

【编者按】图书馆长期致力于文献资源的开发利用,随着信息技术的发展,为了满足用户服务的新需求,其开发利用的方法和模式也在不断发展。文献资源的载体形态从单一的纸质形态和二元的纸电并存向多元复合的多模态形态发展,加工的方法经历了从数字化到数据化再到智慧化的发展历程,服务模式从以资源获得为中心向以数智循证和交互体验为中心转向。近年来,随着数字人文成为 GLAMs 机构、信息资源管理学科和历史人文艺术学科的共同议题,人文资源的开发利用也成为一种跨学科的问题领域。图书馆作为人文研究的基础设施,有必要深入洞察人文研究者的真实需求,人文研究者作为图书馆人文资源的最重要用户群体之一,也有必要了解图书馆在资源开发利用方面的方法、技术和流程。双方在深入了解的基础上加强合作和交流,有利于促进数字人文领域中的学科交融和共同发展。

基于此,《图书馆杂志》与上海市图书馆学会学术委员会数字人文专委会合作组织了“人文资源的数智化开发利用”专题,包括4篇文章,分别从资源加工、服务、应用场景和语料库建设等不同的方面探讨了人文资源的数智化开发利用这一主题,期望能为图书馆更深入洞察人文研究者的需求、人文研究者更深入了解图书馆人文资源开发利用的现有方法并介入其流程提供参考。本期先刊出其中2篇文章:朱武信等的《面向全流程管理的专家协同知识生产系统设计——以家谱联合编目系统为例》一文以上海图书馆开发的“家谱联合编目系统”为例,提出了一种全流程管理和专家协同知识生产的历史人文资源开发利用方法,并创新性地引入了基于大模型的 AI 专家机制。潘威等的《语料库与近代西北水利档案的数智化应用方式探索》一文从西北边疆省份水利档案开发利用的角度出发,提出并实现了一套语料库建设的方法、技术和流程。其余2篇文章将于《图书馆杂志》第9期刊出,敬请关注!

面向全流程管理的专家协同知识生产系统设计与实现*

——以家谱联合编目系统为例

朱武信 夏翠娟 吴建伟(上海图书馆)

摘要 在数智时代,传统编目系统面临着挑战,难以满足基于智慧数据生成的知识生产要求。上海图书馆对传统编目流程进行改进,融入基于众包理念的专家协同知识生产与全流程管理理念,设计了面向全流程管理的专家协同知识生产系统——家谱联合编目系统。该系统实现了:利用上海图书馆数字人文建设已有的人、机构、地、时、事等基础知识库,实现了编目过程中跨网域的规范控制,不同机构的家谱编目专家,共享基础知识库中已有的知识、共同补充丰富编目过程中缺失的知识,实时构建知识关联和语义链接;并将编目流程和知识生产与家谱知识服务平台打通,由此知识生产与知识服务无缝地链接起来,实现了从采购、捐赠、著录、查重、质量检测、规范控制、知识服务、关联数据发布的全流程管理;在此基础上,探索了基于大语言模型的 AI 家谱编目数据审校的路径。

关键词 数字人文 全流程数据加工 联合编目 众包 AIGC

DOI: 10.13663/j.cnki.lj.2024.08.001

* 本文系国家社会科学基金重大项目“编纂《1949年以来中国家谱总目》”(项目编号:18ZDA329)的研究成果之一。
通信作者:朱武信, E-mail: wxzhu@libnet.sh.cn

Design and Implementation of Expert Collaborative Knowledge Production System for Full Process Management: Genealogy Joint Cataloging System

Zhu Wuxin, Xia Cuijuan, Wu Jianwei (Shanghai Library)

Abstract In the age of digital intelligence, traditional cataloging systems are facing challenges and are unable to meet the requirements for knowledge production based on intelligent data generation. The Shanghai Library has improved the traditional cataloging process, incorporated the concept of expert collaborative knowledge production and full-process management based on the concept of crowdsourcing, and designed an expert collaborative knowledge production system for full-process management—the Genealogy Joint Cataloging System. This system could realize the standard control across network domains in the cataloging process by using the existing basic knowledge base of people, institutions, places, times, events, etc. of the Library. Genealogy cataloging experts from different institutions can share basic knowledge. The existing knowledge in the library can jointly supplement and enrich the knowledge missing in the cataloging process, build knowledge associations and semantic links in real time, and connect the cataloging process and knowledge production with the genealogy knowledge service platform, so that knowledge production and knowledge services are seamlessly linked. Together, we have achieved full-process management of procurement, donation, description, duplication checking, quality inspection, normative control, knowledge services, and related data release. On this basis, the path of AI genealogy cataloging data review based on the big language model was explored.

Keywords Digital humanities, Full-process data processing, Joint cataloging, Crowdsourcing, AIGC

0 引言

数字人文带来了人文研究范式的革新，推动了知识生产方式和知识交流模式的变革。本研究旨在为作为知识中介的图书馆重新思考知识生产方式、重组知识生产流程、提升知识服务效果，以更高效快速地适应数字人文研究者对图书馆从提供文献服务转变为提供知识服务的新需求^[1]。数智时代的图书馆的知识生产包含资源的数字化、文本化、结构化、语义化与智慧化，最终形成可操控、可理解、可推理的数据之过程，即本文所研究的全流程管理的专家协同知识生产模式。

本研究基于上海图书馆（下文简称“上图”）“家谱联合编目系统”的设计开发实践，试图探索一种面向全流程管理和专家协同知识生产路径。该系统实现了家谱采购、捐赠、编目、基于互联网的规范控制、专家审校与AI审校、工作量计算和不同维度的家谱编目数据统计分析、编目数据的导入导出、与FOLIO馆藏管理系统和中国家谱知识服务平台无缝集成的全流程管理；还实现了支持多机构在统一的系统上基于一致的著录规范进行家谱编目，利用上图数字人文建设已有的人、地、时、事等

基础知识库，在编目过程中对知识库中的URI参引、新知识注入和语义关联的补充丰富；并初步探索了如何将大模型技术引入到专家协同知识生产的方式。

本文通过对家谱联合编目系统设计实施过程的再现和反思，来探索适应图书馆数字人文服务的、面向全流程管理的专家协同知识生产系统的设计思路 and 实施方案，以为同类项目提供参考。

1 现状调研

数字人文视域下，历史文献资源的知识生产模式一直是一个值得关注的问题。欧阳剑等着重阐述了知识和数据在数字人文研究中的基石地位与核心功能，指出图书馆作为知识生产者需从数字馆藏、数据管理、数据呈现、数据分析的基础上拓展，总结了人文数据需要具备完整性、可计算、可用性、重用性、可发行的特征^[2]。刘炜等论述了基于数字人文的知识生产体系，以知识图谱为代表的语义数据对历史文献资源所产生的积极作用与深层挖掘，分析了数字人文相比于传统人文在研究过程、素材内容、研究方法、技术应用、科研协作与成果

交流上的扩展优势。通过语义化知识的数据架构所产生的智慧数据将更进一步展现历史人文领域的价值^[3-4]，Haslhofer B. 等阐述了知识图谱在图书馆学与数字人文方面通过人名、地名与事件之间的语义关系从而在历史文献资源中产生新的发现^[5]。

跨机构的联合编目系统在图书馆界已经历多年的发展。1997年，全国图书馆联合编目中心 OLCC 成立。OLCC 经过 20 多年的努力，以计算机技术和网络通信优势实现了书目数据资源共建共享的重要途径^[6]。2012年，全国古籍普查登记工作启动，古籍普查编目系统被作为全国范围内的跨机构的古籍联合编目平台，平台著录内容化繁为简、字段完备，深入基层多渠道获取资源，古籍普查编目系统对实现全面高质量著录系统有模范作用。高校图书馆界也有 CALIS 联合编目中心这样的案例^[7]。但这些联合编目系统主要是为数字图书馆时代的特藏知识库建设而服务的，其著录规范主要为揭示文献的载体信息而非细粒度的知识而设计，尽管这些系统也有对规范控制的支持，但其规范控制的范围主要在系统内部，而不是基于互联网的，无法实现跨域网的规范控制。

随着计算机技术的发展，图书馆对大规模、细粒度数据计算和知识服务的需求日益增加，单家机构和单个专家的知识生产无法满足这样的需求，基于众包理念的协同知识生产进入图书馆。2015年，Terras M. 提出数字人文可以结合众包模式来解决知识生产难题，将各方人士的知识与经验通过网络汇聚到一起，并结合数字人文的关联数据，从而在知识资源上取得大而精的成效^[8]。2019年，席运江等论述了传统历史文献加工系统存在的问题，包括数据加工的局部性、重复性、劳动力大、成本高、精度低，并提出了基于历史文献专业性的特点，在采用众包的模式下，除了终端用户的知识盈余贡献外还需要引入专家协同^[9]。2019年，上海图书馆将馆藏历史文献以众包模式进行数据加工，将难以机器识别的手稿以众包模式分发文本识别任务，并引入激励机制，有效地实现了基于众包的知识生产过程^[10]，但由于缺乏领域专家的协同参与，终端

用户生产的内容无法得到快速的反馈和有效的质量控制，在可持续性发展方面遭遇了一定的困难。

2023年，基于 GPT 的大模型技术飞速发展，在各领域都掀起了 AIGC 的研究热浪。张慧等在论文中指出生成式大语言模型是 AI 2.0 时代智慧图书馆的技术驱动与创新，GPT 类技术应用将为图书馆带来新的应用场景^[11]。大语言模型因其强大的自然语言处理能力衍生了各种应用场景，如翻译、摘要、问答与写作。智能推荐、智能搜索、自动摘要与分类、智能翻译、辅助写作与编辑等新场景在图书馆应用中相继出现。与此同时，挑战仍然存在，刘炜等在文章中指出领域大模型应用的重要性，指出通用模型对于领域知识与面对复杂情况存在不足，应对此问题，需要将领域知识喂给大模型，此方法也在 GPT4.0 的模型中得以体现^[12-13]。图书馆界的一些研究者提出了将 AIGC 引入图书馆各业务应用场景中，形成 GPT+ 的创新模式，在图书馆的整个知识生产和知识服务流程中，从数据加工层、模型层、服务层都可以与 AIGC 有效融合，以此进一步提升图书馆业务能力^[14-15]。

综上所述，已有的联合编目系统存在知识生产到知识服务的实时性不足，没有将知识生产与知识服务的全流程打通；在引入众包机制的协同操作系统中，由于参与者自身对领域知识掌握的情况参差不齐，缺乏专家协同支撑，在文献编目和知识生产的效率、质量与专业性方面都难以达到要求。而对于如何将大模型技术应用在知识生产的过程中，大多数图书馆界的研究主要关注大模型的生成能力，而没有将大模型的泛化能力与实际场景有效结合，由于大模型技术仍然存在的幻觉问题，其生成的内容还难以直接用于用户服务，但可以在优化传统编目 workflow、提高效率方面助一臂之力。

2 系统设计

本章将从以下几个方面再现家谱联合编目系统的设计思路。①系统应用场景和需求概述，以用例图说明系统参与者、用例与使用场

景之间的关系。②系统总体架构，基于业务需求的全流程知识生产系统架构设计。③家谱著录规范设计，兼顾基于家谱本体的关联数据生成。④全流程知识生产 workflow，以时序图说明家谱编目工作的主要流程。⑤规范控制，在编目流程中融合数字人文基础知识库数据重用、关联和丰富。⑥人类专家与 AI 专家编目审校功能设计。

2.1 系统应用场景和需求

家谱联合编目系统是家谱馆藏机构对家谱

进行编目和管理的关键工具，为家谱编目工作提供跨机构的联合编目支持，同时作为家谱知识服务平台提供知识生产支撑。各馆藏机构可在平台上进行家谱知识的共享与共建，完成家谱数据编目，实现数据的出版与服务。系统根据加工流程分为 3 类参与者，分别是编目人员、机构专家、编委会专家。组织结构图如图 1 所示，专家协同编目与审校为知识生产给予助力，通过审核反馈机制提升了知识质量与知识生产者的能力。

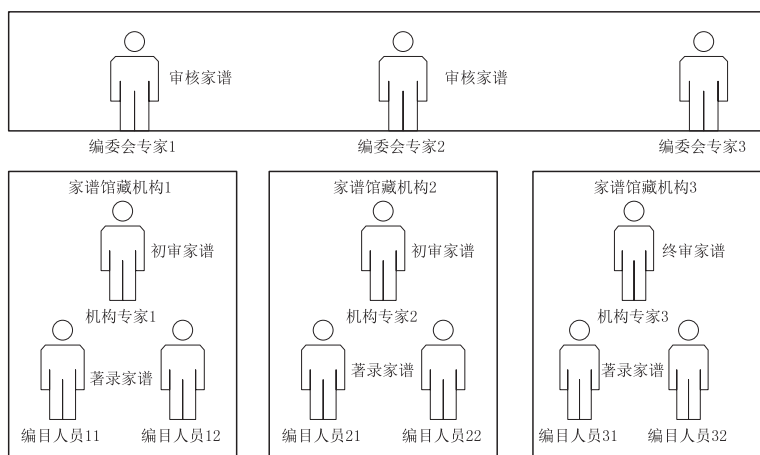


图 1 系统组织图

系统应用场景如图 2 所示，联合编目机构的著录专家负责家谱关联数据的著录工作，其中包括了数据查重、著录与修正，作为整个流程中的知识生产者与发起者承担着重要的生产力供给。机构专家对本机构下的编目数据进行初步审校，主要包括检查数据的完整性、规范性、正确性，作为编目规范的责任者，其熟知本机构馆藏，因而有效地提高了编目的衔接性

与统一化。编委会专家则对来自各机构的数据进行审校，跨机构地以出版标准来对编目数据进行质量把关，将编目数据汇总起来，实现数据共享与服务。

2.2 系统总体架构

系统架构图如图 3 所示，上图的系统设计遵循微服务设计理念，以功能为单元进行模块划分，以 API 的方式实现数据互通^[3]。家谱联合编目系统与上图的用户统一认证平台、数字人文开放数据平台、家谱知识服务平台互联互通。

(1) 用户统一认证平台是围绕用户信息提供认证服务的平台，主要功能包括用户注册、第三方授权登录、重置密码、用户管理和信息安全等功能。该平台为家谱联合编目系统提供了用户认证与注册功能，支持家谱联合编目系统用户的统一身份管理，而具有不同权限的角

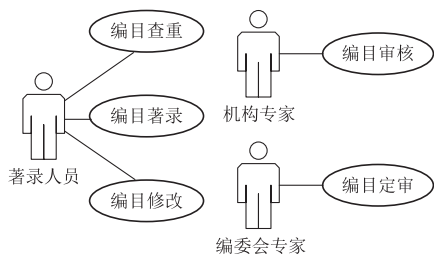


图 2 系统用例图

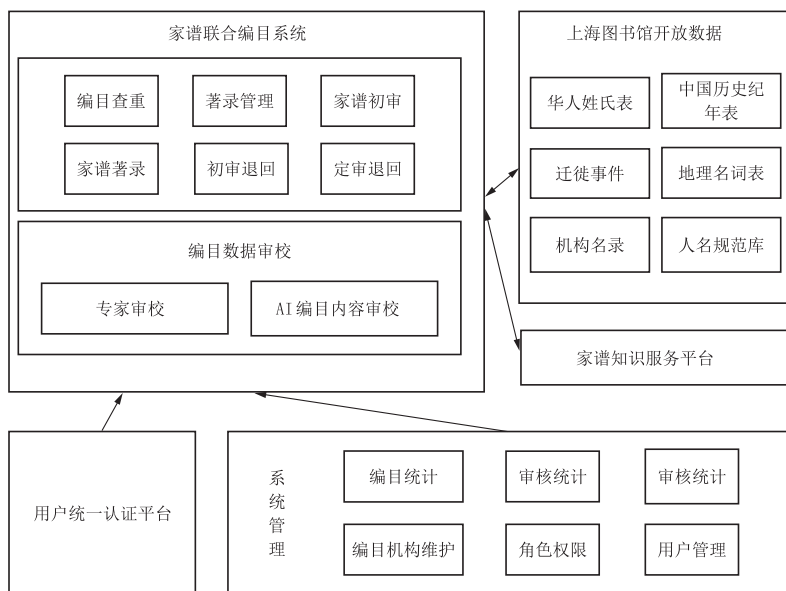


图3 系统架构图

色在家谱联合编目系统中独立管理。

(2) 数字人文开放数据平台是以内容协商与 API 接口方式提供开放数据。包括机构、纪年表、地理名词、姓氏、迁徙事件、历史名人等基础知识库。家谱联合编目系统采用这些数据作为提供规范控制的规范档^[10]。在编目工作界面,基础知识库中的人、地、时、事、姓氏、机构通过 HTTP URI 参引的方式支持规范控制。

(3) 家谱知识服务平台面向家谱爱好者、寻根问祖者、修谱续谱者以及历史人文研究的专家学者,提供家谱文献检索、全文浏览、世系表和迁徙事件可视化等服务。家谱知识服务平台不只是从编目系统获取新的编目数据,同时也作为编目查重的数据中心,著录用户可以通过标题、责任者、谱籍地、姓氏、始迁祖、版本等关键字段进行数据检索,在编目流程发起之前进行数据查重,以此实现已有数据的重用和修改。由于家谱本体是基于作品-实例-单件的 BIBFRAME 模型建模,因而在编目时,可修改作品和版本的数据,也可以重用已有的作品和版本数据,仅增加单件数据。

(4) 家谱联合编目系统分为两大模块,即编目著录模块、数据审校模块。多家机构的家谱整理研究专家在平台上完成编目和审校工

作,数据将实时同步到家谱知识服务平台在互联网上为所有用户提供服务。在编目过程中,新的名人、姓氏、机构、迁徙事件等知识将会实时补充到对应的基础知识库中,作为为所有著录用户在互联网上提供规范控制的规范档。数据审校系统是人类专家与 AI 专家审校编目数据的系统,其中 AI 编目数据审校是基于大语言模型对编目数据的质量打分。AI 以专家与普通用户两个角色对编目数据进行审校打分与评价,人类专家在审校过程中参考 AI 的打分结果与评价,起到降低审校难度提高审校效率的作用。

通过与不同平台的实时交互,家谱联合编目系统实现了采购、捐赠、编目、审校、服务、规范控制的全流程管理。

2.3 家谱著录标准规范

家谱联合编目系统首先采用了项目组设计的《1949 年以来中国家谱总目著录项目规范表》^[16]作为统一的著录规范,通过系统界面的数据验证和提示功能实现了不同机构不同著录用户编目规则的标准化,有效规避了因机构编目标准的不同而出现差异的情况。基于《1949 年以来中国家谱总目著录项目规范表》中的著录规范,对谱籍地、姓氏、书名项、责任者方

式、版本项、内容提要项、装订项等字段进行约束^[16]。主要是对编目时输入文字内容进行自动检测，通过正则表达式与统计工具对著录的文字如简繁体、字长、标点符号进行校验。还支持异常检测与缺省补充，对著录中漏填的信息予以提示，例如必填项责任者字段未填写，提示需补充其内容。

另外，由于家谱联合编目系统中的数据需要与数字人文开放数据平台中的基础知识库和家谱知识服务平台实时同步以提供智慧化服务^[12]，也需兼顾到家谱本体的设计。家谱本体的作品-实例-单件三层模型为编目数据提供了知识组织结构和数据复用的基础。由于家谱资源的特殊性，家谱本体的三层模型在实例化时采用了一对一对多的模式，谱名、责任者、姓氏、谱籍地、迁徙事件、先祖名人等作为作品的属性，版本相关的信息作为实例的属性，存卷、馆藏机构、索书号等信息作为单件的属性。这样的数据结构对同一种家谱具有多个复本、不同复本具有不同的馆藏信息提供了结构支撑，以此支持已编家谱的数据修改、复本添加、基于数字人文开放数据平台基础知识库的规范控制等专家协同知识生产模式。

2.4 全流程管理

与原有的编目流程相比（如图4所示），家谱联合编目系统的全流程管理除了编目工作

外还涉及采购、捐赠、查重、审校、规范控制、与知识服务平台的数据同步等工作。原先的数据加工流程存在以下问题：（1）在编目流程中，对参与人员的知识要求较高，不同机构之间编目理念、著录规则、习惯、质量要求有所差异，虽然有编目规范文件，但实际操作中也发现编目人员理解深度不够，最终导致编目数据产生质量问题。（2）已生产的编目数据不能即时服务于用户，由于数据是以EXCEL文件形式或者编目数据库形式存在，生产系统没有与服务系统形成即时的信息传递，从而导致服务滞后于编目。

相比而言，全流程管理以编目专家为主要参与者，将数据查重、数据著录、质量控制、数据发布等每一个环节形成工作流。具体来说，是当在数据查重中发现已编书目数据时，从新增数据流程变成更新数据流程。之所以将数据查重放在整个工作流程的第一步，目的是剔除重复数据，以期在新编目数据之前就将重复编目的可能性降至最低，同时促进了已编书目数据的重用，提高编目效率。对于整个编目流程来说，一条编目数据可能由多家机构共同参与与维护更新，著录人员是隶属不同机构的编目专家，审核人员也由各机构的专家选出，提供质量控制。编委会专家同理，主要负责家谱数据的最终审核。因此，家谱联合编目系统作为全

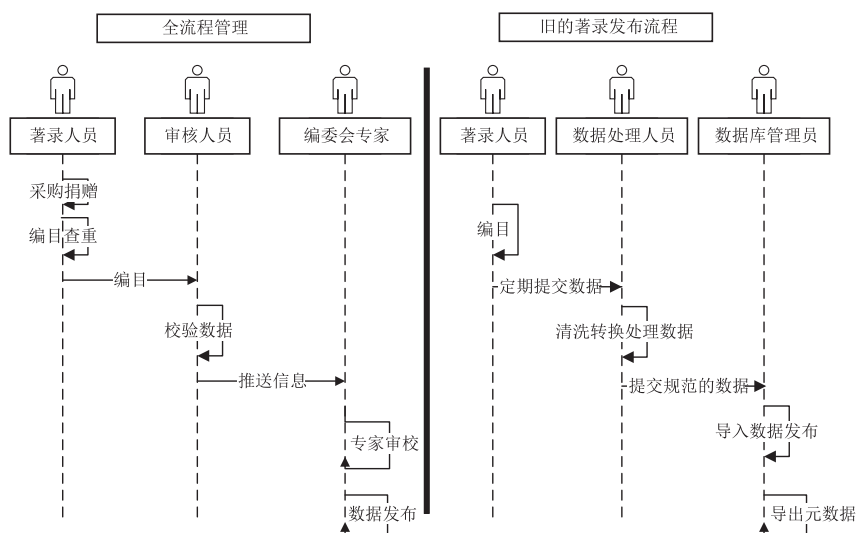


图4 编目时序图

流程管理的协同专家知识生产系统,形成了跨机构、跨地域、跨网域的数据共建共享新模式。

在编目数据质量控制方面,有3个措施:一是培训编目人员,为编目人员提供了编目手册与培训。二是系统层面控制数据质量,根据本体设计对字段的类型、长短、辅以规范词表进行约束。三是加强审核工作,在整个编目数据审核过程中,有两次审核,第一次以编目人员所在机构专家为主,第二次审核引入了AI专家与人类专家共同进行数据审核,AI专家作为人类专家的辅助。

2.5 家谱编目中的规范控制

规范控制一方面是为了提高编目数据质量,确保其准确性与可靠性,是图书馆编目工作中重要的一环,也是全流程知识生产质量控制的重点。规范控制有利于知识生产的规范化与通用化,以数字人文开放数据平台基础知识库作为规范档,利用已有的基于BIBFRAME的家谱本体与数字人文开放数据平台基础知识库人、地、时、事、机构等数据关联,从而实现家谱编目数据与规范数据的语义关联建立^[1],有助于解决不同系统中的数据多源异构问题,也有助于解决不同编目专家由于知识背景、著录习惯、著录经验而造成的数据质量差异化问题。

规范控制有两种方式。一是对部分著录项建立对应的规范词表,如图5所示,系统中的规范词表包含取值词表和数字人文开放数据平台基础知识库中的规范概念或实体。采用基于

取值词表的规范控制的有:责任方式、版本、装订项等,在界面实现中以表单选择器进行筛选。二是基础知识库的引入,采用基于基础知识库的规范控制包括姓氏、谱籍地、朝代、迁徙事件、人名的RDF关联数据。

2.6 专家协同编目审核模块设计

编目的主要参与者是各机构家谱相关的领域专家,专家互相协同编目如图6所示。整个编目行为根据数据是否重复分为著录、复本添加、数据修改。复本添加与数据修改是在其他专家已经著录的前提下,根据掌握的家谱信息,专家对信息进一步完善。当数据出现重复,不同机构的专家将会跨时间、跨地域地对同一家谱信息进行协同编辑。

在AI审核中,基于大模型实现了两方面功能,分别是生成提示词与审核打分。生成提示词是以编目数据的审核要求作为大语言模型的输入,要求其输出审核提示词。其中要求评分是0-100分,以不同角色进行评价,要求大模型注意语言表达、读者体验、结构与组织等。启用大语言模型来生成提示词的思路是在于模型能更好地产出自己能理解的提示词。采用通义千问大模型支持AI专家与AI用户两种角色从不同的角度审核打分,大模型根据提示词扮演了AI专家与AI普通用户的角色,AI专家以风格一致性、语法正确性、词汇丰富度、语言流畅性和逻辑清晰性进行评分。AI普通用户从易读性和整体感受进行评分。人类专家参考两种打分情况给出最终审核结果。

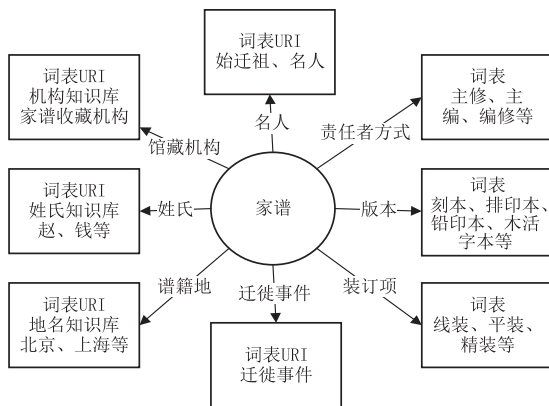


图5 规范控制词表模块

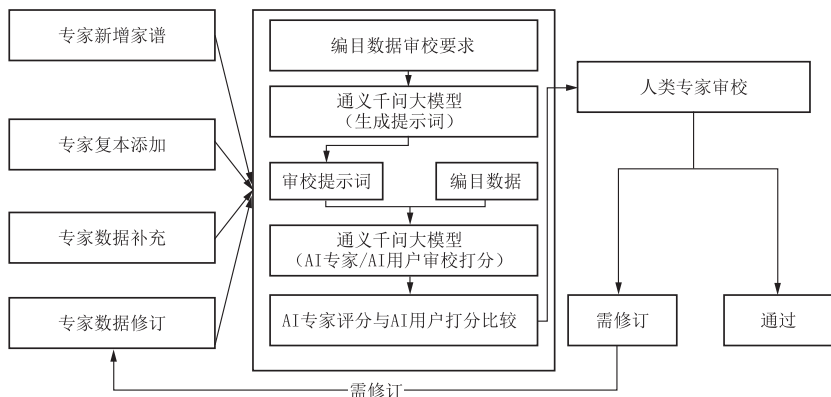


图6 专家协同编目审校模块

3 系统应用情况和效果

3.1 全流程协同编目应用效果

项目历时4年，共计著录了55 000条家谱书目数据，其中经专家协同审核通过的家谱有52 900条，通过率为96.18%。在馆藏机构的查重校验过程中，复本数据共计15 468条。通过引入全流程管理的知识生产模式，项目实现了知识生产与知识服务的实时同步功能。

为了对比协同知识生成的效果，图7是系统在2023年7月至2023年12月的著录数据情况。参与编目的机构有近50家，参与编目著录人员200多人，其中上图编目人员占比12.1%。总编目数量5 000多条，编目工作量是以往上图编目工作量的3倍之多，并且随着参与者的加入，知识生产量还会进一步增速与放大。

为了进一步评价家谱联合编目系统的效果，将其与古籍普查编目系统比较，相对而

言：家谱编目系统基于数字人文开放数据平台基础知识库进行规范控制，是对已有知识生产成果的有效利用，基础知识库基于关联开放数据技术，提供了跨机构、跨地域、跨网域、跨平台的规范控制功能，同时还支持跨机构、跨地域、跨网域、跨平台的专家协同知识生产功能，如不同机构的不同编目人员新增的姓氏、人名、机构将作为新的规范数据，为之后的编目工作重用。家谱联合编目系统实现了家谱的采购、捐赠、编目、审校、服务与规范控制的全流程管理。

3.2 AI专家审校应用效果

AI审校功能目前处于实验阶段。从编目的家谱中，随机选取50条编目数据进行审校打分，如图8所示。其中AI专家的平均分是88.95，AI普通用户的平均分是85.25。AI专家评分高于AI普通用户的原因是AI专家非常重视编目数据提供的信息是否准确无误，以及内

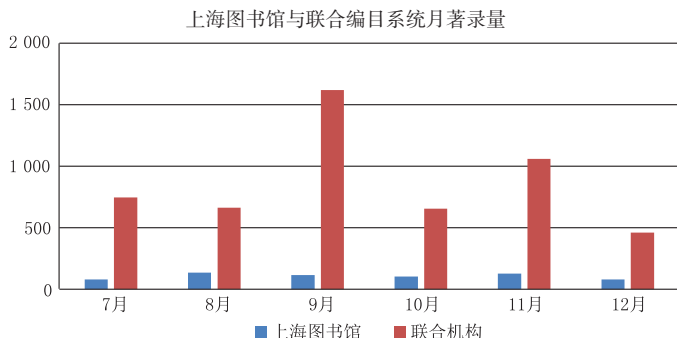


图7 上海图书馆与联合编目系统月著录量对比图

注：本图统计日期为2023年7月至2023年12月。

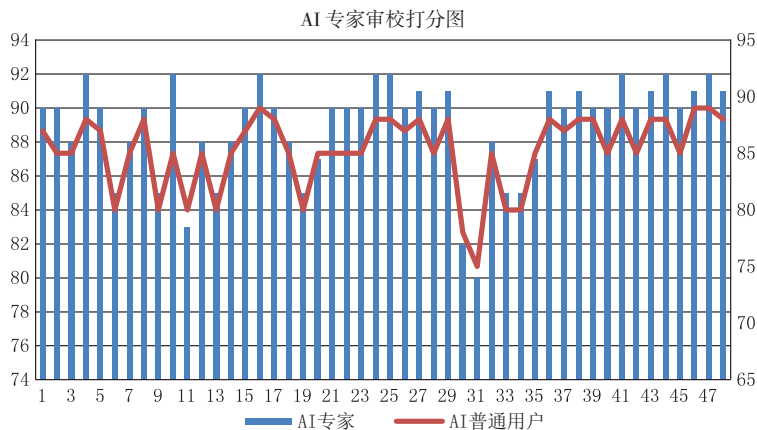


图8 AI 审校打分图

容是否全面覆盖了家族的源流、迁徙、人物、世系等重要历史信息，同时关注家谱在学术和历史研究上的价值。AI普通用户关心内容是否容易阅读，信息是否直观，是否有足够的图表、索引或目录帮助快速查找信息。图中评分有高低，对于分数高的，AI给的主要依据是内容翔实文化价值高。分数低的，AI反馈的是卷数不明确，信息缺失。通过人工校对后，发现AI审校分数较为准确，分数高的编目数据的确比分数低的质量更好。

4 结论与展望

家谱联合编目系统的全流程管理专家协同知识生产模式，是数智时代历史文献资源共建共享模式的一种新的探索。专家协同知识生产的理念是全流程管理加工系统的核心设计思路，此模式与大数据中的任务分发机制（Map-Reduce）有异曲同工之妙。多家机构在同一个编目平台进行联合编目工作，有利于提高知识生产规模和质量，以适应数智时代人文数据基础设施建设和从文献服务向知识服务转型的新需求。具体而言：

（1）对图书馆来说，以传统的编目为基础，引入全流程管理与专家协同知识生产模式，验

证了多家机构协同编目可行性，为同类工作提供可参考的依据。同时在专业数据加工和知识生产中，采取专家为主的协同方式，其效率和质量优于依赖用户贡献知识盈余的众包模式。

（2）全流程管理所带来的效果显著。对编目机构来说，系统所实现的采购、捐赠、跨机构的联合编目与知识服务将整个生态链打通，整体上提高了文献资源管理、编目、服务的效率。

（3）专家协同模式下的规范控制和规范档维护，不同机构在知识生产系统下可以进行统一的规范控制，实时进行数据的校验，在编目过程中有效抑制了机构之间知识生产融合可能会产生的如数据重复、数据错误、数据清洗的问题。通过规范控制知识生产，数据可以实时在家谱知识服务平台上进行服务，以此编目数据在各机构实现了编目成果的共建与共享。

（4）在编目审校中引入了AI审校模块是一种创新，为了规避AI幻觉的问题，采用了AI专家辅助人类专家的方式对编目数据进行质量控制，通过让AI扮演不同的角色来对数据质量进行评价，在知识生产审校环节有效地预估了知识服务的效果。AI专家辅助人类专家的模式也是一种尝试，还需要在未来的工作中进一步检验。

参考文献

- [1] 夏翠娟. 面向人文研究的“数据基础设施”建设——试论图书馆学对数字人文的方法论贡献[J]. 中国图书馆学报, 2020, 46(3): 24-37.
- [2] 欧阳剑, 彭松林, 李臻. 数字人文背景下图书馆

人文数据组织与重构[J]. 图书情报工作, 2019, 63(11): 15.

- [3] 刘圣婴, 王丽华, 刘炜, 等. 数字人文的研究范式

(下转第33页)

- 书情报知识, 2023, 40(6): 68-78.
- [24] 李丹, 杨照坤, 王浩. 新时代古籍工作专业人才保障机制建设的关键问题及建议[J]. 国家图书馆学刊, 2024, 33(1): 44-53.
- [25] 王旭. 用户导向的协同服务模式研究——以北京大学图书馆为例[J]. 河南图书馆学刊, 2024, 44(1): 53-54; 103.
- [26] 娄策群, 卢绍君. 论图书馆系统的协同与竞争[J]. 图书馆学研究, 1986(4): 1-5; 150.
- [27] 王苏卉, 李永卉. 我国古籍保护地方立法可行性及路径分析[J/OL]. 图书馆杂志: 1-12[2024-06-03]. <http://kns.cnki.net/kcms/detail/31.1108.G2.20240511.1901.008.html>.
- [28] 李永卉, 孙悦帆. 建构与边缘: 近代中国古籍保护立法研究[J]. 图书馆杂志, 2023, 42(5): 116-127.
- [29] 唐元玲. 承载、传播与建构——媒体作为社会力量“第三极”在公共文化服务体系建设中的价值[C]// 新时代文化馆: 改革融合创新——2019 中国文化馆年会征文获奖作品集, 2019: 78-81.
- [30] 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.
- [31] 北京市政府采购网. 北京城市图书馆古籍文献馆古籍修复与保护设备购置和智慧空间建设(含进口)中标公告[EB/OL]. [2024-01-04]. http://www.ccbp-beijing.gov.cn/xxgg/sjzfcggg/sjzbggg/t20240104_1559608.html.
- [32] 林通, 郑翔, 李明杰. 古籍的数字赋能与增值利用——“数智时代的古籍活化与利用”高端论坛述评[J]. 图书馆情报知识, 2024(2): 81-86.
- 孙悦帆** 南京农业大学人文与社会发展学院, 博士研究生。研究方向: 数字人文、科学技术史。作者贡献: 论文选题、研究思路、论文撰写、论文修改。E-mail: 354780607@qq.com 江苏南京 210031
- 包平** 南京农业大学人文与社会发展学院, 教授, 博士生导师。研究方向: 数字人文、科学技术史。作者贡献: 思路指导、论文修改。江苏南京 210031

(收稿日期: 2024-04-08 修回日期: 2024-06-04)

(上接第12页)

- 与平台建设[J]. 图书馆情报知识, 2022, 39(1): 6-29.
- [4] 陈涛, 刘炜, 单蓉蓉, 等. 知识图谱在数字人文中的应用研究[J]. 中国图书馆学报, 2019, 45(6): 34-49.
- [5] Haslhofer B, Isaac A, Simon R. Knowledge graphs in the libraries and digital humanities domain[J]. arXiv preprint arXiv: 1803.03198, 2018.
- [6] 杨丽. 全国图书馆联合编目中心重复数据规避策略研究[J]. 图书馆工作与研究, 2021(S1): 108-111.
- [7] 洪琰, 王沛. 全国古籍普查登记工作实践与思考[J]. 国家图书馆学刊, 2014, 23(5): 12-17.
- [8] Terras M. Crowdsourcing in the digital humanities[J]. A New Companion to Digital Humanities, 2015: 420-438.
- [9] 席运江, 林瑶瑶, 廖晓, 等. 基于网络众包模式的海量历史文献数字化处理方法研究[J]. 现代情报, 2019, 39(2): 161-168.
- [10] 夏翠娟, 贺晨芝, 刘倩倩, 等. 数字人文环境下历史文献资源共建共享模式新探[J]. 图书与情报, 2021, 41(1): 53-61.
- [11] 张慧, 佟彤, 叶鹰. AI 2.0 时代智慧图书馆的 GPT 技术驱动创新[J]. 图书馆杂志, 2023(5): 4-8.
- [12] 刘倩倩, 刘圣婴, 刘炜. 图书情报领域大模型的应用模式和数据治理[J]. 图书馆杂志, 2023(12): 22-35.
- [13] 张智雄, 曾建勋, 夏翠娟, 等. 回应 AIGC 的信息资源管理学人思考[J]. 农业图书情报学报, 2023, 35(1): 4-25.
- [14] 蔡丹丹, 宋歌笙, 刘炜. 以 AIGC 创新图书馆知识服务[J]. 图书馆杂志, 2023(12): 36-44.
- [15] 赵瑞雪, 黄永文, 马玮璐, 等. ChatGPT 对图书馆智能知识服务的启示与思考[J]. 农业图书情报学报, 2023, 35(1): 29-38.
- [16] 顾燕. 《1949 年以来中国国家谱总目》著录规则的特点与编纂意义[J]. 图书馆理论与实践, 2022, 258(4): 126-130.
- 朱武信** 上海图书馆(上海科学技术情报研究所), 工程师。研究方向: 数字人文。作者贡献: 数据调研、内容撰写。E-mail: wxzhu@libnet.sh.cn 上海 200031
- 夏翠娟** 上海图书馆(上海科学技术情报研究所), 研究员。研究方向: 知识组织、数字人文、数字记忆。作者贡献: 论文选题和框架指导、文章整体修改定稿。上海 200031
- 吴建伟** 上海图书馆(上海科学技术情报研究所), 研究馆员。研究方向: 历史文献学。作者贡献: 参与文中著录规范制定、专业知识指导。上海 200031

(收稿日期: 2024-04-29 修回日期: 2024-07-01)