智能、区块链等新兴技术实现对浩如烟海的古籍资源的系统性整理与研究、创造性转化与创新应用，让古籍所承载的内容“活起来”，成为一个崭新的时代课题。当前，现有古籍数字资源以原版图像、电子文本等非结构化数据形式为主，虽然在一定程度上满足了学者的资源获取和利用需求，但从使用效果和服务质量来看，仍不尽如人意。数据质量问题较大，研究者对分析工具的认识不足，整理校勘、电子转录和数据分析任务仍然十分艰巨。^[2]随着国家文化产业数字化转型战略的实施，文化大数据体系建设也对古籍的活化利用提出了更多需求，如何借助数智技术赋能古籍活化利用也已经成为一个重要的时代课题。

本文在数字人文研究理念指引下，尝试性提出了“古籍数字化再造”的概念，意在现有的古籍数字资源基础上尝试对其进行“数据化”和“智慧化”的创造性再表示和再组织，使现有数字资源升级转变成古籍“智慧数据”，进而为古籍创新性整理、发布、阅读、分析和利用提供基础，使古籍发挥出更大的学术价值、文化价值和社会价值。在以大数据、人工智能等新兴技术为代表的数智化时代，古籍数字化再造将为文化遗产活化利用提供一条新颖的发展进路。

1 古籍数字化再造的基本逻辑

“再造”一词在“在线汉语字典”^①中解释为：重新给予生命；泛指再生、复活；重新创建。英文中的 rebuild、reengineer、reconstruct 等单词都可以与之相对应。“rebuild”意为重建、恢复到以前的状态，“reengineer”意为重新设计、重组运作（以提高效率），“reconstruct”则意为重新构造、重新创造或重新想象过去的事物。从这些词义可以看出，“再造”意味着更新、再生和复活。早在 2002 年，我国就针对古籍的二次利用问题，提出过“中华再造善本”文化工程。“再造善本”是纸质文献环境下古籍的原版“再造”，是将分藏于不同机构的珍贵古籍善本利用现代印刷技术影印复制，再按照古代书籍的装订排版方式印刷出版，为学界所利用、大众所共享。^[3]在“再造善本”工程中，古籍的知识载体形式不曾改变，原汁原味地再现了古籍的内容与形式。

数字化对古籍的整理与出版工作具有重大影响。作为古籍文献资源保护与利用的重要手段之一，古籍数字化是对实体古籍进行数字化转换，涉及古籍著录、分类标引、文本识别、检索利用、网络传播与共享等众多环节的一个系统工程。^[4]古籍数字化也是围绕已存的古籍实体展开，强调古籍的数字化保护、整理和出版。随着实践发展的不断深入，人们对古籍数字化的认识先后经历了存储介质的转换、文本组织与检索、数据加工与知识服务三个阶段。^[5]在此过程中，将古籍数字化文献仅仅当作方便查询的“书”来使用的观念开始转变^[6]，不断深化的则是挖掘古籍蕴含的知识体系与文化内涵，创新古籍数字化产品和服务方式。伴随数字人文的发展，分散独立的古籍数字化资源实现语义关联和知识重组，并在此基础上提供智慧化的知识服务

① <https://cd.diyifanwen.com/>

成为新的需求。

古籍数字化再造,则是脱离古籍纸质载体、对数字化古籍内容的再挖掘、再组织与再表达,是数智时代古籍数字化实践的进一步拓展。它以实现古籍的“活化”为目标,核心要义是利用本体、语义出版、数据建模、知识组织等数字技术和方法对古籍数字资源的开发与利用;通过挖掘、关联、重组,表达、展示与传播古籍所蕴含的历史知识和文化内涵,将古籍数字资源转化为智慧数据资源,实现古籍内容形式、呈现形态、应用场景的数字化再造与活化,为人文学术研究、文化传播等提供有力支撑和保障。

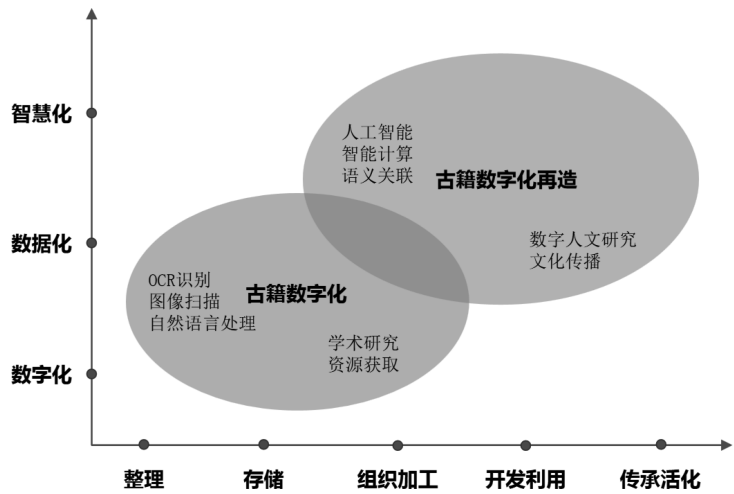


图1 “古籍数字化再造”与“古籍数字化”

进一步两相比较,古籍数字化的主要目的是古籍的保护与共享发布,它通过建立书目索引、影印拍照、扫描 OCR、全文数字化等方式对古籍原文进行加工和再现,实现古籍资源从物质载体到数字媒介的迁移。在以往的古籍数字化过程中,古籍的数字资源形态以数字图像为主,也包括转录后的电子数据,以及整理后的全文数据,但资源整体的颗粒度较粗,仍然是非结构化的、计算机难于处理的数字资源。同时,对古籍数字文献的管理也延续了纸本时代的观念,以古籍的物理形态为单元进行加工整理,而未能深入到古籍所包含的信息单元和知识内涵,忽略了用户的深度使用需求,也没有发挥数字媒介、数据库和数字工具的可供性。

而古籍数字化再造是在古籍数字化的基础上,以古籍所包含的信息或知识单元为基础,通过这些知识单元的分合与重组来挖掘和解构古籍内容,创新古籍呈现方式,实现古籍内容在数字媒介上的再表达,满足数字人文研究和数字文化传播的新兴需求。古籍数字化实现了古籍从“藏”到“用”的过程,古籍数字化再造则旨在实现从“用”到“活”的过程。在数字化技术的不断进步下,古籍数字资源可以实现更深层次和更广领域的文献加工整理,古籍数字化再造正是古籍数字化工作向数据化、智慧化方向的提升。同时,通过满足古籍相关数字人文研究、文化传承与传播、文化产业发展等方面的数据资源需求,古籍数字化再造以高质量的数据资源和多样的呈现形式,为古籍的创造性转化和创新性发展提供支撑与保障。

表 1 古籍数字化与古籍数字化再造的比较

| 视角 | 阶段 | |
|----|-------|---------|
| | 古籍数字化 | 古籍数字化再造 |
| 目标 | 藏→用 | 用→活 |

续表 1

| 视角 | 阶段 | |
|------|-------------------|--------------------------|
| | 古籍数字化 | 古籍数字化再造 |
| 加工对象 | 古籍实体文献 | 古籍数字资源 |
| 技术方法 | 图像扫描、OCR 识别、自动标点等 | 人工智能、智能计算、语义关联、虚拟现实、元宇宙等 |
| 数据粒度 | 粒度较粗,文献层面 | 粒度更细,知识层面 |
| 利用方案 | 单一孤立的古籍信息查阅 | 复合多维知识挖掘与价值发现 |
| 服务模式 | 文献服务与数据服务 | 知识服务与智慧服务 |
| 开发路径 | 人工编校、机器辅助 | 专家、大众以及机器的协同参与 |
| 面向用户 | 以相关研究学者为主 | 研究学者及大众用户 |

2 古籍数字化再造的理论基础

古籍数字化再造是一项系统性工程,其理论体系仍在不断完善。它与大数据、数据科学、数字人文等领域都有着密切的关联,理论基础主要有知识组织、文化基因、数字记忆、智慧数据等。

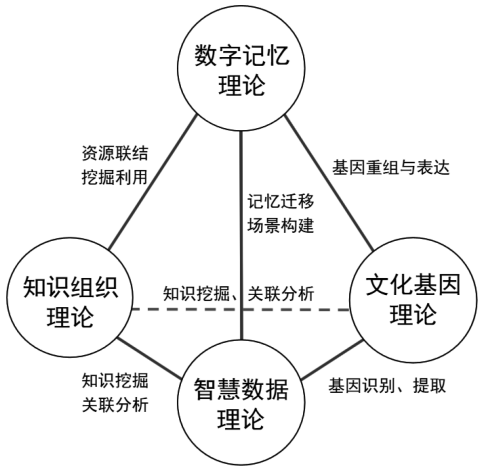


图 2 古籍数字化再造的理论基础

2.1 知识组织理论

知识组织是以知识为对象的诸如整理、加工、表示、控制等一系列的组织化过程及其方法。从经、史、子、集分类法的应用到当代分类法、主题法、元数据、叙词表、概念图、主题图、本体等,无一不是对知识组织进行的探索和尝试。^[7]从知识组织的理论视角出发,古籍数字化再造是对古籍资源进行深度加工的知识组织,其理论可以为古籍数字化再造的知识挖掘、关联分析等提供指导和支撑。

在现有古籍知识组织实践中,利用叙词表、本体及开放关联数据构建的国史本体^[8]、二十四史本体^[9]、先秦古籍本体^[10]等揭示了不同古籍文献的知识结构。一些馆藏机构也通过对古籍版本、典故、注释、人名、地名等的知识标注、历史事件抽取,构建了如中国历史地理信息系统(CHGIS)、中国历代人物传记资料库(CBDB)、

《左传》知识库^[11]、方志物产知识库^[12]、《史记》的多维知识重组^[13]、典籍知识图谱^[14]、中国古典知识库(CCKB)^[15]等相关专题知识库。这些探索性实践在古籍内容的自动分词、词性标注、句法分析、关联知识发现等方面取得了优异成绩,推动着古籍文献资源向深层次知识组织和知识管理迈进。

2.2 文化基因理论

和生物进化过程中基因的作用类似,对于人类社会的发展而言,文化基因是文化传承与传播的最小因子,是文化系统存在、变革、演化和发展过程中最本源的要素,也是知识遗传与变异的最小功能单元^[16-18],它可以以语言符号、思想观念、行为方式等有形或无形的任何方式呈现。古籍承载了丰富的古代文化知识体系和精神内核,它是文化基因的一种外在表现,而文化基因的发掘与呈现是古籍数字化再造的核心要义。

从信息论的视角来看,文化基因是可以复制、选择、传播、交换且具有语义的信息单元,可以实现信息处理或计算。^[19]古籍数字化再造需要借助大数据、语义组织、知识图谱等数字手段,如相关传统文化的语义表征模型,开展古籍文化基因的特征、挖掘、提取和量化,实现古籍外在表现形式与文化内涵的映射,从而支撑中华民族文化基因库的构建^[20]。文化基因理论作为古籍智慧数据建设和数字化再造的基础理论之一,将突破原有的学科藩篱,使古籍焕发出更强大的文化再生能力。

2.3 数字记忆理论

在文化记忆的视域下,古籍就是文化记忆的一种载体形式,是通过文字、图像、符号等信息所承载的文化记忆。数字记忆即是数字形态的文化记忆,通过识别文化内涵的基因特征,将特定历史文化信息以数字化方式采集、组织、存储或展示,并在网络空间承载、再现和传播。^[21]

“数字记忆”的发展,从实践、方法和技术层面使文化记忆机构的一体化发展成为正在发生的现实。^[22]以数字记忆理论为指导,数字空间下的古籍文化记忆建构将成为古籍数字化再造的一种实践场景。古籍数字化再造可面向记忆建构和文化传承,根据特定功能或需求构建项目型实践,在古籍中的各类记忆元素之间建立关联与联想,创造出体验化、交互性、虚实融合的数字记忆场所,改变人们感知历史和文化的的方式,创新文化知识传承与传播的路径。

2.4 智慧数据理论

智慧数据是信息资源的高级编码方式,是一种包括了高级资源组织形态(未来的形态)和最原始资源组织形态(历史上的形态)的混合数据状态,它具有富语义性、可计算性、可解释性和可互操作性等核心特征。^[23-24]智慧数据不仅结构化、语义化和关联化程度更高,也更为强调数据的可信度和可溯源性,是通过已知来试图揭示未知的未知^[25]。

构建智慧数据是建设数据网络、发掘数据资源隐藏价值的关键^[26],它通过数字资源的语义链接,实现不同知识类型的汇聚,从而打通从实物(或事实)到数据、到信息、到知识,再到智慧数据的生产链条,以此形成一个跨领域、跨模态、跨时空的知识网络。智慧数据借助多源异质数据的融合、关联、分析等活动,实现决策辅助和智慧服务。在智慧数据的理论指导下,古籍数字化再造通过对古籍内容资源的整合、清洗和再组织,尝试对其所蕴含的知识和文化进行更进一步的挖掘、表达、关联、重组、呈现和传播。构建智慧数据将深化古籍的内容

层面,把古籍信息转化为结构化、语义化、关联化、计算机可分析、可计算的数据资源,为古籍数字化再造的文化基因提取和数字记忆建构提供智能化的数据资源基础。

综合来看,知识组织和智慧数据理论可共同指导古籍数字化再造的资源建设,使古籍信息与知识体系在智慧数据建设中实现进一步的关联、融合、分析与发现;文化基因和智慧数据理论则促进了古籍研究的跨领域融合,识别和提取古籍文化基因,不仅是古籍实现传承和发展中华优秀传统文化的重要前提,所提取的文化基因也是古籍智慧数据的一种表现形态,可以进一步丰富古籍知识体系,并提升相关的智能化知识服务水平;同时,在更高的知识呈现和应用层面,以数字记忆理论为基础,由文化基因的重组与再表达、智慧数据的场景构建和记忆迁移所共同形成的古籍数字记忆,也为古籍数字化再造提供了实践场景。在古籍数字化再造研究与建设中,该理论框架可以为多层次、多维度的古籍数字化再造实践提供理论指导。

3 古籍数字化再造的堆栈模型

古籍数字化再造包括了对古籍内容载体、知识组织、呈现形态等多个层面的创新与再造。参考数字人文堆栈^[27],本文提出古籍数字化再造的堆栈模型(图3),阐释古籍数字化再造过程中资源开发、技术应用、成果呈现等多方面的内容与目标。古籍数字化再造将依托数字技术,将原始的古籍数字化内容转换为数字空间下便于再现和传播的数据形态,以智慧数据承载古籍的文化内涵与文化表征,满足数字环境下跨学科、跨地域、跨机构的人文学术研究和多元文化传播的需求,实现对古籍资源新的视角解读、新的知识组织和新的呈现方式。

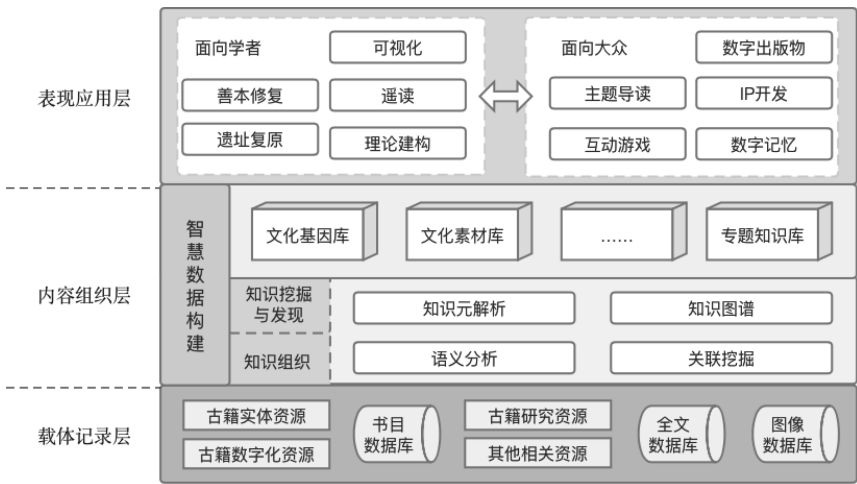


图3 古籍数字化再造堆栈模型

3.1 载体记录层:从纸张载体到数字媒介的再造

古籍文献资料从纸质载体到数字媒介的迁移,不仅是古籍数字化的实现目标,也是古籍数字化再造开展的前提和基础。随着数字技术的不断升级,大量古籍资源被数字化,形成了数量庞大、内容丰富的现代人文学术资源池和数据库。不同于古籍数字化阶段,古籍数字化再造在信息载体迁移的基础上,突破以文本为单位的封闭式知识管理方式,通过相对开放的网络式知识管理模式,将原先由于不同知识分类而隔绝于不同

知识领域、组织结构中的古籍文献资料,经过进一步的数据化处理、规范且统一的聚合关联,实现古籍信息结构的重组与再造。在载体记录层,它不是存粹的资源存储空间,而是以内容为根本、以先进技术为支撑、互联互通的资源创新管理体系,体现了对古籍实体资源和数字资源的多元管理能力。

古籍数字化再造的资源建设不仅解决古籍数据粒度不高、描述信息不全、循证信息缺失的资源管理问题,也将通过古籍原生数据和相关数据的共生融合来避免资源的重复建设,规范古籍数据的机器可用性和互操作性,为相关人文研究与文化传播提供资源基础。从传世文献到出土文献,从汉文典籍到少数民族文字典籍,在古籍整理范围、保护水平不断提升,以及研究内容不断深入的情境下,古籍数字化再造的载体记录层旨在提升古籍数据的流通能力、强化资源协同管理,为古籍数字形式的保存和传播提供支撑。

3.2 内容组织层:从孤立文本到知识网络的再造

古籍承载的是庞大的古代文化知识系统,对古代文化的正确解读需要对其知识元及知识元群间的语义关系进行准确描述,形成完整的知识体系。^[28]随着传统线性因果式研究思路逐渐延伸至网状式相关分析视角,古籍知识简单以外部特征进行存储和排序的组织方式,难以满足用户和研究人员信息查询和知识发现的深层次需求。^[29]将信息组织成能够提供知识服务并保证人们有效利用的古籍知识,是古籍数字化再造在内容组织层面的根本任务。

结合人文领域的认知模式和研究特点,古籍数字化再造首先对多版本、多模态、多媒介的古籍资源进行加工汇集,从语义层面进行标注、聚合、关联、重组等;其次,通过知识元解析、语义分析、关联挖掘等方式,对古籍的概念实体和知识单元进行细粒度识别与表示,将古籍知识或文化记忆细粒到数据、事实和知识的最小单元,形成人物知识库、文化基因库等古籍智慧数据资源集合;最后,再借助计算思维、数字人文、智能计算等,建立古籍知识图谱,将人、时、事、地、史、典故、名物等要素组合,实现跨越不同题材、不同文本的知识重组。在古籍数字化再造的内容组织层,古籍数字资源将进行升维再构,以全知型的视角系统、整体地还原和呈现古籍的知识体系,通过古籍智慧数据的建设创新传统的思维方式和研究范式,拓展我们对文献、文本、知识的认知路径和把握尺度,促进古籍数字化向更深的知识组织、知识发现、知识服务方向发展。

3.3 表现应用层:从线性阅读到多维交互的再造

古籍资源与技术的充分融合不仅能够辅助相关学科发现新的研究问题,也有利于进一步丰富古籍文化的再现场景,满足大众日益多元的数字文化消费需求。古籍数字化再造通过对古籍资源的系统性观照,实现大规模、跨时空、细粒度的内容挖掘,通过多维度、多模态知识的再组织和再利用,丰富古籍内容的呈现与表达方式。在把握数字化活动规律和古籍特性的基础上,古籍数字化再造充分挖掘古籍的知识内涵和资源价值,丰富古籍的传承性保护手段。

在全媒体技术和数字内容产品飞速发展的环境下,古籍数字化再造可借助超文本、虚拟现实、互动游戏等媒介载体改变古籍阅读方式,通过打造古籍沉浸式阅读、遥读、VR/AR 阅读、交互阅读等形式,使读者得以具身方式深入古籍文本所建构的数字体验空间,用游戏创作、IP 开发等手段丰富数字文化产业创新发展。同时,古籍数字化再造所提供的知识服务,也为基于古籍文本研究成果的虚拟场景转化、3D 古籍修复、出土文献缀合复原等人文研究的深入开展提供支撑。古籍数字化再造通过古籍智慧数据相关的知识服务与场景化应用,

扩展了古籍的媒介呈现形态,使古籍资源在新时代数字文旅、在线教育与文化传播方面发挥更大的价值和作用,实现优秀传统文化的“活态传承”。

4 古籍数字化再造的进路

古籍数字化再造的提出,无疑会开启新时代古籍数字化工作的新阶段,而古籍数字化整理、研究、出版、利用需求的不断提升,必将要求在以上框架的基础上,进一步发展古籍数字化再造的理论与方法,探索可行的实施方案和技术路线,找到合适的应用场景和可持续发展模式。此处,从理论、技术、用户、机制、模式五个维度,提示古籍数字化再造的进路。

4.1 理论维:发展古籍数字化再造理论

基础理论是科学研究纲领的核心部分或本质特征,决定着科学研究的发展方向。随着数智化的深入,不断强化理论与方法论研究,使古籍数字化工作获得整体的理论性指导与支撑,才能将古籍数字化工作推向一个新台阶。^[30]一方面,知识组织、文化基因、数字记忆、智慧数据这些古籍数字化再造的基础理论,本身还是动态演化和不断完善的过程。尽管学界已经开始关注智慧大数据建设、文化基因表达和数字记忆建构等领域,但现有研究中理论稍显滞后于实践,具体到古籍就更为匮乏,极易导致古籍数字化再造由于建设方向与范畴的不清晰而出现偏颇。另一方面,古籍数字化再造仍然需要多学科理论的不断融入,比如数字叙事、文化组学、文化信息学等,只有不断推进基础理论研究的深入,才能科学指导古籍数字化实践,为古籍整理与保护提供科学的指导思想和清晰的发展方向。同时,“数字文献学”“数字版本学”等新的学科分支发展,也与古籍数字资源的活化利用有着无法割舍的关系。

4.2 技术维:突破古籍智能分析与计算关键技术

技术的发展始终推动着古籍整理研究工作的不断进步。当前,古籍数字化关键技术主要包括图像扫描、光学字符识别、自动句读与标点、古文分词以及命名实体识别等^[31],而专门针对古籍内容如词义标注、句法分析等不同层级任务的尚不多见;机器学习、知识图谱、地理信息系统(GIS)、虚拟仿真、数字孪生等技术与古籍内容挖掘、场景呈现的结合仍待进一步强化;新兴的元宇宙、拓展现实(VR/AR/MR)、区块链、NFT(Nonfungible Token,非同质化代币)/NFR(Non-fungible Right,非同质化权益)等技术在古籍数字资源的智能化管理与应用还需要拓展探索。同时,针对简帛、甲骨文、敦煌残卷等出土文献的古籍文本恢复、古文字自动补全与推理等也需要进一步提高计算机技术的辅助功能,例如借助深度神经网络来预测古希腊铭文中缺失的文字,并评估其地理属性和时间属性^[32]等。总之,古籍数字化再造需要在知识分析、知识挖掘、可视化呈现等方面不断突破核心与关键技术,探索不同技术与古籍数字资源融合的可能性,增强古籍数字化再造各个环节的技术支撑。

4.3 用户维:分析不同目标用户的再造需求

古籍数据资源建设和知识服务的现实需求是古籍数字化再造不断发展的导向和依据。古籍数字化利用主体包括研究型 and 一般阅读型^[33],他们对古籍的使用目的、使用方式等都存在差异。因此,了解和明确不同目

标用户的古籍使用习惯和需求,有助于进一步完善和推进古籍数字化再造。面对数字人文驱动的人文学术研究变革,古籍数字化再造需要进一步深化古籍数字阅读、学术交流、细粒度内容分析、知识挖掘等方面的用户需求,提高科研人员资源获取的效率和准确率。而面对社会公众对优秀传统文化的需求,古籍数字化再造应通过多元化、立体化的大众传播方式,利用有声阅读、知识讲解、VR/AR 图书、3D 古籍、互动游戏、历史复原等形式提供更丰富、更优质的古籍数字产品;或借助短视频等互联网平台,以公众喜闻乐见的形式讲解和传播古籍里的知识,创新和演绎古籍文化内涵,服务于针对大众的传统优秀文化素养培养和基本知识普及,真正实现古籍文化创造性转化和创新性发展的目标。

4.4 机制维:设计不同主体机构协同参与机制

由于古籍资源的分散性和不易流通的特点,不少图书馆、档案馆、博物馆等古籍收藏机构开始了以馆藏资源为基础的项目驱动型古籍数字化再造实践。从早期对莎士比亚作品、《红楼梦》等单一经典文学和历史文献资源的文本可视化呈现,到当下的“明代吴门书画家书札精品展”^①、“董其昌书画艺术大展”^②,及“数字故宫”“云游敦煌”“唐宋文学编年地图”^③等项目,都在实践运用中探索和丰富了古籍的数字多元表现形态。但与此同时,由于古籍数字化多主体、多领域、跨学科等特性,不同主体间的合作方式、资源融通标准、资源互操作原则等问题也在项目实践过程不断凸显出来。因此,古籍数字化再造需要进一步探索不同主体机构协同参与的发展机制,例如以示范性项目为代表,建立统一的数据管理原则、共享的资源利用模式、可持续性的项目建设标准以及可探索性的功能设计等,实现古籍数字资源建设和服务应用中现实问题、理论难题的兼容解决。

4.5 模式维:探索机构与用户之间的价值共创模式

在古籍数字化再造的理论探索与项目实施中,不仅需要文史哲领域专家团队的系统性挖掘和解读,更需要图档博等文化机构和科技企业的通力协作。只有多元力量的协同发力,才可以破解和改善当前古籍数字化与活化过程中仍然面临的技术、人才、资金等难题。事实上,价值共创作为一种新型的数据资源价值创造模式正在逐渐兴起,数字人文领域已围绕资源、组织、用户三个层面展开了相关研究^[34]。在古籍内容基础和技术应用的双重保障下,探索价值共创的合作方式同样有利于古籍数字化再造的可持续发展。首先,通过价值生产者之间项目合作、联合开发等形式促进彼此的协同创作、资源共享;其次,在具体实践中鼓励社会力量的积极参与,使用户和相关机构成为古籍文化价值的共同创造者,这不仅可以降低成本、提供更好的产品和服务,也可以提高公众的参与积极性,并最终改善用户体验。

5 结语

本文围绕深入推进新时代古籍工作这一指导目标,提出并阐释了“古籍数字化再造”的概念内涵,通过辨析古籍数字化与古籍数字化再造的区别,提炼分析古籍数字化再造的基础理论,从载体记录、内容组织、表现

① <https://www.shanghaimuseum.net/museum/shuangli/index.jsp>
② <https://www.shanghaimuseum.net/museum/dongqichang/index.html>
③ <https://sou-yun.cn/poetlifemap.aspx/>

应用三个层次构建了古籍数字化再造的堆栈模型,并论述古籍数字化再造过程中资源开发、技术应用、成果呈现等多方面的内容与目标。最后从基础理论、技术方法、需求分析、机制模式等方面提出古籍数字化再造的现实进路。

古籍数字化再造是在古籍数字化基础上开展的古籍内容再发现、再组织、再表达的系统性工程,是未来古籍数字化工作新的探索性阶段。古籍数字化再造工程的实施不仅可以加速古籍数字资源的转化利用、创新古籍数字出版模式、提供更丰富的古籍知识服务体系、推动相关人文研究范式的转型升级,还将深入挖掘古籍所蕴含的文化内涵和时代价值,扩展和丰富古籍文化的传播范围、应用场景,促进数字文化产业的高质量发展,更好发挥古籍工作在传承和弘扬中华优秀传统文化中的地位和作用,从而推动新时代古籍数字化工作的开展和古籍整理出版研究范式的转型升级。

参考文献

- [1] 杨牧之. 古籍整理与出版专家论古籍整理与出版[M]. 南京: 凤凰出版社, 2008: 278
- [2] 黄水清, 王晓光, 夏翠娟, 等. 推进新时代古籍工作, 加快创新智能化发展[J]. 农业图书情报学报, 2022, 34(5): 4-20.
- [3] 何维义. 从《中华再造善本》谈再造古籍的开发利用[J]. 图书馆界, 2008(2): 31-33.
- [4] 陈力. 数字人文视域下的古籍数字化与古典知识库建设问题[J]. 中国图书馆学报, 2022, 48(2): 36-46.
- [5] 李明杰, 张纤柯, 陈梦石. 古籍数字化研究进展述评(2009-2019)[J]. 图书情报工作, 2020, 64(6): 130-137.
- [6] 郑永晓. 加快“数字化”向“数据化”转变——“大数据”、“云计算”理论与古典文学研究[J]. 文学遗产, 2014(6): 141-148.
- [7] 成全, 罗栋, 钟晶晶. 知识组织的理论缘起及演进路径探析[J]. 图书馆论坛, 2014, 34(11): 26-34.
- [8] 王颖, 张智雄, 孙辉, 等. 国史知识的语义揭示与组织方法研究[J]. 中国图书馆学报, 2015, 41(4): 55-64.
- [9] 董慧, 徐雷, 王菲, 等. 语义分析系统研究(Ⅱ)——史籍推理机制[J]. 情报学报, 2014, 33(2): 195-203.
- [10] 何琳, 陈雅玲, 孙珂迪. 面向先秦典籍的知识本体构建技术研究[J]. 图书情报工作, 2020, 64(7): 13-19.
- [11] 李斌, 王璐, 陈小荷, 等. 数字人文视域下的古文献文本标注与可视化研究——以《左传》知识库为例[J]. 大学图书馆学报, 2020, 38(5): 72-80+90.
- [12] 徐晨飞, 包平. 面向农史领域的数字人文研究基础设施建设研究——以方志物产知识库构建为引[J]. 中国农史, 2019, 38(6): 40-51.
- [13] 张琪, 王东波, 黄水清, 等. 史书多维知识重组与可视化研究——以《史记》为对象[J]. 情报学报, 2022, 41(2): 130-141.
- [14] 欧阳剑, 梁珠芳, 任树怀. 大规模中国历代存世典籍知识图谱构建研究[J]. 图书情报工作, 2021, 65(5): 126-135.
- [15] 刘石, 孙茂松. 关于建设“中国古典知识库”的思考[J]. 人民政协报, 2020-08-24(8).
- [16] 吴福平, 李亚楠. 文化基因概念、理论及学术史批判[J]. 深圳社会科学, 2020(6): 96-103+122.
- [17] 赵传海. 论文化基因及其社会功能[J]. 河南社会科学, 2008(2): 50-52.
- [18] 孔宪铎, 王登峰. 基因与文化[J]. 东岳论丛, 2010(2): 5-11.
- [19] 刘长燕, 酆全民. 文化进化的模因视角及其哲学意蕴[J]. 哲学分析, 2022, 13(1): 154-164+199.
- [20] 赵海英. 文化基因研究缘起、进展与未来研究思考综述[J]. 中国传媒大学学报(自然科学版), 2021, 28(5): 1-10.
- [21] 冯惠玲. 数字记忆: 文化记忆的数字宫殿[J]. 中国图书馆学报, 2020, 46(3): 4-16.
- [22] 夏翠娟. 构建数智时代社会记忆的多重证据参照体系: 理论与实践探索[J/OL]. 中国图书馆学报: 1-21[2022-07-01].
http://kns.cnki.net/kcms/detail/11.2746.G2.20220413.1505.003.html.
- [23] 曾蕾, 王晓光, 范炜. 图档博领域的智慧数据及其在数字人文研究中的角色[J]. 中国图书馆学报, 2018, 44(1): 17-34.

- [24] 王晓光. 数字人文与智慧数据[J]. 上海高校图书情报工作研究, 2018, 28(2): 25 + 24.
- [25] ZENG M L. Smart Data for digital humanities[J]. Journal of data & information science, 2017(1): 3 - 14.
- [26] 李宗俊, 范炜. 面向智慧数据的语义丰富化方法及应用评述[J]. 情报科学, 2021, 39(7): 186 - 192.
- [27] 大卫·M. 贝里, 安德斯·费格约德. 数字人文数字时代的知识与批判[M]. 王晓光, 等译. 沈阳: 东北财经大学出版社, 2019: 24.
- [28] 常颖聪, 路程, 翟军平. 基于关联数据的古文知识组织应用研究[J]. 图书馆理论与实践, 2019(2): 55 - 59.
- [29] 刘石, 李飞跃. 大数据技术与传统文献学的现代转型[J]. 中国社会科学, 2021(2): 63 - 81 + 205 - 206.
- [30] 王珏. 新中国七十年古文献学的发展: 特征、趋势及问题[J]. 郑州大学学报(哲学社会科学版), 2020, 53(2): 97 - 104.
- [31] 苏祺, 胡韧奋, 诸雨辰, 等. 古籍数字化关键技术评述[J]. 数字人文研究, 2021(3): 83 - 88.
- [32] ASSAEL Y, SOMMERSCHIED T, SHILLINGFORD B, et al. Restoring and attributing ancient texts using deep neural networks [J]. Nature, 2022, 603: 280 - 283.
- [33] 王立清. 关于多元古籍数字化主体的探讨[J]. 图书馆学研究, 2011(7): 53 - 58 + 18.
- [34] 赵宇翔, 张妍, 全冲, 等. 面向价值共创的数字人文类开放数据竞赛行动者协作行为研究[J]. 情报理论与实践, 2022, 45(1): 112 - 119.

The Logic and Approach of Digital Reconstruction of Ancient Books in the Data Intelligence Era

Lei Jueying Hou Xilong Wang Xiaoguang

Abstract With the continuous development of the ancient books protection plan and the digital publishing of the ancient books, China has accumulated a huge number of ancient books digital resources. Under the environment of data intelligence, how to apply new technologies and new ideas to realize the extraction, display, activation and utilization of excellent cultural genes contained in ancient books has gradually become the main goal and core task of ancient books digitalization in the new era. Based on the digitalization of ancient books, this paper puts forward the concept of "digital reconstruction of ancient books", and analyzes the differences between the two concepts. Then it refines the theoretical basis of digital reconstruction of ancient books, and constructs a stack model from the carrier record layer, content organization layer and application layer. The paper also puts forward the realization approach of digital reconstruction of ancient books from the five dimensions of theory, technology, users, mechanism and mode, which provides a reference for the in-depth development and activation of ancient book resources. Digital reconstruction of ancient books is an important part of the activation of ancient books in the future, which can provide a new guiding concept for the work of ancient books in the new era.

Key words digital reconstruction of ancient books; smart data; digital humanities; knowledge organization