DOI:10.13530/j.cnki.jlis.170025

面向循证实践的中文古籍数据模型研究与设计*

夏翠娟　林海青　刘　炜

摘　要　在数字人文逐步成为数字图书馆建设新常态的大背景下,本文通过借鉴“循证实践”和“循证社会学”的思想,提出了“古籍循证”的概念。利用文献调研、需求分析、数据建模、实验验证等方法,调研古代目录、现代联合目录的编排体例和古籍元数据标准规范的结构框架,分析在互联网和机器智能时代,基于古籍循证的版本学、校勘学、分类学及历史人文学等特定领域的研究需求,设计一个可将不同来源、不同格式的古籍目录、元数据记录、古籍文献全文和各类古籍知识融合为一体的古籍数据模型。依托“中文古籍联合目录及循证平台”的建设,利用此模型和本体词表融合 14 种典型的古籍目录和古籍数据库中的数据,实现古籍的不同版本、分类和提要的聚类与比较、古籍著者和其他责任者及其相关关系的统计分析等初步的古籍循证功能,以验证该模型的可行性、开放性和可扩展性,并进一步提出需要解决的问题,探讨可能的解决方案。图 6。表 5。参考文献 18。

关键词　数字人文　古籍循证　数据建模

分类号　G254

Designing A Data Model of Chinese Ancient Books for Evidence-based Practice

XIA Cuijuan, LIN Haiqing & LIU Wei

ABSTRACT

Ancient book catalogs and ancient literature are important sources and evidence material for many Humanities and Social Science research. Traditional research related to ancient books usually relies on experts' expertise or subjective judgment. The emerging Digital Humanities can help scholars to gather relevant information as completely as possible. It can help to raise research questions from bigger spatio-temporal scenes and conduct intensive research across a variety of subjects with unprecedented perspectives. This requires developing a digital humanity platform with relatively complete data and more applicable advanced technologies. A data model that can integrate different formats of different kinds of ancient book catalog data is the basis of this platform.

In this paper, we are proposing a data model of Chinese ancient books using cutting-edge ontology and linked data technology to support researchers to accomplish a so called “evidence-based practice”. The data model is based on the knowledge of classical bibliography combining with philology, bibliology, and so on. This research also intends to explore the new methods of the use of the ancient catalog and documentation to support researches in various disciplines, such as historical research, linguistics, sociology, literature, culture and arts. Web ontology and linked data are the latest achievements of the semantic technologies. They are the most suitable and applicable technologies for developing “authority control” and “evidence-

* 本刊“青年学术论坛”特约稿(Special contribution for the Youth Academic Forum sponsored by this Journal)

通信作者:夏翠娟,Email:cjxia@libnet.sh.cn,ORCID:0000-0002-1859-6979(Correspondence should be addressed to XIA Cuijuan, Email: cjxia@libnet.sh.cn, ORCID:0000-0002-1859-6979)

based" applications. It has the advantages of flexibility and scalability that the traditional relational database does not have. It is very important especially in the distributed environment of massive semi-or non-structured data applications. The advantage of having such data model can directly deal with semantic data (machine understandable), but also support knowledge-based queries with reasoning function.

The data model takes into account of the design method and the aspects of the data model, including the bibliographic framework, creators and contributors, classifications, seals, taboo term and so on. The bibliographic framework consists of 3 + 2 model which stands for "Work- Instance-Item"+"Annotation"+"Classification" based on the needs of evidence-based research of Chinese ancient books, with the reference of the four-tier model of FRBR's "WEMI" and three-tier model of LOC's BIBFRAME2.0. It can adapt flexibly to any kinds of ancient book catalogs and metadata schema based on MARC or DCAP; it also can integrate the full texts of ancient literature. It has an appropriate ability to represent the classification and its multiple comments of different time periods in the records of ancient books. For the description of creators and contributors, the BIBFRAME "Contribution" model is used to clarify the relationship between the responsibility and the document, the relationship between the principle responsibility and the shared responsibility. The knowledge of ancient books is structured into fine-grained semantic units in order to facilitate the machine processing.

Using this model and the vocabularies to integrate data from 14 titles of typical ancient book catalogs, including historical catalogs, official catalogs, private catalogs, large modern joint catalogs and Shanghai Library's ancient book database, the platform realizes key functions for evidence-based research of ancient books. The functions include the search of the versions and classification of ancient books, the clustering and comparing of different versions or copies of an ancient book, the relations of authors and contributors, and the statistical analysis of ancient books with a given time period, area and topic. By the construction of "Chinese Ancient Book Union Catalog Platform for Evidence-based Research", the availability, flexibility and scalability of the data model has been verified. The paper also puts forward the problems that need to be further resolved, such as the identification of a "Work", the establishment of the relationships between the "Instances", the extraction of structured and fine-grained data from the content of ancient books, and so on. 6 figs. 5tabs.18 refs.

KEY WORDS

Digital humanities.　Evidence-based practice on ancient books.　Data modeling.

1　引言

1.1　什么是古籍循证?

循证实践(Evidence-Based Practice),也叫"循证学",本意是"基于证据的实践",源于"循证医学(Evidence-based Medicine)",意为"遵循证据的医学",又称"实证医学",循证医学与传统医学的不同之处在于它强调医疗决策应建立在最佳科学研究证据的基础上[1],传统医学则以经验为主,而不是主要依靠实验性的临床案例、临床资料和疾病的基础知识为依据来诊治病人。随着"循证实践"向人文社会科学领域的延伸,逐渐形成了循证教育学、循证管理学、循证经济学、循证犯罪学、循证软件工程、循证图书情报学等新兴学科领域。以古籍目录的记载和古籍文献中的内容为依据,将科学的研究方法与研究人员的经验结合起来,解决特定研究

问题的过程,可称之为“古籍循证”。

古籍循证的核心是构建古籍研究的证据链,即各种事实性证据及其关联关系,可分为以下几种类型。

(1)物理证据:有关古籍的各种物理性特征,如古籍的装订、尺寸、行款版式、纸张材料等。

(2)内容证据:内容本身和各种基于内容特征的描述,如书名、分类、序跋、避讳等。

(3)历史证据:与古籍相关的各种历史记录,如历代古籍书目的记载、版本变化、收藏历史等。

(4)关联证据:与古籍相关的各种关联关系,包括古籍作品和人物、机构的关系,如作者、批校序跋者、刻印者、收藏者等;也包括古籍版本之间的相关关系,如重刻、增刻、注疏等。

传统的古籍循证过程是专家学者通过收集、遴选和比对来构建这些证据间的逻辑关系。随着数字技术在古籍研究领域中的应用,古籍循证过程越来越多地依靠自动化的数据挖掘和推理来揭示各种证据间的逻辑关系,并将这些逻辑关系形式化地表现出来,这就是古籍研究领域的数字人文研究方法。

1.2 古籍循证的意义

古籍是历史的见证,具有历史文物性、学术资料性和艺术代表性,是研究古代文化、人文、历史、社会等的门径,不仅存在于当前各大图书馆、档案馆、博物馆等公共文化机构里和私人收藏者手中,还被记录在目录学典籍中。虽然大量已经散佚的古籍难再见真颜,但前人在文献中留下有关它们的辑录、提要和评述,不仅可以从中窥见原书的踪迹源流,还能与当前各级各类图书馆、博物馆、档案馆中存续的古籍相互印证,并可反映出大量善本、珍本、孤本之间的丰富联系。

当前已进入大数据时代,包括人文研究在内的各领域逐渐重视以数据为驱动的“第四范式”,除了古籍文献本身,古籍目录学成果也是研究古代学术传承的重要数据来源,可为研究者提供文献证据,是“辨章学术、考镜源流”的主要依据。不论是古代目录学典籍,还是新中国成立后编纂的大型古籍联合目录,均以印刷品的形式存在,无法利用计算机进行智能化的处理,也难以适应当前的互联网环境。而且目前各机构提供的中文古籍数据库,一方面只是以各自馆藏资源的揭示和查找为目的,没有和其他机构建立联合目录,提供一站式检索;另一方面,以记录现存的古籍为主,少有记录曾经存在过而如今已散佚的古籍;再者,这些古籍数据库主要是基于一定的古籍元数据提供查询阅览服务,对于古籍研究中相关的编纂、批校、提拔、藏印、刻工、避讳字等信息虽有所著录,但却缺少规范控制和结构化,不利于机器处理,无法实现大规模数据分析、个性化推荐、深度揭示和智能推理。

图书馆正在经历从提供文献服务到提供知识服务的转型,以互联网和人工智能时代的知识组织为手段建立具有古籍循证功能的联合目录平台,能帮助研究者进行大规模、长时间范围内的古籍资料收集、内容结构化、数据建模、多源数据融合,以及各类古籍知识如版本、刻工、避讳字、藏印、版式等知识库构建并提供知识服务,同时提供各种研究工具以支持大规模数字化古籍文献资料的遴选、聚类、统计、分析、数据挖掘、知识推理等。

1.3 古籍数据模型的作用

要实现“古籍循证”的目的,需要构建一个系统性的古籍数据模型,它不仅能容纳现存古籍的书目数据和馆藏数据,还能容纳已经散佚的古籍书目数据;不仅能揭示古籍作品、版本、馆藏的各项文献特征,还能深入揭示与古籍相关的人、地、时、事等内容特征,以及不同版本、不同责任者之间的关联关系,并将编、著、批校、序跋、藏印、刻工、避讳字等信息转换为结构化的数据;甚至能将古籍书目数据与文本化的古籍全文融合在一起,以便利用数据挖掘、数据可

视化、文本分析等方法和技术手段辅助人文研究。同时它能基于 Web 标准建立古籍著者、刻工、收藏家、目录学家等重要人物的人名规范库,以及古籍版本知识库、藏印知识库、避讳字知识库等,以支持知识关联、发现和推理,实现互联网环境下的知识共享和重用。下文以“中文古籍联合目录及循证平台”的建设实践为依托,试图利用知识本体等新时代的知识组织方法和技术,设计一个灵活可扩展的、易于共享和重用的、面向循证实践的古籍数据模型。

2　古籍目录内容结构及元数据标准规范调研

为设计面向循证实践的古籍数据模型,首先需要对“古籍循证”所依赖的数据来源的内容结构进行深入调研,这些数据来源可归为三种:没有标准化内容结构的“古代目录”,在编排体例上有高度相似性的“现代古籍联合目录”,以及制定了古籍元数据描述和著录规则的中文古籍元数据标准规范。

对于以书籍形式出现的古籍目录,主要研究其编排、体例,分析其共性和特性;对于古籍元数据标准规范,主要研究其元数据元素的结构、关系等。整合古籍目录中的内容结构和古籍数据库中的元数据元素,以便在数据的层面反映古籍作品、版刻、收藏、批点、序跋、校勘等情况,以及古籍作品、版本、责任者之间的关系。

2.1　古代目录

古代目录是古典目录学研究的对象和成果,汪辟疆在《目录学研究》中提出了目录学的“四说”,从内容方面总结了目录的四种类型:目录家之目录、史家之目录、藏书家之目录、读书家之目录[2]。一般来说,古代目录按编纂者和编纂目的可分为史志目录、官修目录、藏书楼目录及私人目录,按照这个分类可以梳理出古代目录编纂体例的发展历史和相应的内容结构变化情况。

本研究的目的是将古籍目录中的内容结构化,提取其中的知识,因而重点研究古籍目录的内容多寡和编排体例(见表 1)。通过对近 40 种史志目录(包括 7 种正史艺文志、经籍志和 32 种补志)、24 种官修目录、44 种私家目录的调研,大致可分为两类。

一是简单的无解题式目录,或称登记式目录,即只记录书名、卷数、著者、类别,《汉书・艺文志》《旧唐书・经籍志》《明史・艺文志》属于此类。

二是解题式目录,即除了记录书名、卷数、著者、类别外,还有关于该书的著者生平、内容提要、版本源流、收藏历史和相关掌故,所谓“提要”“书志”“经眼录”“知见录”“解题”等可归入此类。但解题式目录并没有统一遵循的体例,姚名达按照解题的详略和侧重点的不同,将解题式目录分门别类:一一详论的《崇文总目》《四库全书总目》,在类目之下还有小序,来新夏认为其主要目的是“辨章学术”,对某一部类图书的学术流派、演变和特点加以论述,因而体现了一本古籍目录的学术价值[3];注重版本的《读书敏求记》《百宋一廛书录》《艺风藏书记》;专述书籍内容的《郑堂读书记》;而《郡斋读书志》《直斋书录解题》则较为自由,不是对每种书都专于同一个方面。后世目录学家一般将叙录体、传录体和辑录体作为解题式目录的三种类型[4]。刘向的《别录》开叙录体目录之先河,此后直到清代一直占据主流;叙录体目录的条目内容会涉及校雠原委、著者生平、著书背景、辨书真伪、评论思想、叙述源流、价值判定,等等。传录体目录始自南朝王俭《七志》,在书名之下有著者的传记,如生平事迹、学术流派和终身抱负等,便于帮助读者理解书的内容。辑录体目录以元马端临的《文献通考・经籍考》为代表,主要是抄录书的原文(节选)或序跋、题记原文,并汇编在一起,客观起到了辑佚的作用[5]。

表 1 古代目录的内容结构分析

题名	分类	小序	题名	著者	其他责任者	解题	辑录
汉书·艺文志	有	有	有	有			
隋书·经籍志	有	有	有	有	有		
旧唐书·经籍志	有		有	有	有		
新唐书·艺文志	有		有	有	有		
宋史·艺文志	有		有	有	有		
明史·艺文志	有		有	有	有		
清史稿·艺文志	有		有	有	有		
崇文总目	有	有	有	有	有	有	
文献通考·经籍考	有		有	有	有	有	有
四库全书总目	有	有	有	有	有	有	
郡斋读书志	有	有	有	有	有	有	
直斋书录解题	有		有	有	有	有	
读书敏求记	有		有	有	有	有	
郘亭知见传本书目	有		有	有	有	有	
皕宋楼藏书志	有		有	有	有	有	有
铁琴铜剑楼藏书目录	有		有	有	有	有	
士礼居藏书题跋记	有		有	有	有	有	有
百宋一廛书录	有		有	有	有	有	
贩书偶记	有		有	有	有	有	

2.2 现代古籍联合目录

新中国成立后,集全国古籍专家之力、囊括多家馆藏机构的馆藏目录修纂而成的《中国丛书综录》《中国地方志联合目录》《中国古籍善本书目》《中国古籍总目》《中国家谱总目》等大型古籍联合目录,综合了各省(自治区)、直辖市的公共图书馆、博物馆、档案馆、文物保管委员会、大专院校和中等学校图书馆、科学院系统图书馆,甚至包括名人纪念馆和寺庙藏书单位等众多机构的馆藏古籍目录,将同一古籍的不同版本聚集在一起,列出其收藏地点,是反映中国古籍流传与收藏状况的联合目录(见表 2)。

《中国丛书综录》是一部关于古籍丛书的联合目录,详细列出丛书所含子目及分类,包括来自 41 家单位的馆藏,所收录的丛书都经过必要的考订与整理,所录之书都存于世,其中包括不同版本,对据旧板编印的丛书加以注明,同时亦注明了丛书的藏所[6]。《中国古籍善本书目》的编纂按照事先拟定的“著录条例”,将来自众多单位的卡片目录统一规格,查明著者的时代,同一古籍有多馆报送卡片的,经审查如属同一版本的则合并为一条款目著录。在审校过程中遇到问题时,必通过实地考察、函调书影等方式查阅原书解决问题[7],其中记载的古籍都有馆藏信息,是仍存于世的古籍目录。《中国古籍总目》是在数量和范围上远远超出以往任何一部古籍目录的大型联合目录,收录来自 1 182 家收藏机构的馆藏目录[8],还有近 1/3 的古籍目录没

表 2　当代主要古籍联合目录

题名	出版时间	藏书机构数量	收录范围	收录数量
中国丛书综录	1959—1962 初版，中华书局；1982 再版，上海古籍出版社	41 家	唐宋至民国历代出版的丛书	丛书 2 797 部，子目 7 万多种
中国地方志联合目录	1985 年，中华书局	190 家	自南朝宋至 1949 年的方志	8 200 余种
中国古籍善本书目	1989 年 10 月至 1998 年 3 月，上海古籍出版社	781 家	清乾隆前以及至辛亥革命前有特殊价值的刻本、抄本、稿本、校本	约 6 万多种
中国古籍总目	2009 年，中华书局、上海古籍出版社	1 182 家	至民国初撰著并经抄写、刻印、排印、影印的历代汉文书籍	约 20 万种
中国家谱总目	2009 年，上海古籍出版社	597 家 608 个姓氏	至 2003 年各机构收藏的家谱	52 401 种

有注明馆藏单位。

这些古籍联合目录大都采用“经、史、子、集、丛”五部分类法，简明扼要地梳理了某一古籍作品的不同版本，同一版刻又因批点、校勘、题跋的不同分为不同的条目，每一条目下又列出收藏情况。表 3 是《中国古籍善本书目》中集部（卷二十一）楚辞类的《楚辞章句十七卷》的部分不同版本及其收藏情况，颇有代表性。

《中国家谱总目》《中国地方志联合目录》这样大型的专科目录也收录了大量的古籍书目，但其体例与前三者稍有不同。《中国家谱总目》作为一部提要式目录，除了著录家谱的谱名、责任者、版本年代、册数、收藏者等信息，还撰写了该家族的始祖、始迁祖、迁徙路线、历代名人等内容，以及该谱的特色、谱中有价值的资料等。而《中国地方志联合目录》的著录项有书名、卷数、纂修者、版本、藏书单位和备注，还在备注项中说明了存佚情况、卷数分合、记事起讫、地名古今变迁、书名异称、内容详略、流落异域孤本的国别与收藏单位等。专科古籍目录虽与综合古籍目录有交叉重合，但其侧重、详略各有不同，如能将其中的条目一一对应后融合在一起，便可互为补充。且家谱与方志也是与正史并重的研究史料，可为循证研究提供更为全面的证据支撑。

2.3　古籍元数据标准规范

古籍元数据是关于古籍文献的描述性数据，是面向循证实践的古籍数据模型设计的基础和参考，在 20 世纪后半叶，大多以 MARC 格式存在于图书馆的自动化系统中。随着 20 世纪末 DC 元数据的盛行和数字图书馆建设的兴起，古籍元数据标准规范被作为数字图书馆元数据标准规范的一部分，得到了高度的重视，以科技部基础性工作专项资金重大项目“我国数字图书馆标准规范建设”子项目“专门数字对象描述元数据规范”最为突出，并推出了国家数字图书馆工程标准规范系列成果之一《国家图书馆古籍

表 3 《中国古籍善本书目》的内容结构——以《楚辞章句十七卷》部分内容为例

分类	题名	著者	版本	其他责任者	馆藏机构代码
集部(卷二十一)楚辞类	楚辞章句十七卷	汉王逸撰	明正德十三年黄省曾、高第刻本		0101, 0144, 0201, 0341, 0841, 1501, 1913, 2401, 2702,2901
		汉王逸撰	明正德十三年黄省曾、高第刻本	清袁廷梼校并跋	0101
集部(卷二十一)楚辞类	楚辞章句十七卷	汉王逸撰	明隆庆五年豫章夫容馆刻本		0101, 0141, 0144, 0163, 0201, 0241, 0301, 0502x, 0841, 1601, 1901, 2302x, 2401,2441,2702
		汉王逸撰	明隆庆五年豫章夫容馆刻本	清傅承霖跋	0101
		汉王逸撰	明隆庆五年豫章夫容馆刻本〔卷五至六配清同治十一年蒋曰豫抄本〕	清蒋曰豫跋	1601
		汉王逸撰	明隆庆五年豫章夫容馆刻、天启三年丛桂堂重修本		0182,1901
		汉王逸撰	明万历十四年冯绍祖观妙斋刻本	清彭孙遹批校并跋	1705

元数据规范与著录规则》和文化行业标准(WH/T66-2014)《古籍元数据规范》[9]。一些研究能力较为突出的高校图书馆如北京大学图书馆也制定了适合本馆的古籍元数据规范[10]。

基于 DC 的古籍元数据标准规范充分借鉴了 DC 的元数据方案设计方法和思想,摒弃了 MARC 的繁琐和封闭,强调简单易用、可扩展、互操作和对互联网的适应性,利用元数据元素核心集加扩展集的方式将通用性与专指性统一起来,通过对元数据元素进行"是否必备""是否可重复"和元素取值的约束,使得著录过程得到进一步的规范。这不仅是古籍描述的一大进步,而且极大地改善了古籍元数据记录的规范性,更方便机器读取、处理、转换和传输。但这种以古籍文献为核心描述对象的元数据规范,其主要目的是对文献特征的揭示,虽然也有提要、时空范围、版本类别、收藏历史(包括收藏沿革、题跋印记、获得方式、购买价格)等元数据元素,甚至有记载过该古籍的重要古代目录的元数据元素,但就单个元素的取值来看,是非结构化的文本,粒度还不够细致。另外,对同一古籍不同版本的关联关系的定义和规范,对著者,批、校、序、跋者、收藏家,对版本、分类、藏印、避讳字等知识的进一步描述和规范则不是元数据的重

点,而这些对于古籍循证研究有着重要价值。

3　古籍循证需求分析

汪辟疆在《目录学研究》中提出了古典目录学"四说":纲纪群籍簿属甲乙之学,辨章学术、剖析源流之学,鉴别旧刊、雠校异同之学,提要钩玄、治学涉径之学,总结了古典目录学的治学方法——文献分类、剖析源流、校勘异同、治学门径,也反映了作为古典目录学研究对象的古籍本身和作为古典目录学研究成果的古籍目录对于研究人员的文献价值[2]。被誉为"用现代学术的理念和方法对古代目录学进行开创性研究"的梁启超总结了目录的四大功用,大致可分为:"周知古今著作之大凡";对"散亡代谢"之书籍,可"识其名数、见其学风";通过"博稽诸家著录",可了解书籍的"流传有序",依据书籍的"展转储藏之所在"来"按图索骥";通过"区分门类"来"博观互校""知类通方"[11]。梁氏的思想,与 IFLA 于 1998 年在"书目记录的功能需求(FRBR)"中总结的书目记录要能满足"查找、识别、选择和获取"四项任务是一致的,除此之外,还体现了古籍目录的"辨章学术、考证源流"之于学术研究的重要性。

到了互联网时代,由于语义万维网技术、大数据技术、社会化网络技术、人工智能技术的加持,古籍目录有望焕发新的光彩,其功用将超越传统目录学甚至图书馆学领域而扩展到人文历史社会研究领域,进而成为人文研究数据基础设施的一部分。

3.1　书目控制

"书目控制"一直是图书馆的使命,OCLC 等联合编目机构的存在使得全世界范围内的联合编目已成为书目控制的常态。书目控制也是目录学研究的重要组成部分,中文古籍的书目控制可追溯到五世纪阮孝绪的"穷天下之遗书",及十二世纪郑樵的"通录古今之有无",可见古代的目录学先贤已经有了历史性的视角和"天下"的眼界,到了梁启超的"周知古今著作之大凡""展转储藏之所在",更是具备了现代"书目控制"的思想。中文古籍的书目控制不仅要通录现存古籍,还要为那些散佚的古籍编号注册,使之有目可查。

当前,单个机构的古籍收藏单位一般已完成馆藏古籍编目,并以元数据为基础建设古籍数据库,由此实现本机构的书目控制和书目服务,而全国范围内的中文古籍的书目控制主要是通过编纂大型的古籍联合目录来实现。在互联网时代,纸本目录的阅读和传播受到诸多限制,2008 年美国国会图书馆发布的《书目控制未来报告》草案提出:"书目控制未来将是合作性的、分散的、世界性的和基于万维网的。"[12] 这也同样适用于中文古籍的书目控制,依托互联网进行全网域范围内的中文古籍书目控制逐渐成为共识,由国家古籍保护中心建设的全国古籍普查登记基本数据库就是这一认识的体现。

互联网时代的古籍书目控制要实现的具体功能有:查找——在互联网上提供跨平台跨终端的查询服务,并在大规模数据查询中保证查准率和查全率;识别——每种古籍(无论现存或散佚)都有一个全网域范围内唯一识别的 URI(统一资源标识符);选择——实现同一古籍作品不同版本、同一版本不同馆藏所有者的聚类与区分;获取——实现分布式存储、一站式获取。

3.2　版本学研究

版本学研究的主要内容包括书籍的版本历史——源流、演变、传抄,各种版本的异同优劣,版刻、印刷、装帧各方面的技术。清代著名藏书家黄丕烈以能灵活运用"审内容、看字体、辨刀法、视纸墨、察版式、验图记、查目录、细比勘"等多种方法鉴别版本而著称于世,体现了版本学研究的方法和依据。传统的版本学研究一方面需要依赖研究者的个人经验,另一方面需要看到原书进行肉眼观察,极为不便,使得版本学研

究成为越来越小众的研究领域。

本文所提出的“古籍循证”试图从以下几个新的方面来支持版本学研究。

一是以大规模、多来源古籍目录的实际记载为研究依据,专家的个人经验成为系统知识库的一部分而可以共享。

二是提供细粒度、结构化数据,如从提要或元数据记录中提取人物、时间、地点、印刷、纸张、装订、版式、行款、字体、藏印、避讳字、序跋内容等信息并结构化,便于机器自动化处理,其中藏印和避讳字涉及丰富的历史文化知识,可建立藏印知识库和避讳字知识库。

三是提取责任者项或提要中的重要人物(著者和校勘者、序跋者、批点者、写刻者、藏印所有者等其他责任者)、地(成书地、版刻地、印刷地)、时(成书时间、版刻时间、印刷时间)、事(相关的历史事件)等信息,并提供关于人、地、时、事的更多背景知识而不仅仅是名称,以帮助研究人员更好地理解、辨别、区分版本的异同和梳理其源流。

四是聚集同一作品的所有版本并能灵活地按责任者、地点、时间、版式、内容等结构化的信息进行版本的区分、聚类、比较和统计分析。

五是能够将古籍文本化的全文、序跋内容,或整书扫描影像、书影、藏书印、人物肖像等图片与相应的版本关联在一起,随时随地为研究人员提供参考。

3.3 校勘学研究

传统校勘学的研究内容可以用“勘同异、定是非”来概括,其作用有:“正事实”——通过考订不同文献记载的事实或取得史实的旁证;“通文字”——对照不同的版本疏通古文献中由于多次传抄而发生的词意曲折和谬误;“惠后学”——通过校勘改订记录心得帮助后人理解书籍内容。陈垣在《校勘学释例》中提出了四种校勘的方法:对校法——用同书的不同版本对读,遇不同之处就在旁加注;本校法——用本书的前后互证而抉摘其异同,从而了解其中的谬误;他校法——用他书校本书;理校法——根据一书的内在逻辑和机理来察看其不合理之处。

上述基于古籍循证的版本学的五点需求也是校勘学的基本需求,但校勘学更多地依赖文本化的古籍全文,以及如实反映古籍版刻情况和前人校勘成果的全文影像或书影,并要求提供更方便快捷的自动化和智能化工具。如通过版本聚类帮助研究者找到同书的不同版本用于对校;快速地定位到同书的不同页面进行本校;通过人、地、时、事之间的关联帮助研究人员推荐有关联的不同古籍实施他校;通过文本分析和基于大量结构化语义数据的智能推理来帮助研究人员发现古籍内在的逻辑问题,实现理校。除此之外,还需提供方便易用的标注工具支持研究人员在线标注,标注的内容经同行认可后可以融入到整个数据结构体系中,成为原有数据的补充,也将作为古籍循证的新证据。

3.4 分类学研究

分类学是目录学的一个重要分支,古典目录学随着我国最早的一部图书分类目录——《七略》的诞生而产生,到宋代郑樵提出“学之不专者为书之不明也,书之不明者为类例之不分也”,把图书的分类提到了与学术研究同等重要的高度。然而,从《七略》和《汉书·艺文志》的六分法,到《中经簿》和《隋书·经籍志》的四分法、《五部目录》的五分法、《四库全书总目》的四分法,最后到《中国古籍善本书目》《中国古籍总目》的五分法,古籍分类法不仅在大类上经历了一个不断变化的过程,在细类的划分上更是因学术的发展、思想的变化、图书种类的增加而不断变化。在不同的分类法中,采用的分类标准并不一致,甚至细类的名称相同,但意义和范围却不尽相同。

从内容结构来看,古籍的分类除了一个有着上位类的分类词外,目录学家常在每一类目下作一小序来评述某一类图书的学术流派、演变和特点,因而古籍的分类往往可以反映当时

的学术成就和思想源流,是古籍循证的重要证据。

古籍循证为分类学研究提供了新的研究视角和研究方法,也对数据模型提出了新的需求。

第一,如实地记录每种古籍目录或某机构古籍数据库中用到的每一个分类词,以及该词的上位类,方便分类学研究人员了解每种古籍目录或古籍数据库的整体分类体系。

第二,如实地记录每种古籍目录中每一个类目下所作小序的内容,作为反映古籍目录编纂的那个时代学术发展状况的佐证。当以某一分类词为研究对象时,能够根据时间顺序探索该分类词在不同时代不同古籍目录中是如何被定义、被理解、被评述的。

第三,如实地记录古籍目录中每一种古籍被分到哪一大类、哪一小类。当以某一古籍的分类为研究对象时,可以了解该古籍在不同的时代或不同的古籍目录中是被如何分类的。

满足上述需求后,可从不同的维度探索不同古籍分类的原貌,并进一步探索某个时代的学术潮流和思想认知的变迁轨迹。将分类词和词间关系结构化后,便于实现基于大规模分类数据的统计分析,以发现例外和不合理之处,或者提供统计数据,帮助学者对没有分类的古籍进行分类推荐。

3.5　历史人文等特定领域研究

与版本学、校勘学、分类学这些与目录学有密切关系的研究领域不同,历史人文等特定领域的研究,对古籍循证既有通用的需求,也有各自特殊的需求。通用的需求主要在于研究方法和研究工具的支持,例如时空分析法、社会关系分析法、文本分析法等,以及更好地把这些研究方法和研究数据结合起来的数据可视化技术[13]。要在数据的层面支持这些研究方法,对数据模型也提出了新的需求。

时空分析法:需要提取古籍目录或元数据记录的时间和地点信息,并实现规范控制,如中国历史纪年和公元纪年的对照,历史地名与今地名的对照。对于历史纪年,需要标注某一年号纪年的起止时间、所属朝代等。对于地名,需要标注地名的经纬度,古今地名在时间序列上的变化情况。针对此需求,需进一步完善上海图书馆的“中国历史纪年知识库”和“历史地理知识库”[14],并将古籍文献中的时间和地名与该知识库中的时间和地名建立关联[15]。

社会关系分析法:需要详细地描述古籍的著者、批校、序跋、收藏者,并以机器可处理的方式建立人与人之间的关系,需进一步完善上海图书馆的“人名规范库”,将文献与图书馆人名规范库中的人名建立关联。

文本分析法:除了古籍目录数据和元数据外,需要大量的古籍全文文本,包括题记、序跋、正文等,并与古籍的书目数据和扫描图像建立关联。

4　面向循证研究的古籍数据模型研究与设计

面向循证研究的古籍数据模型,不能受限于繁杂多样的目录体例编排的不同,或因目录学家个人领域的专业性而导致的目录侧重点的不同,而是利用基于语义万维网的知识组织方法和技术,将各种类型的古籍目录有机地融合在一个一致性的数据结构框架中,使之结构化、语义化后便于机器读取和处理。

4.1　书目框架的数据结构设计

书目框架的数据结构是书目控制的基础,笔者在借鉴了 FRBR 的“作品—内容表达—载体表现—单件(WEMI)”四层模型和美国国会图书馆 BIBFRAME 的“作品—版本—单件(WII)”三层模型的基础上,提出了中文古籍书目框架的“作品—版本—单件”+“注释”+“分类”的“3+2”模型(见图 1)。

“作品”是一个抽象的概念,由题名和著者(主要责任者)决定,如王士禛著《渔洋山人精华

录》。“版本”为同一古籍作品的不同物理表现形式,一般由版刻时间和版本类型决定,如《渔洋山人精华录》的“康熙三十九年写刻本”,“版本”包括版式、版刻机构、刻工及除著者之外的其他责任者等信息;“单件”是某一版本的不同复本,与收藏者相关,包括索书号、全文扫描图像或全文文本的 URL 等信息,决定了古籍的获取方式;“注释”则是关于作品、版本和单件的各种解释性信息,如提要、书志、研究观点或研究结论,一般来源于某部古籍目录、论文专著、用户的评论等;“分类”是“版本”和“注释”在某一分类体系中的具体类目,这种分类体系往往来源于某一古籍目录或某个机构的古籍数据库。

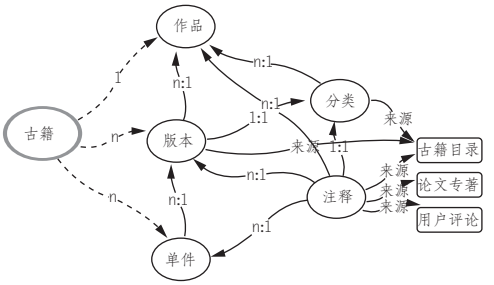

图 1　古籍书目框架抽象模型

“作品”是从“版本”中抽象出来的概念,其目的是为了将同一古籍的不同版本聚集在一起,因而一个“作品”可能对应多种“版本”,一种“版本”可能有多个复本,有不同的收藏者,因而一个“版本”可能对应多个“单件”。而每一个作品、版本、单件则可能分别对应多个注释,每个注释对应一个“分类”术语,每个分类术语来源于某种具体的古籍分类体系(如某一古籍目录)。一个作品同时也与其所有的版本和注释的分类相关联,因而一个作品可能对应多个分类。

“作品—版本—单件”的三层模型能很好地适应当代联合目录的内容结构。而对于古代目录,笔者将其分为三种类型:无解题式、解题式和辑录式,分别用不同的方式处理。不同于目录学家将叙录体、传录体与辑录体相提并论作为解题式目录的不同类型,之所以将辑录式目录剥离单列,是考虑到辑录体目录中的内容一般是来自原书,反映具体某一古籍版本刊印的客观内容,而非后世目录学家融入自己的学识和主观观点后,对原书内容的再加工和对著者、版本的说明、认识和看法[16]。所以辑录式目录中的条目可认为是客观地反映古籍版本的内容特征,而作为版本的内容属性。其他类型的解题式目录中的条目则可认为是主观的,作为与作品相关联的“注释”来处理。图 2 是以《韩诗外传》的“明天启六年刻本”为例构建的古籍书目框架数据模型。

需要强调的是,为了支持版本学的循证研究,“版本”需要高度细粒度化的结构化数据以便于机器处理。可分为内容属性和版式属性,内容属性记录的是某一版本的内容,如摘要、目录、凡例、附录、牌记、序、跋、正文的内容文本,和古籍文献的文本全文一起作为文本分析的数据基础。版式属性用于描述某一版本的外在特征,如装订方式、纸张、每行字数、半页行数、鱼尾、象鼻、版框类型、版框尺寸、天头、地脚、界行等。同一古籍的不同版本往往在内容属性和版式属性上存在一定的差别。

4.2　责任者项的数据结构设计

在基于 MARC 或 DC 的古籍元数据描述与著录规范中,一般将责任者分为著者和其他责任者来著录,在每个责任者的姓名后跟着责任方式。由于古籍的责任方式非常复杂多样,上海图书馆的古籍元数据中,关于责任方式的著录就有超过 100 种,如著、撰、编、纂、修、批点、校对、注疏、序、跋、写、抄、刻,等等。如何在数据的层面,理清每个责任者和与之相对应的责任方式,不同的责任者之间有何关系?责任方式之间存在何种关系?是较难解决的问题。

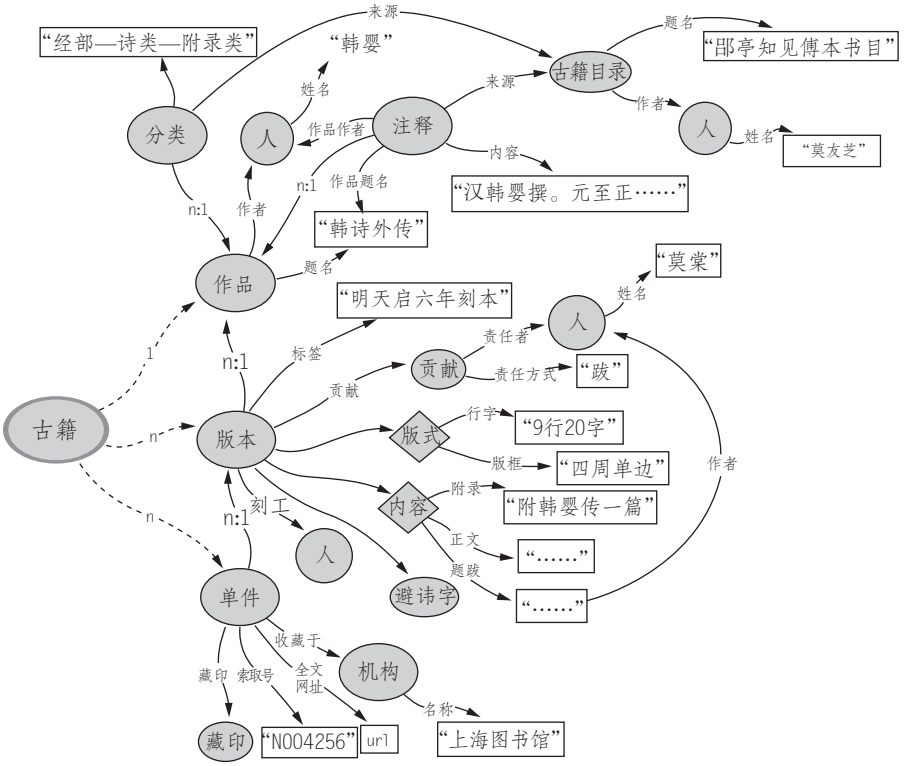

图 2　中文古籍书目框架数据模型——以《韩诗外传》为例

笔者借鉴了 BIBFRAME 的"贡献(Contribution)"这一数据模型来设计解决方案。"贡献"由一一对应的责任者和责任方式构成:责任者是机构或人物实体,而不是字符串形式的姓名,在"中文古籍联合目录及循证平台"中,古籍的责任者都来自于上海图书馆"人名规范库"(names.library.sh.cn),责任方式的取值由一个开放性的受控词表(Scheme)来规范。

人名规范库中的每个人,都被当作一个真正存在过的人物实体来处理。这样做的好处是:一方面可为著者添加丰富的背景知识,除了姓名之外,还有字、号等各种别名,生卒年、职业、任职经历、生平大事、小传(说明其主要著述和学术流派),以及有亲属关系或社会关系的其他人物等,这些都是关于人的结构化的描述数据。这些数据在循证平台中既可被研究人员方便地调阅以作参考,也可被机器处理,用于分面、聚类和个性化推荐。另一方面,可以实现人名的规范控制,将同一著者的不同名称聚集在同一人物实体中,区分同名不同人和同人不同名的情况,进而实现将同一著者有过贡献的所有古籍文献聚集在一起,或按照不同的贡献方式分面呈现。

责任方式的取值来自于一个可更新维护的取值词表,每个取值词被赋予一个全局 URI,既可进一步说明其含义,也可在不同的取值词间建立关系,比如可以用"owl:sameAs"来表示"著"和"撰"之间的等同关系(见图 3)。

4.3　分类的数据结构设计

为了满足上文中提到的三个如实记录分类原貌的需求,在分类的数据结构设计中,将每一部具有分类的古籍目录或每个机构的古籍数据库作为一个分类体系,将其中的每一个分类词作

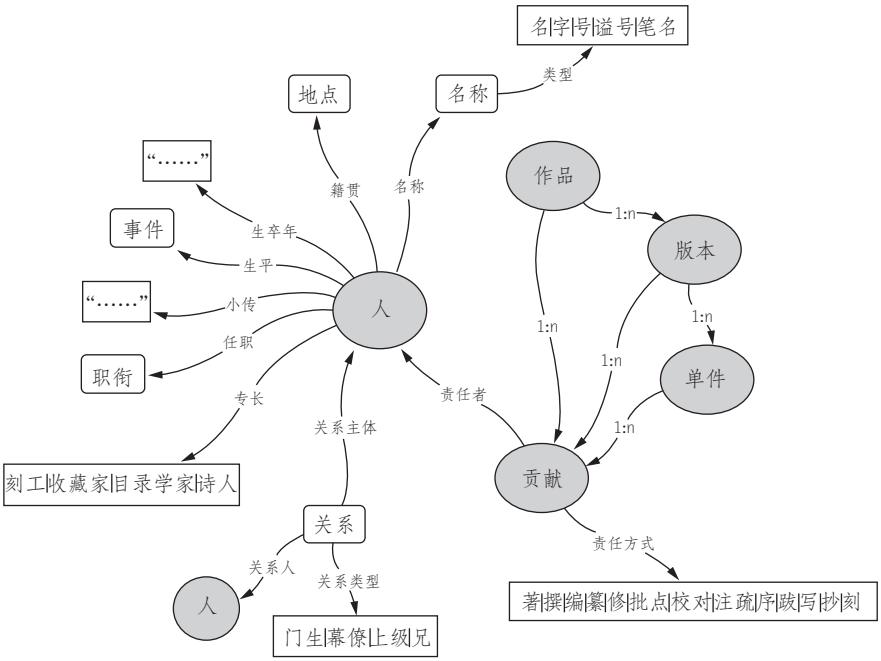

图 3　责任者项的数据模型

为一个对象,并赋予一个全局 URI。分类词由以下属性描述。

(1)分类词标签,即每一个分类词在原分类体系中的名称。

(2)分类小序,某一类图书的学术流派、演变和特点的评述。

(3)上位类分类词的 URI,该级分类词的上一级分类词 URI。

(4)分类词的来源,该分类词所属的分类体系,如某一古籍目录或某机构古籍数据库,以 URI 来表示,如《中国古籍善本总目》的 URI。

(5)分类词的编号,反映分类词层级和顺序的编号,如刘国钧分类法中的编号或杜威十进制分类法的编号,该编号主要用于排序。

在同一分类体系中,若分类词的字面表述相同而上位类不同,则被认为是不同的分类词。在不同的分类体系中,如果分类词的字面表述和上位类相同,也被认为是不同的分类词。不同的分类词以不同的 URI 标识。将分类词作为对象处理后,分类词与作品、版本和注释相互关联,其关联关系如图 1 所示,不同分类体系中的分类词之间也可以建立关系。

表 4 以上图古籍数据库和《中国古籍总目》中的分类词为例来说明每个分类词的数据结构。

表 4　上海图书馆古籍数据库和《中国古籍总目》中分类词的数据结构示例

级别	分类词标签	来源	上位类分类词 URI	URI	小序
第一级	集	上图古籍数据库 URI		URI_1	
第二级	别集	上图古籍数据库 URI	URI_1	URI_2	
第三级	唐五代	上图古籍数据库 URI	URI_2	URI_3	
第一级	集部	《中国古籍总目》URI		URI_4	
第二级	别集类	《中国古籍总目》URI	URI_4	URI_5	
第三级	唐五代之属	《中国古籍总目》URI	URI_5	URI_6	

4.4　藏印、避讳字的数据结构设计

在古籍目录中,藏印和避讳字常以一段说明性文字出现。在古籍元数据著录中,藏印字段的值一般是反映印文内容的简短文字,避讳字也是如此,不足以揭示藏印和避讳字包含的丰富知识,难以实现其循证价值。故在藏印和避讳字的数据结构设计中,藏印和避讳字的相关知识是细粒度化的结构化数据。藏印有印文内容、印文字体、刻印类型(阴刻或阳刻)等,避讳字有避讳对象、避讳方式、避讳字朝代等属性。因其均被作为实体对象处理,可方便地和其他实体对象如作为藏印所有者的藏书家和古籍文献的版本、单件等建立关联,其中藏书家和避讳对象(一般为某位帝王)也和责任者一样来自于上海图书馆"人名规范库",具有丰富的个人背景资料。图 4 是黄丕烈的一枚藏书印的数据结构示例,图 5 是清嘉庆间抄本《翰林记二十卷》中两个避讳字的数据结构示例。

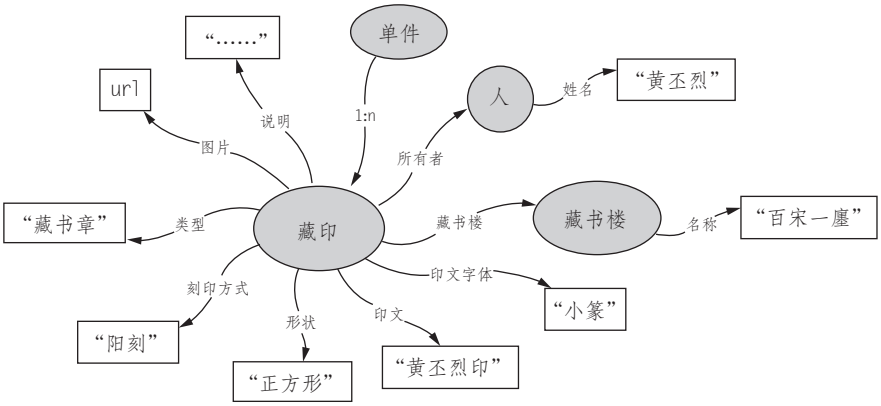

图 4　藏印的数据结构示例

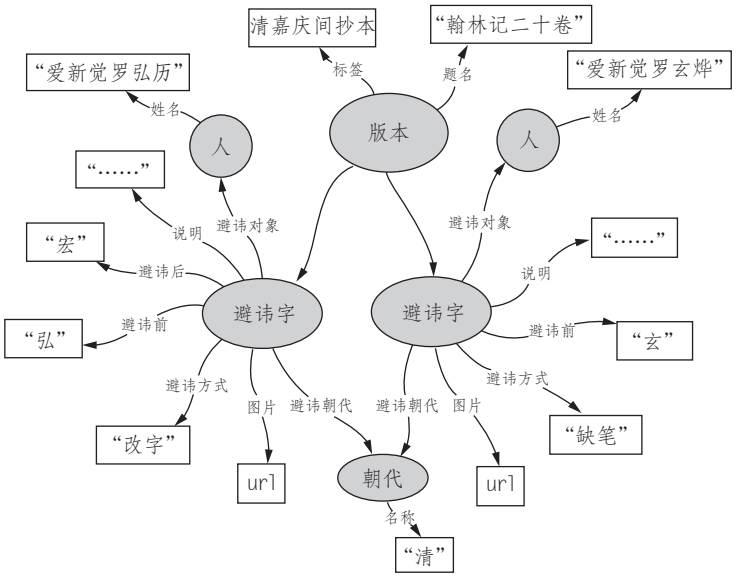

图 5　避讳字的数据结构示例

基于上述数据结构,可将具体的藏印或避讳字和古籍文献相关联,当藏印或避讳字的数量达到一定规模,可形成独立的藏印知识库和避讳字知识库,支持相关领域如历代藏印文化等课题的研究。

5 实施与验证

5.1 实施情况

面向循证的古籍数据模型采用与上海图书馆家谱本体相类似的知识本体方法[17]设计了古籍本体术语词表,目前共有近 40 个类和 160 个属性,其定义和约束以 RDFs 编码后公开发布在网站(http://gj.library.sh.cn/ontology)上。又选取了有代表性的史志目录 7 种、官修目录 2 种、私家目录 1 种、联合目录 2 种、机构藏书志 1 种、机构古籍数据库(即上海图书馆古籍数据库(善本))1 种,外加《明清著名藏书家藏书印》用于建立藏印知识库,通过文本分析、拆分、清洗和结构化处理,将来自 14 种不同数据来源、不同内容结构的古籍书目数据融合在一个数据模型中,并以古籍本体术语词表描述,最后以 RDF 序列化格式编码,以便于跨平台、跨领域的机器读取和处理[18]。

表 5 是 14 种不同来源的古籍目录数据融合前和融合后的数量情况。其中古代目录原书的书目种数和生成的注释种数基本一致,而联合目录原书的目录种数是以作品为单位的,一个作品对应着多个版本,故而得到的版本数量远大于原书的目录种数。柏克莱的善本书志同时

表 5 14 种不同来源的古籍目录数据融合情况

数据融合前			数据融合后		
类别	题名	书目种数	版本数量	注释数量	作品数量
史志目录	汉书·艺文志	614		614	290 005
	隋书·经籍志	3 168		3 168	
	旧唐书·经籍志	2 974		2 974	
	新唐书·艺文志	5 247		5 247	
	宋史·艺文志	9 188		9 188	
	明史·艺文志	3 673		3 673	
	清史稿·艺文志	7 176		7 176	
官修目录	崇文总目	3 389		3 389	
	四库全书总目	10 249		10 249	
私家目录	郘亭知见传本书目	3 681		3 681	
联合目录	中国古籍善本书目	约 6 万	96 728		
	中国古籍总目	约 20 万	407 581		
机构藏书目录	柏克莱加州大学东亚图书馆中文古籍善本书志	802	802	802	
	上海图书馆古籍数据库(善本)	11 125	11 125		

也是馆藏目录,其中每条目录数据生成一个版本(书目数据)、一个注释(书志内容)。上海图书馆古籍数据库(善本)仅是馆藏书目记录,无解题,故一条书目记录生成一个版本。所有的目录数据经过题名和著者查重合并后得到 290 005 种古籍作品。

基于上述数据,构建了“中文古籍联合目录及循证平台”的原型(gj.library.sh.cn),为了实现基于互联网的古籍书目控制,上述 14 种古籍目录中著录的古籍作品及其不同版本、涉及的责任者等被赋予了 HTTP URI 并在 Web 上发布。该原型系统分为三个功能模块:联合目录——作为现存古籍的联合目录,解决查询、选择和获取的问题;古籍目录——按古籍目录原书的内容实现全文浏览和查询;循证研究——为版本学研究、分类学研究以及其他人文历史领域的研究提供证据、方法和工具。这三种功能所依赖的数据是表 5 中 14 种不同来源的书目按照本文所设计的古籍数据模型融合而成的一个整体,但各自的范围有所区别,体现了数据模型的灵活性:联合目录——两种现代联合目录和两种机构藏书目录;古籍目录——除上海图书馆古籍数据库之外已出版的印本古籍目录;循证研究——所有已融合的数据。目前循证研究模块实现了细粒度化的语义检索、多维度分面统计分析,据此可实现以下功能:同一作品的不同版本的聚类和比较以辅助版本学研究,同一作品的不同注释的聚类和比较以辅助作品的源流探析,同一作品的不同分类的聚类和比较以辅助古籍分类的研究,不同责任方式的责任者与文献的关系统计分析、责任者合作关系统计分析以辅助从人出发的人文历史研究。图 6 以《韩诗外传》为例说明如何将不同来源的分类、注释和版本聚集在同一作品下,实现一个作品的版本聚类和源流探析功能。

图 6 《韩诗外传》的分类聚类、版本聚类和源流探析

5.2 结论与问题

通过“中文古籍联合目录及循证平台”原型的建设,验证了本文提出的面向循证研究的古籍数据模型和本体词表基本能满足预期需求。但尚存在以下问题,主要来自于数据处理和实施层面的困难。

一是数据的结构化和语义化。主要由于古籍目录的复杂性和多样性,从古籍目录中提取的数据难以准确地对应到本体词表中相应的类和属性中。

二是版本和注释的区分。被作为注释的古籍目录如《四库全书总目》中也含有大量的版本信息,如何从注释中提取版本信息生成相应的版本实体,也是一个需要继续探索的问题。

三是版本和版本间关系的建立。不同的古籍目录中提取的版本可能存在多种多样的关系,有的可能是同一个版本,有的版本可能是另一个版本的底本,这样的关系在目前的古籍数据模型中没有定义,但可以在现有的模型上扩展。

四是古籍作品的认定。目前由机器自动以“题名+著者”去重合并,虽然能将大部分古籍的作品与其版本和注释聚类,但不够准确,比如在有的古籍目录中,著者和其他责任者没有明显的区分,导致本应属于同一作品的版本没有准确地归并到同一作品下。

五是大量文本全文的支持。上述 14 种数据主要来源于古籍目录及描述性元数据,这些都是重要的研究依据,尤其是对于已经散佚的古籍,是间接的证据。对于现存于世的古籍,古籍全文是循证研究的直接证据,特别是文本格式而非扫描图像的古籍全文,可支持文本分析。在本文所提出的数据模型中,已经考虑到了文本化全文数据的融合,还需要进一步实施,这也是“中文古籍联合目录及循证平台”下一步的建设重点之一。

上述问题的解决,既要依赖古籍全文的文本化、古文字识别、自然语言处理、名称实体识别、自动文白互译等技术的进步,也对数据模型的设计提出了新的需求,要求模型是开放的、可扩展的,而这正是以语义万维网为基础的知识本体方法的长处,本文所提出的数据模型的设计正是在数据加工、清洗、分析的过程中逐渐完善的。例如第二个问题的解决可以采取两种方式,一是扩展古籍数据模型,在不同的版本之间添加关系属性和关系类型取值词表;二是引入众包的思想,在数据实例的层面,允许专家用户将一个具体的版本与某个作品的关联改成另一个作品。

6 总结与展望

随着大量古籍的数字化和现代信息技术的进步,以古籍目录和古籍文献为研究证据的人文、历史、社会科学研究领域将逐渐改变依赖个人经验和个人搜集整理资料的研究模式,而更多地依靠大规模文献资料中蕴含的数据、事实和知识,以及基于这些数据、事实和知识的统计、分析和推理。古籍目录和古籍文献的数据化,将促进研究人员在研究方法和研究模式上的更新和变革。首先,可实现基于互联网的、全球范围内超越时空限制的古籍书目控制。其次,与目录学密切相关的传统研究领域如版本学研究、校勘学研究、分类学研究,以及人文历史等特定领域研究,也可充分利用大数据技术的大规模、分布式存储和计算能力,实现远远超出个人和单个机构能力的大规模、自动化的数据统计、分析、推理和机器学习。

本文提出的“古籍循证”概念和面向循证实践的古籍数据模型的设计,以及“中文古籍联合目录及循证平台”的建设,正是顺应这一趋势的尝试。目前原型系统主要涉及的还是“书目”数据,下一步还将联合目录与古籍原文(主要是扫描图像和古籍文本)勾连,提供更多语义层面的内容分析和素材,结合各类工具提供更丰富的研究“证据”。

古籍目录的编纂和古籍书目数据建设是一项复杂且困难的工作,需要具备对书籍分类的认识,对书籍版本的鉴定,对书籍完整或残缺的

认定,对书籍反映的历史、文化、思想的不断发现和认识。这就要求古籍循证平台是开放的,支持模型的完善、数据的丰富和知识的进化。在数据层面,除了囊括古籍目录和古籍文献的描述性数据外,还需要整合文本全文,利用文本分析技术提供直接的研究证据。在服务层面,对于人文历史等特定领域的古籍循证还需要更多的调研和典型场景模拟,提供有针对性的方法和技术支持,以满足不同研究领域的不同需求,甚至将平台的功能嵌入研究人员的研究活动,保存研究过程中产生的数据,并与已有的数据有机融合,相互完善。随着"中文古籍联合目录及循证平台"的建设,将要容纳更多的古籍目录、古籍数据库和古籍数字化全文,本模型也将进一步得到完善。

参考文献

[1] Sackett D L, Straus S E, Richardson W S, et al. Evidence-based medicine: how to practice and teach EBM[M]. 2nd ed. Edinburgh & New York: Churchill Livingstone, 2000:10-52.

[2] 汪辟疆. 目录学研究[M]. 上海:华东师范大学出版社, 2000:1-13.(Wang Pijiang. Research on bibliography [M].Shanghai: East China Normal University Press, 2000:1-13.)

[3] 来新夏. 古典目录学浅说[M]. 北京:中华书局, 2003:52-57.(Lai Xinxia. On the classical catalogue[M]. Beijing: Zhonghua Book Company, 2003:52-57.)

[4] 王重民. 中国目录学史论丛[M]. 北京:中华书局,1984:80.(Wang Chongmin. On the history of Chinese bibliography[M]. Beijing: Zhonghua Book Company,1984:80.

[5] 李书. 对我国书目提要演变的思索——兼论机读目录提要项的著录[J]. 图书馆建设, 2004(2):48-49.(Li Shu. Thinking about Chinese development of bibliography abstract: discuss the summary item of MARC concurrently[J]. Library Development, 2004(2):48-49.)

[6] 陈秉仁.《中国丛书综录》对中国古籍目录索引的贡献[J]. 中国索引, 2005(3):2-5.(Chen Bingren. The contribution of *China Series Collection* to index of Chinese ancient catalog[J]. Journal of The China Society of Indexers, 2005(3):2-5.)

[7] 冀淑英.《中国古籍善本书目》后记[J]. 北京图书馆馆刊, 1996(2):79-83.(Ji Shuying. The postscript of *The Catalogue of Chinese Rare Ancient Books*[J].Journal of the National Library of China,1996(2):79-83.)

[8] 时永乐.《中国古籍总目》的得与失——以《论衡》为例[J]. 图书馆工作与研究, 2014, 1(11):79-81.(Shi Yongle. The advantage and disadvantage of *The Total Catalogue of Chinese Ancient Books*: an example of *Lun Heng* [J].Library Work and Study, 2014, 1(11):79-81.)

[9] 葛红梅, 徐晶晶.基于 DC 的古籍元数据评述[J]. 兰台世界, 2015(26):33-34.(Ge Hongmei, Xu Jingjing. A review of DC-based ancient books metadata[J].LanTai World, 2015(26):33-34.)

[10] 姚伯岳, 张丽娟, 于义芳,等. 古籍元数据标准的设计及其系统实现[J].大学图书馆学报, 2003, 21(1):17-21.(Yao Boyue,Zhang Lijuan,Yu Yifang,et al. On the design of rare book metadata standard and its system implementation[J].Journal of Academic Libraries, 2003, 21(1):17-21.)

[11] 戴丽琴, 彭树欣, 柯平. 梁启超:中国古代目录学研究现代第一人[J]. 图书馆论坛, 2012, 32(3):190-193.(Dai Liqin, Peng Shuxin, Ke Ping. Liang Qichao: the first classical bibliography scholar in modern times in China[J].Library Tribune, 2012, 32(3):190-193.)

[12] 顾犇. 关于《书目控制未来报告》草案[J]. 国家图书馆学刊, 2008, 17(1):76-78.(Gu Ben. About *Report on the Future of Bibliographic Control* [J]. Journal of the National Library of China, 2008, 17(1):76-78.)

[13] 刘炜,谢蓉,张磊,等. 面向人文研究的国家数据基础设施建设[J]. 中国图书馆学报, 2016, 42(5):29-39.(Liu Wei,Xie Rong,Zhang Lei, et al. Towards a national data infrastructure for digital humanities[J]. Journal of Library Science in China, 2016, 42(5):29-39.)

[14] 夏翠娟. 中国历史地理数据在图书馆数字人文项目中的开放应用研究[J]. 中国图书馆学报, 2017, 43(2):40-53.(Xia Cuijuan. The opening and application of Chinese historical geography data in digital humanities projects of libraries[J]. Journal of Library Science in China, 2017, 43(2):40-53.)

[15] 夏翠娟.以关联开放数据服务为基础的数字人文平台建设方案研究[J].图书馆学与资讯科学,2017(4), 43(1):47-70.(Xia Cuijuan. Building a digital humanities platform by using linked open data services[J]. Journal of Library and Information Science, 2017(4),43(1):47-70.)

[16] 宋光宇. 从《四库总目》和《文献通考·经籍考》看叙录体与辑录体的区别[J]. 南风, 2016(20):13-14.(Song Guangyu. On the differences between "Ji Lu Ti" and "Xu Lu Ti" from *Si Ku Zong Mu* and *Wen Xian Tong Kao· Jing Ji Kao*[J].South Wind Folk Literature of Guizhou, 2016(20):13-14.)

[17] 夏翠娟,刘炜,张磊,等. 基于书目框架(BIBFRAME)的家谱本体设计[J]. 图书馆论坛, 2014(11):5-19.(Xia Cuijuan, Liu Wei, Zhang Lei, et al. A genealogical ontology in the rorm of BIBFRAME model[J]. Library Tribune,2014(11):5-19.)

[18] 夏翠娟,刘炜,陈涛,等. 家谱关联数据服务平台的开发实践[J]. 中国图书馆学报, 2016, 42(3):27-38.(Xia Cuijuan, Liu Wei, Chen Tao, et al. A genealogy data services platform implemented with linked data technologies[J].Journal of Library Science in China, 2016, 42(3):27-38.)

夏翠娟 上海图书馆系统网络中心高级工程师。上海 200031。

林海青 美国加州大学柏克莱分校东亚图书馆技术部主任及中文编目图书馆员。美国 加州 94720。

刘 炜 上海图书馆副馆长,研究员。上海 200031。

(收稿日期:2017-09-26)