



古籍知识库互联互通框架研究与设计^{*}

□陈涛 杨鑫 夏焱 苏日娜

摘要 加强古籍数字化资源汇聚共享和互联互通是新时代古籍工作的必要之举,现有的中心化的古籍资源融合方式存在维护成本高、更新不及时、重复使用难等症结。文章依据数字人文应用中较为成熟的语义技术,从元数据转换和发布、图像资源互联互通、内容资源互联互通、语义资源互联互通等方面构建古籍知识库互联互通框架。该框架在确保各存藏机构现有古籍数据库结构不变的前提下,采用分布式架构实现古籍资源去中心化的、实时的在线互联互通。这种互联互通方式既可以用于古籍数字化基础设施建设,也契合文化数字化战略的建设需求,具有一定的应用前景。

关键词 古籍知识库 互联互通 去中心化 文化数字化

分类号 G253.5

DOI 10.16603/j.issn1002-1027.2023.04.008

1 引言

古籍是中华民族智慧的结晶和宝贵的精神财富,是坚定文化自信的重要源泉,是中华千年文明传承和发展的重要载体。习近平总书记在中国人民大学考察时强调,“要运用现代科技手段加强古籍典藏的保护修复和综合利用,深入挖掘古籍蕴含的哲学思想、人文精神、价值理念、道德规范,推动中华优秀传统文化创造性转化、创新性发展”^[1]。2022年4月,中共中央办公厅、国务院办公厅印发《关于推进新时代古籍工作的意见》,在加快古籍资源转化利用方面提出“积极对接国家文化大数据体系,加强古籍数据流通和协同管理,实现古籍数字化资源汇聚共享”的具体要求^[2]。2022年10月,全国古籍整理出版规划领导小组印发《2021—2035年国家古籍工作规划》,其中第六项要求“统筹古籍数字化建设”,对古籍数字化工作提出具体政策举措:“加强古籍数字化基础设施建设,推进古籍数字化总平台、骨干平台建设,加快构建古籍资源中心、骨干资源库,形成以平台为牵引、资源中心和资源库为依托的古籍数字化资源体系,实现互联互通和共享利用”^[3]。”

由此可见,古籍数字化是国家古籍工作部署的重点,古籍数字化的进一步开发和利用是未来古籍保护工作的关键基础。古籍原生性保护、再生性保

护和传承性保护构成了古籍保护的完整体系^[4],而古籍数字化理应在整个体系中处于核心地位。过往关于古籍数字化的相关研究多集中于古籍数字化专题数据库的构建、各项具体技术的研究和实施等方面,鲜有从技术统筹和互联智慧的宏观视角出发的相关研究。从国家近几次古籍工作指导意见中可以看出,加强古籍资源互联互通和共享利用是开展古籍深度利用和持续研究的根本,这也契合《关于推进实施国家文化数字化战略的意见》中关于文化数字化基础设施和服务平台建设的战略要求。本文从顶层设计的思路出发,试图提出古籍资源的互联互通概念,并从古籍图像互联、内容互联、语义互联三个层面探讨构建古籍知识库的互联互通框架,以实现古籍数字化内容的深度挖掘和智慧共享,进一步挖掘、延续和发展古籍的时代价值。

2 古籍数据库应用现状

从当前古籍工作现状来看,提升古籍数字化水平、加快古籍智能化发展、推进古籍数据库和相关平台建设是目前古籍数字化工作的重点。

古籍数字化和智能化研究是古籍数据库建设的基础,古籍数字化程度直接影响着古籍数据库的智能程度以及应用的推广程度,其重要性不言而喻,可

^{*} 2022年度古籍工作立项重点课题“古籍数字化资源平台互联互通研究”(编号:22GJK008)的研究成果之一。

通讯作者:苏日娜,ORCID:0000-0002-1060-6511,邮箱:surina@mail.sysu.edu.cn。



视为对古籍发现、保护、利用的延续、传承与创新,是人文学术研究的基础性、时代性工程^[5]。狭义的古籍数字化主要是对实体古籍进行数字化转换,表现为对古籍资源进行计算机编码汉字和计算机图像识别并转换为字符的能力^[6],常用的方法是借助光学字符识别技术进行古籍图像版式和文字字符的识别^[7],这些是进一步深化古籍研究的前提。而广义的古籍数字化还包含针对古籍数字资源的一系列整理、利用和研究,如使用自然语言处理、机器学习等方法对古籍进行断句、标点、词语切分、实体识别^[8-10],进而实现自动摘要生成^[11-12]、关系图谱分析^[13]、地理信息系统时空分析^[14]。在古籍数字化发展的基础上,有学者指出对古籍内容进行再挖掘、再组织与再表达,应用本体、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进行再造,并转化为智慧数据,以提供更为智慧的数据支撑^[15-16]。

古籍数据库建设是使古籍为人所熟知的重要手段。根据建设和应用情况来看,诸如国家图书馆牵头建设的“中华再造善本”数据库、“中华古籍资源库”以及爱如生公司开发的“中国基本古籍库”“中国数字方志库”等重大项目,都在近二十年间不断发展,除为专业学人提供研究渠道和途径之外,也让社会公众有机会接触到古籍文献资料,从中获取古籍知识。从近些年古籍数字化建设发展变化来看,数字化建设对象由基本常见类逐步向更具学术性和稀有性的古籍转变。《2011—2020 年国家古籍整理出版规划》中“古籍数字化类”共包括中国基本古籍数据库、两汉全书、魏晋全书、全唐五代诗、全宋诗、全宋文、全元文、全元戏曲、全明诗、全明文、二十四史及《清史稿》修订本电子版、二十四史知识分析系统 12 项基本古籍文献类,和殷商甲骨文知识库、殷周铜器铭文知识库、古代简帛文献知识库、古代碑刻文献知识库、《中国文物地图集》电子出版物、中国古代科技典籍集成及数字化工程、续修四库全书(电子版)、乾隆版大藏经 8 项专门古籍类^[17]。从古籍资源互联互通的方式来看,线下古籍数据融合并提供联合目录是主要方式。于 2003 年启动的“CALIS 古籍联合目录”系统致力于揭示、整合国内各高校的古籍资源。2004 年 CALIS 组织建设的“高校古文文献资源库”,努力为全国乃至海内外的“古籍联合编目”提供一个良好的组织模式及技术平台,截至 2019 年,已有 31 家国内外高校图书馆参加组织建

设。该数据库作为 CALIS 三期建设的子项目之一,共收录元数据 67 万余条、书影 30 余万幅、电子书 8.35 万册,已成为世界上规模最大的中国古籍书目数据库之一^[18]。上海图书馆“中文古籍联合目录及循证平台”^[19],收录有上海图书馆、加州大学伯克利分校东亚图书馆、哈佛燕京图书馆等 1400 余家机构的古籍馆藏目录,并实现了各馆古籍珍藏的联合查询和规范控制。华东师范大学“数字方志集成平台”^[20]收录有华东师范大学、北京师范大学、上海师范大学等十余家师范院校的方志藏品资源,并提供方志资源的联合查询。此类数据融合方式主要通过将不同存藏机构的古籍资源提交到某一中心机构,来构建具体的古籍专题平台。

通过上述分析可知,现有古籍数据库实现资源融合的主要方式是由特定中心存藏机构牵头构建,其他参与机构向中心存藏机构提交各自的馆藏目录,通过统一格式转换实现数据的聚合。以上方案是典型的 Web 2.0 时代的数据互联互通方式,其所带来的弊端也显而易见。

(1)中心机构维护成本增加。所有存藏机构提交各自资源到中心机构,数据维护由中心机构负责,增加了中心机构的运维成本,尤其需要针对古籍资源的不同格式进行归一化处理。

(2)资源原始状态难以跟踪。开放数据原则中强调必须开放资源的原始数据,而非加工和修改过的数据。经过中心机构运维的数据在重复使用时将难以对原始数据进行跟踪和循证。

(3)数据更新存在严重滞后。数据更新时,需要各参与机构将各自更新的古籍数据提交到中心机构,并依据流程进行转换更新,整个流程的时效性和统一性难以保证。

(4)数据副本版本难以统一。当某个机构的馆藏资源被多个机构使用时,就会存在严重的数据副本版本不一致问题,尤其是伴随资源的更新,这些数据副本之间的资源同步将异常困难。

3 古籍知识库互联互通框架设计

文章所提及的古籍知识库即古籍数据库的进一步发展形式,古籍知识库的建立是实现古籍资源互联互通的关键,唯有打破传统的元数据组织方式,从知识元的角度对古籍数据进行重组,才能实现知识的互联互通。回溯古籍知识组织的方式,可粗略将



其划分为三个阶段:一是以传统官修目录和私家藏书目录为主的古籍书目著录形式,具体分为叙录体、辑录体和传录体等著录形式,可视为传统的著录形式;二是古籍元数据著录方式,是以标准化、简捷化为著录标准,适用于方便机器读取的著录形式,可视为现代的著录形式;三是古籍本体著录,当前古籍本体的设计和编制尚未形成系统性理论成果,还属于有限范围内的尝试,这一方式可以视为未来的著录形式。

本文所依循的古籍著录方式为古籍本体著录,主要考虑对象为古籍元数据。图1为古籍资源互联互通框架总体设计图,左侧部分为古籍元数据资源自动发布与互联互通框架,通过对不同存藏机构古籍数据库(关系型数据库为主)中结构化数据的映射与重组,实现跨格式古籍资源的统一表示;通过知识库的发布,打破机构间的资源壁垒,实现跨机构资源

的汇聚关联;具体应用中,通过对关联资源的调用,实现跨机构资源的数据融合和知识发现。图中右侧为古籍知识互联互通框架,由古籍图像互联框架、古籍内容互联框架、古籍语义互联框架三部分构成。鉴于古籍领域存在不同形式和不同结构的数据资源,为保证不同种类资源在互通共建下的一致性和通用性,并充分揭示其中的异同,文章确定以图像、内容、语义三者为框架核心内容。其中,古籍图像互联框架主要借助国际图像互操作框架(International Image Interoperability Framework, IIIF)实现不同存藏机构中古籍图像资源的在线交互和互操作;古籍内容互联框架使用Web注释数据模型(Web Annotation Data Model, WADM)对古籍图像以及古籍全文进行内容注释;古籍语义互联框架则使用关联数据的技术和方案,实现跨机构古籍实体资源间的互联,以及与外部关联开放数据集(知识库)的关联。

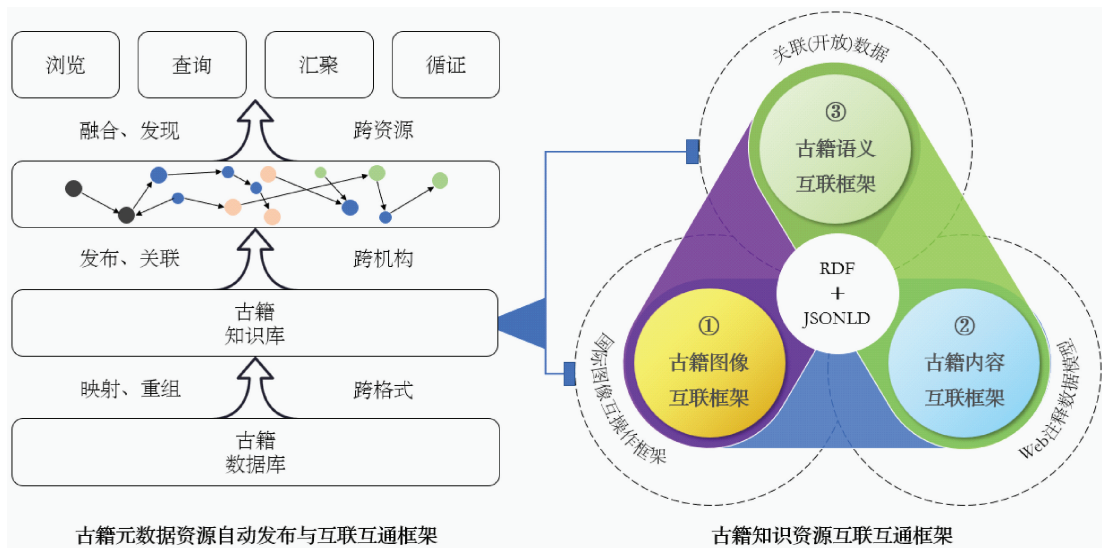


图1 古籍资源互联互通框架

3.1 古籍元数据资源自动发布与互联互通框架

传统的文献资源互联方案(如联合目录)主要采用资源的物理导入,即将不同机构来源数据按照某一固定的元数据格式导入统一的物理空间。该物理空间一般位于某一中心机构中,因此这种汇聚方式并没有实现去中心化。图2为古籍元数据互联互通框架,该框架中无需将各个存藏机构的古籍资源导入到某一中心机构中。各存藏机构仅需部署相同的中间件程序,以实现各自馆藏元数据信息到三元组数据的转换。中间件程序主要通过古籍通用本体和关系型数据到三元组数据的映射语言,如R2RML、

D2R Mapping,将存藏机构结构化数据表转为各自的三元组图,并通过配置SPARQL端点方式提供对外访问,实现跨机构古籍元数据资源的互联互通。相比传统的资源汇聚,该框架有以下优点:

(1)资源去中心化。任一存藏机构通过SPARQL联邦查询(Federated Query)均可实现跨机构古籍元数据资源的在线集成,无需进行线下汇聚,从而去除中心节点,各存藏机构扮演相同的节点角色。

(2)资源访问的实时性。采取中间件程序进行数据的转换并进行发布,可在不同存藏机构间实时

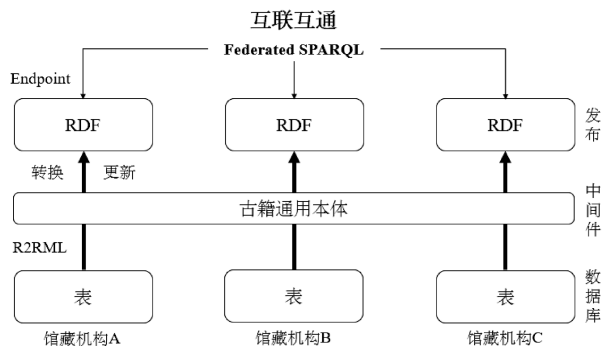


图2 古籍元数据互联互通框架

获取最新的数据资源,尤其是当原始数据发生变更时,更能体现在线、实时获取资源的重要性。

(3)数据交互的便捷性。与传统 API 接口方式相比,通过 SPARQL 端点方式进行古籍元数据的调用,能获得到原生的数据资源,为数据的进一步整合与应用提供便利,也利于数据的溯源。

3.2 古籍图像资源互联互通框架

随着古籍数字化工作的有序推进,越来越多的古籍图像资源逐渐开放,在实现古籍数字化回归、数字化聚合的同时,也带来了古籍研究的新变革——网络化研究。借助相关互联互通框架,可以较好地实现跨机构古籍资源的共享与交互。2015年6月,大英图书馆、哈佛大学等29个非营利图像资源存储机构共同成立了国际图像互操作框架组织^[21],由于该框架在解决跨机构图像资源兼容与互操作问题上具有独特优势,一经提出便在全球文化记忆机构中展现出持续的影响力,众多图书馆、档案馆、博物馆、艺术馆等机构都开始使用 IIIF 框架来呈现各自的馆藏图像资源,同时还构建了庞大的全球图像库^[22-23]。因此,文章选用 IIIF 框架来组织和呈现古籍存储机构间的图像资源,并以此来搭建不同机构图像资源之间的桥梁。

古籍图像互联互通框架如图3所示,主要采用 IIIF 框架对各存储机构中的古籍图像数字化资源进行发布和呈现。此处运用 IIIF 框架中的可编程接口(API),主要是图像 API 和呈现 API,图像 API 通过 Web 服务方式提供面向图像的交互和响应,通过调用图像 URI 实现跨机构古籍资源的在线复用;呈现 API 提供了组织和呈现图像 URI 的接口规范,该 API 中使用的是图像 URI 地址,因此可以实现跨机构图像资源的在线重组。

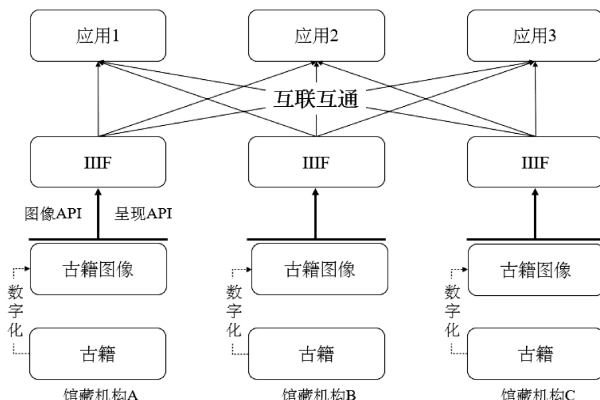


图3 古籍图像资源互联互通框架

3.3 古籍内容资源互联互通框架

古籍内容互联互通研究主要考虑古籍全文和古籍图像两种资源类型,这两类资源的内容注释也是最常见的需求。图4为古籍内容资源互联互通框架,框架统一采用 Web 注释数据模型,该模型主要包括 target 和 body 两部分, target 表示标注的区域, body 表示标注的内容^[24]。对于古籍文本资源,标注粒度可以从整本古籍到每页古籍,甚至是到章、句、词、字等;对于图像资源,则可以针对整张图像或是图像中的任意区域进行注释。标注的内容则可以是多模态类型,如文本、图像、音频、视频、三维模型,甚至可以将一个数据集作为标注的内容添加到特定注释对象中。

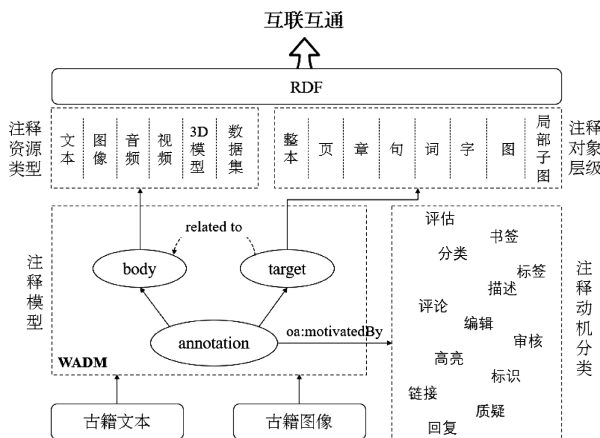


图4 古籍内容资源互联互通框架

此外,内容标注的动机在该模型中也有详细描述,常用的标注动机有评论(Commenting)、描述(Describing)、链接资源(Linking)、高亮(Highlighting)、标签(Tagging)等。对于使用 IIIF 框架进行组织和呈现的图像,其图像本身也将以绘画(Painting)动机形式组织到标注(Annotation)类中。经过该模



型标注后的注释内容为三元组数据,可通过 SPARQL 端点方式提供对外访问接口,实现与其他馆藏古籍资源内容层面的互联互通。

3.4 古籍语义资源互联互通框架

考虑到与其他古籍库和外部开放链接资源间的资源交互,本文采用三元组数据库而非 Neo4j 等图数据库。不同古籍知识库之间的关联可以借助自动关联工具,如 LIMES、SILK 等,也可以使用图向量 RDF2vec 计算多图之间的相似度。图 5 为古籍语义资源互联互通框架,此处的语义资源互联互通主要指不同存藏机构中古籍实体资源的链接关联,即以三元组形式链接不同古籍库中的相关实体,包含本体对齐、模式制定、链接抽取等关键流程。

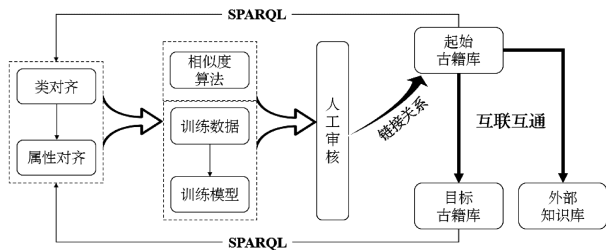


图 5 古籍语义资源互联互通框架

(1)本体对齐。不同古籍知识库建设时本体设计存在差异,因此互联互通首先需要将待关联的起始古籍库和目标古籍库进行本体对齐,通过对齐建立不同古籍本体类和属性之间的映射关系。

(2)模式制定。古籍知识库之间实体(语义)关联映射规则可以采用实体链接关系发现框架(Link Discovery Framework for Metric Spaces, LIMES)中的度量表达式(Metrics)或机器学习方法。①度量表达式通过图形化生成映射语句,如 $\text{cosine}(s.\text{rdfs:label}, t.\text{foaf:name})$ 表示对起始知识库 s 和目标知识库 t 中某类下所有实体的 rdfs:label 属性和 foaf:name 属性值进行余弦相似度计算。②采用机器学习的方式, LIMES 支持有监督和无监督学习方式。事先准备好已经人工关联的实体链接语料, LIMES 通过机器学习自动抽取最优的训练模型。

(3)链接抽取。根据制定的抽取模式(度量表达式或训练模型),从起始古籍库中自动抽取相关联的链接。抽取时可以设定通过阈值(如 0.98)和审核阈值(如 0.9),相似度大于通过阈值的为关联实体,介于通过阈值和审核阈值之间的链接需要进一步进行人工审核。对于有更新数据的古籍知识库,可以定

期进行链接抽取,以完成古籍知识库之间实体资源的语义互联互通。

4 古籍互联互通模式思考

本文从宏观技术统筹的视角,以元数据发布、图像资源互联互通、内容资源互联互通、语义资源互联互通为主要维度构建古籍资源互联互通框架。根据上述技术阐述,古籍知识库互联互通框架的目的和意义在于实现以下四个转变:自建转为共建、自用转为互用、独立转为互联、所有转为所用。

4.1 自建转为共建

古籍资源在较长时期内还未真正实现共建共享,关键原因在于不同存藏机构各自建设不同主题的古籍数据库,在人员配比、数据库选型、技术方案制定、平台实施路径等方面都依据各自的实际情况开展,不仅造成数据结构、技术平台、知识体系的重复建设,更加剧了机构之间古籍资源的差异性和杂糅性。导致这一现状的因素是多方面的:部分存藏机构“重藏轻用”,不愿意开放资源;构建平台的技术体系标准存在异同,缺少统一的技术选型和指导方案,这其中包括古籍数字化流程、著录标引、检索语言的差异性和特殊性,这也是最核心的因素;古籍著录方式多样,加之版本纷杂,底本选取标准不一,以及著录中的异体字和避讳字等问题,导致古籍数据库构建过程中存在多重标准,古籍资源之间难以协调和统一。由此可见,使用统一的结构框架、技术体系构建物理分布、逻辑关联的平台体系,是古籍互联互通的基础,也是古籍资源长久发展、深度利用的前提,更是打破资源壁垒的关键。本文所提古籍互联互通框架的实施即以此为设计出发点,通过统一的知识元和可扩展的语义技术模型,在保持现有古籍数据库结构的基础上,以搭建中间件的形式构建格式统一、语法统一、语义统一的互联空间,实现古籍机构资源自建和资源互联空间共建。

4.2 自用转为互用

各古籍存藏机构已通过联合目录、统一检索等方式实现跨机构的古籍资源的相互访问,然而,这种古籍篇章级的共享粒度在数智时代严重影响了知识的传播力和创造性。正如 WEB 2.0 时代网页超链接逐步发展成知识图谱一样,古籍资源的共享粒度也逐渐向纵深发展,不断追求知识的细粒度化以及知识元间的知识融合。古籍机构构建的数据、创造



的知识在满足自身业务应用的同时,更应为同行机构提供知识复用的可能,由此构建古籍知识复用循环体系。本文的互联互通框架采取三元组模式作为资源描述框架模型,结合关联数据、国际图像互操作框架、Web 注释数据模型、SPARQL 等万维网联盟推荐的多种语义技术方案,并以开放共享的理念贯穿框架始终。

4.3 独立转为互联

技术更迭和信息爆炸的时代特征决定了单一机构难以拥有所有数据,唯有打通机构间、资源间的壁垒,形成共享的古籍资源宝库,才是中华优秀传统文化源远流长、不断创新发展的有力举措。知识的价值并非孤立存在,互联才能促成量变,量变进而引发质变,唯有如此才能更好地推动中华文化的创造性转化和创新性发展。学界针对古籍的语义技术应用研究,主要集中于古籍数字化和数据库构建层面,未深入到古籍资源和知识组织本质,未能探索其资源内容涉及的多个方面以及古籍知识元之间的复杂关联,且缺乏全盘的理论构建研究作为标准和尺度,导致各地各机构所构建的古籍数据库等电子资源难以实现共享共用。因此,探讨古籍资源的语义互联原理,从数字人文的视域下提出古籍资源互联互通的体系框架,不仅有重要的学术意义,对于未来的古籍保护事业发展也具有重要的现实价值。

4.4 所有转为所用

古籍保护的目的是为了让古籍束之高阁。作为全人类的文化遗产,古籍也不应只属于某个收藏机构或个人。为了保护古籍文献,不少图书馆对于商业出版和学术研究不加区分,一概采取“藏”的办法,限制古籍使用,让古籍更加难见天日。更大地发挥古籍的价值才是古籍保护的应有之义。随着数字技术的迅猛发展,很多古籍都已有数字化的替代版本,这些数字版本理应在新时代发挥传承中华文明、延续文化命脉的作用。让古籍走出“深闺”,走进大众视野,让更多人了解古籍、熟知古籍、应用古籍,才是古籍保护工作内在的持续动力。因此,各个收藏机构、文化记忆机构作为古籍文献的管理方,在管理好各自藏品的同时,更有责任开放资源,变所有为所用,做好古籍资源的数字化与知识组织,进一步实现跨机构资源共享,让古籍知识、优秀传统文化知识服务大众、惠及全人类。

5 结论与展望

加强古籍数字化基础设施建设,实现古籍资源互联互通,是新时代赋予古籍研究的新使命。本文结合多种语义技术,在确保收藏机构现有古籍数据库结构的前提下,通过中间件形式实现跨机构古籍资源的互联互通。整体框架的核心技术来源于长期的数字人文应用实践,并已在国内外众多数字人文实践项目中得到成熟应用。将这些技术整合应用于古籍资源设施建设,实现去中心化的跨机构古籍资源的互联互通是一次全新的尝试,这也将改变现有古籍资源交互共享的线下整合模式。线下整合改为线上互联,通过 Web 方式进行互联互通资源的配置,更符合数字时代的资源交互和应用模式,在确保古籍资源实时交互的同时,也带来了应用上的便捷性。

当然,本文框架的实施效果受限于古籍收藏机构的实际情况,如古籍数字化标准规范程度、古籍资源数字化和光学字符识别程度、元数据著录规范化程度等。近年来,国家在古籍工作战略规划中,一直注重标准规范方面的建设,如《2021—2035 年国家古籍工作规划》强调要加快古籍数字化标准建设,建立健全古籍数字化标准规范体系;《关于推进新时代古籍工作的意见》也提出要加强古籍工作标准体系建设,制定修订相关国家标准,完善古籍保护、修复、整理、出版、数字化等工作规范。本课题也将持续关注古籍数字化标准规范方面的政策和行业进展,以期将来可以将数字化标准规范融入古籍互联互通框架研究。

依循文中框架构建互联互通平台的思路除应用于古籍资源建设外,还可以应用于搭建其他文化数据服务平台,以促进文化机构数字化转型升级,推动不同层级、不同平台、不同主体文化数据的共享与交互。世界文明因互鉴而多姿多彩,文化数据也将因互联互通而增值赋能。Web 3.0 时代是万物互联的时代,是价值再造的时代,新的数字技术也将赋予古籍文献更多意义,让中华优秀传统文化在新时代背景之下焕发生机。

参考文献

- 1 新华社.习近平在中国人民大学考察时强调 坚持党的领导传承红色基因扎根中国大地 走出一条建设中国特色世界一流大学新路[N].人民日报,2022-04-26(1).



- 2 中共中央办公厅 国务院办公厅. 中共中央办公厅 国务院办公厅印发《关于推进新时代古籍工作的意见》[J]. 中华人民共和国国务院公报, 2022(12): 30—33.
- 3 全国古籍整理出版规划领导小组. 关于印发《2021—2035 年国家古籍工作规划》的通知[EB/OL]. [2022—12—21]. <https://www.nppa.gov.cn/nppa/contents/279/105197.shtml>.
- 4 姚伯岳, 周余姣, 王莺嘉. 古籍传承性保护再认识[J]. 中国图书馆学报, 2023, 49(1): 58—67.
- 5 王秋云. 我国古籍数字化的研究现状及发展趋势分析[J]. 图书馆学研究, 2021, (24): 9—14.
- 6 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.
- 7 杨海林. 基于深度学习的古籍文字识别和检测[D]. 广州: 华南理工大学, 2020.
- 8 苏祺, 胡初奋, 诸雨辰, 等. 古籍数字化关键技术述评[J]. 数字人文研究, 2021, 1(3): 83—88.
- 9 刘江峰, 冯钰童, 王东波, 等. 数字人文视域下 SikuBERT 增强的史籍实体识别研究[J]. 图书馆论坛, 2022, 42(10): 61—72.
- 10 李娜. 面向方志类古籍的多类型命名实体联合自动识别模型构建[J]. 图书馆论坛, 2021, 41(12): 113—123.
- 11 李惠, 陈涛, 侯君明, 等. 钩玄提要——古籍目录智能分析工具构建[J]. 中国图书馆学报, 2021, 47(4): 97—112.
- 12 徐润华, 王东波, 刘欢, 等. 面向古籍数字人文的《资治通鉴》自动摘要研究——以 SikuBERT 预训练模型为例[J]. 图书馆论坛, 2022, 42(12): 129—137.
- 13 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]. 图书馆情报工作, 2021, 65(5): 126—135.
- 14 李明杰, 杨璐嘉. 基于 GIS 的明代古籍版刻地理信息系统的设计与实现[J]. 信息资源管理学报, 2020, 10(3): 125—133.
- 15 雷珏莹, 侯西龙, 王晓光. 数智时代古籍数字化再造的逻辑与路径[J]. 数字人文研究, 2022, 2(2): 46—56.
- 16 陈涛, 苏日娜, 张永娟, 等. 智慧数据驱动的古籍智慧性保护体系研究[J]. 中国图书馆学报, 2023, 49(1): 68—81.
- 17 新闻出版总署, 全国古籍整理出版规划领导小组. 关于印发实施《2011—2020 年国家古籍整理出版规划》的通知[EB/OL]. [2022—12—21]. http://www.guji.cn/web/c_00000018/d_12200.htm
- 18 姚伯岳. “高校古文献资源库”的建设与发展[M]//国家古籍保护中心. 古籍保护研究(第三辑). 郑州: 大象出版社, 2018: 206—215.
- 19 夏翠娟, 林海青, 刘炜. 面向循证实践的中文古籍数据模型研究与设计[J]. 中国图书馆学报, 2017, 43(6): 16—34.
- 20 鲁丹, 李欣, 陈金传. 基于 API 技术的数字人文基础设施的构建[J]. 图书馆学研究, 2019(13): 42—46, 57.
- 21 国家图书馆研究院. 国际图像互操作组织成立[J]. 国家图书馆学刊, 2015, 24(4): 66.
- 22 单蓉蓉, 陈涛, 刘炜, 等. 国际图像互操作框架及拓展应用[J]. 图书馆杂志, 2021, 40(5): 89—95, 123.
- 23 陈丹, 张毅. 国外大学图书馆国际图像互操作框架应用实践与启示[J]. 图书馆情报工作, 2021, 65(4): 135—143.
- 24 World Wide Web Consortium. Web annotation data model[EB/OL]. [2022—12—21]. <https://www.w3.org/TR/annotation-model/>.

作者单位: 陈涛、杨鑫、夏焱, 中山大学信息管理学院, 广东广州 510006

苏日娜, 中山大学图书馆, 广东广州 510275

收稿日期: 2022 年 12 月 31 日

修回日期: 2023 年 3 月 19 日

(责任编辑: 支娟)

Research and Design of the Interconnecting Framework of Ancient Book Knowledge Base

Chen Tao Yang Xin Xia Yan Su Rina

Abstract: Strengthening the integration, sharing and interconnection of digital ancient book resources is a necessary move for the work of ancient books in the new era. The existing centralized solution of integrating ancient book resources caused problems such as high maintenance cost, untimely updating and difficult reuse. Based on the semantic technologies maturely applied in the field of digital humanities, this paper designed a framework for the interconnection of ancient book knowledge bases, from perspective of metadata conversion and publication, image resources interconnection, content resources interconnection, and semantic resources interconnection. Relying on a distributed architecture this framework has achieved decentralized, real-time online interconnection of ancient book resources without changing the structure of existing databases within collection institutions. It can not only be used in the construction of digital infrastructure of ancient books, but also fits the construction needs of the national cultural digitization strategy. Thus this interconnection approach has certain application prospects.

Keywords: Ancient Book Knowledge Base; Interconnection; Decentralization; Digitization of Culture