



基于关联数据的方志物产史料语义化知识组织研究*

□徐晨飞 包平 张惠敏 姜霖

摘要 旧方志中的物产史料是传承中华农业文明五千年历史的重要载体,数据规模大且史料价值高,在语义网环境下对其进行知识组织,是后续基于方志物产史料深度利用以及知识发现的前提。本研究以机构特藏《方志物产》为研究对象,首先提出基于关联数据的方志物产语义知识组织框架;其次设计并构建方志物产知识本体模型,对其概念、属性及关系进行规范化描述;然后基于关联数据技术,从数据预处理、生成 RDF 数据、实体关联、关联数据存储与发布等步骤探讨方志物产关联数据集的构建与发布过程;最后,以《方志物产》云南卷为例,构建基于关联数据的方志物产知识库,为相关领域用户和社会公众提供各类知识服务。本研究可以激活蕴藏在方志物产史料中的多维度知识,也为其他方志汇编史料语义化知识组织提供参考,同时在一定程度上丰富了我国数字人文研究的应用场景。

关键词 方志物产 语义 关联数据 本体 知识组织 数字人文

分类号 G254

DOI 10.16603/j.issn1002-1027.2020.06.011

1 引言

地方志(方志)是中国古代文献的重要组成部分,是以一定体例,全面系统地记述一定区域各个方面或一个方面历史与现状的资料性文献^[1],具有“资治、存史、教化”之功能。物产是方志中的一个重要类目,简称方志物产^[2],往往被纂修者冠以标明土宜、因民所利、以供后来者借鉴之目的,除记载当地的农作物、蔬菜、果木、动植物等内容外,还涉及土地状况、民生日用、手工业经营等,蕴含着诸多的信息。张舜徽先生曾在《中国文献学》中评价:“方志里面对于赋役、户口、物产、物价等方面的记载,最为可贵”^[3]。

整理并利用旧方志资源始于明末清初学者顾炎武,其征引 1000 多部方志资料汇编而成《天下郡国利病书》。新中国成立以后,党和政府提倡“古为今用”,呼吁开展志书二次文献的编纂工作,如编纂方志目录、方志提要、方志索引以及方志分类资料等^[4],许多机构积极响应并在农业、天文、气象、地震和灾害等诸多领域汇集类编了大量旧志史料。在此

背景下,中国农业遗产研究室在著名农史学家万国鼎先生领导下查抄、整理、汇编国内 7000 余种方志的物产专题性资料,于 1960 年初编成《方志物产》(文中凡加书名号特指该手抄本材料)449 册、《方志综合资料》111 册、《方志分类资料》120 册,共 3600 余万字。抄本《方志物产》为海内孤本,《方志物产》与《中国农业史资料》及续编被称为“红本子”而闻名学界,李约瑟博士等许多国际知名学者曾专程到访查阅。

在计算机信息技术日新月异的高速发展浪潮中,历史研究也迎来了新的机遇与挑战。史料的获取、存储、交流、再生等诸环节均发生了深刻变化^[5]。同时,伴随着数字人文研究的深入与不断发展,人文学者对提供支撑跨学科研究的资源、工具、数据管理与检索等通用解决方案的研究基础设施的需求也迫在眉睫^[6]。目前在中国历史研究领域已有一些较为成熟且提供开放数据集的研究基础设施,如上海图书馆开发的“家谱知识库”“中文古籍联合目录循证平台”等文献知识库以及“人名规范库”“地

* 国家社会科学基金重大项目“方志物产知识库构建及深度利用研究”(编号:18ZDA327)、教育部人文社会科学研究青年基金项目“基于语义的方志物产资料知识组织与知识聚合实证研究”(编号:19YJC870027)的研究成果之一。

通讯作者:包平,ORCID:0000-0002-1131-5169,邮箱:baoping@njau.edu.cn。



理名词表”“中国历史纪年表”等基础知识库^①;哈佛大学与北京大学、台湾“中央研究院”三方合作共建的“中国历代人物传记资料库”(China Biographical Database Project, CBDB)^②;台湾“中央研究院”研制的“中华文明之时空基础架构”(Chinese Civilization in Time and Space, CCTS)^③等。针对网络环境下的地方志资源,华东师范大学图书馆推出的“数字方志集成平台”^④将多所高校的集成系统与“爱如生”“超星”等商业数据库进行了方志元数据的整合,可提供统一的方志资源检索以及关联书目数据的开放服务^[7]。但是,有关方志特定类目下的汇编史料如物产、气象、灾害、天文等,大多仍旧为馆藏纸质文献,属于非结构化的文本状态,数字化乃至语义化的网络开放数据集还较为缺失。

方志数量巨大,而多数机构又“重藏轻用”,方志分类资料的编纂可以充分发挥方志应有的作用与价值,这项艰巨繁重的工作也一直薪火传承至今,最新的成果当属北京师范大学“985工程”重大建设项目《中国地方志分类史料丛刊》(共1663册)的出版^[8]。但是面对这些古今诸多学人前赴后继、呕心沥血之作,我们不禁要思考一个新的研究命题,即如何在大数据及数字人文研究背景下,进一步检索、挖掘与利用这些方志分类资料信息,形成互联网时代可拓展、可关联及可展示的新史料与新史实,将这些珍贵资源重新激活,为各学科领域研究乃至经典传统文化推广发挥更大的作用。基于此,本研究将《方志物产》这份独一无二的方志汇编资料作为研究对象,深入剖析蕴含其中的如物产、来源方志、人物、引书、地名、时间、物产类别等知识元素并展开语义化知识组织,构建方志物产领域知识本体;利用关联数据技术实现与其他外部开放数据集的知识关联,提供领域知识的序化组织以及可视化展现;以《方志物产》云南卷为例,将相关知识进行多维度聚合,构建方志物产知识库系统,以期为领域用户和社会公众提供知识发现、知识考证等一系列知识服务应用,同时也为其他方志类编史料的知识组织与知识管理提供参考。

2 方志物产知识组织研究与关联数据应用研究现状

近20年来,围绕《方志物产》手抄本资料的数字化、知识组织,相关学者开展了多项富有成效的研究

工作。21世纪初,王思明与惠富平团队采用人工录入的方式初步实现了《方志物产》资料的电子化文本文档的生成,该项数字化工作作为后续《方志物产》的智能化处理研究奠定了语料基础。衡中青最早开启了《方志物产》知识组织与内容挖掘的研究,构建了《方志物产·广东卷》信息系统用于检索方志物产文献全文,并生成物产与引书索引^[9]。朱锁玲以《方志物产》广东、福建、台湾三省为研究对象,运用命名实体识别技术,对语料中的地名进行自动识别,并挖掘了物产与地名之间的关系^[10]。李娜以《方志物产》山西分卷为例,构建了一套符合《方志物产》自身特点的物产分类体系,实现类目名称的智能化规范处理^[11],并采用社会网络分析技术对物产名与别名之间的网络关系进行知识关联分析^[12]。综上,学者们着重对《方志物产》文献的体例格式及内容特征进行了深入剖析,且均以物产实体为主线展开了多种知识组织与知识发现研究,不足之处是忽视了《方志物产》中显性或隐性的知识元素以及知识元素之间的语义关系,也未能将这些知识元素与外部知识进行有效关联,使得相关研究分析存在一定的局限性。因此,在相关研究基础上进一步构建规范的方志物产知识组织本体模型尤为必要。

“关联数据”概念最早由蒂姆·伯纳斯·李在2006年提出^[13],旨在推动万维网上资源间的语义关联以及数据开放。关联数据为领域知识描述提供了技术和语义资源,为知识序化与知识聚合提供了保障。近年来,关联数据技术发展迅速,在诸多领域尤其是在图书馆学情报学与数字人文研究中有许多成功应用的案例,如全球众多国家的图书馆相继将数据发布成关联数据^[14],欧洲数字图书馆(Europeana)、威尼斯时光机器项目(Venice Time Machine)、芬兰数字人文关联开放数据基础设施(Linked Open Data Infrastructure for Digital Humanities in Finland, LODI4DH)等著名项目都应用了关联数据技术。在国内,近些年关联数据研究呈现出从理论研究向实践应用发展的趋势^[15],如上海图书馆推出的一系列数字人文项目^⑤均采用关联数

① 网址: <http://data.library.sh.cn/index>.

② 网址: <https://projects.iq.harvard.edu/chinesebdb>.

③ 网址: <http://ccts.ascc.net>.

④ 网址: <http://fangzhi.ecnu.edu.cn/>.

⑤ 网址: <http://data.library.sh.cn/index>.

据作为核心技术。周育彬开发了基于关联数据的数字博物馆语义融合平台,可实现博物馆领域及相关数据资源的关联与整合^[16]。祝帆帆对中国十大传世名画相关的馆藏数据进行关联数据发布,发现将关联数据技术应用在文物数据管理方面是行之有效的^[17]。曾子明提出基于关联数据的视觉资源知识组织方法并应用于敦煌文化遗产项目^[18]。侯西龙运用关联数据技术构建了非物质文化遗产知识关联数据集与知识服务平台^[19]。陈涛将知识图谱应用于数字人文研究的系统框架,并构建了“中国历代人物传记资料库”的关联数据平台^[20]。从上述研究中可以看出,关联数据技术被广泛地应用于图书馆、档案馆和博物馆,以及非物质文化遗产和数字人文研究等领域资源的语义组织与知识聚合等方面,但对方志内容尤其是方志类编集成资料却很少以关联数据的方式进行知识组织,这类资料具有数据量大、专题性强以及史料价值突出等特点,采用关联数据技术不仅可有效组织与管理方志资料这类狭义的领域知识小网络,还可与其他相关多源知识节点链接形成更丰富的广义知识大网络。

3 基于关联数据的方志物产语义知识组织框架构建

万国鼎先生曾撰文指出旧志中载述物产的一些通病:“但载物名,不详其分布轻重;但引录前志及《本草纲目》、《广群芳谱》之属,而不载现状;或更偏于考据,辩证各物,而无一字涉及其在本地之生产情形”^[21],可见其对物产的时空分布与历史演变过程颇为重视;加上万先生又是近代“索引运动”的主将^[22],因此,他在领导编纂《方志物产》过程中,有意地将物产来源志书按省府州县乡的区域位置以及方志编纂年代的先后制成索引,分类编排,以便于他人检阅使用。万先生的精心设计,不仅造就了《方志物产》独特的文献载体形态,还提升了资料内容之间的关联性,同时也为后人在网络时代进一步展开语义化知识组织提供了极佳的基础材料。《方志物产》中包含多类实体,如物产、志书、时间、行政区划,物产描述信息中又包含大量的别名、引书、人物、产地等实体,如何将分散的知识元素进行深度序化组织,同时还需考虑如何与外部其他异构知识源进行关联,最终形成一个富含语义的可扩展、可共享、可聚合的有机整体是目前亟待解决的问题。关联数据作为当前语义网的轻量级实现方式可为实现上述

研究目标提供有效方案,因此本文提出一个基于关联数据的方志物产语义知识组织框架,其核心工作主要包括构建方志物产知识本体对领域资源各类实体以及实体之间关系进行语义化描述,采用资源描述框架(Resource Description Framework, RDF)数据模型及 URI 链接机制实现方志物产知识关联数据的存储与发布。该框架分别由源数据层、本体层、关联数据层以及应用层组成,如图 1 所示。

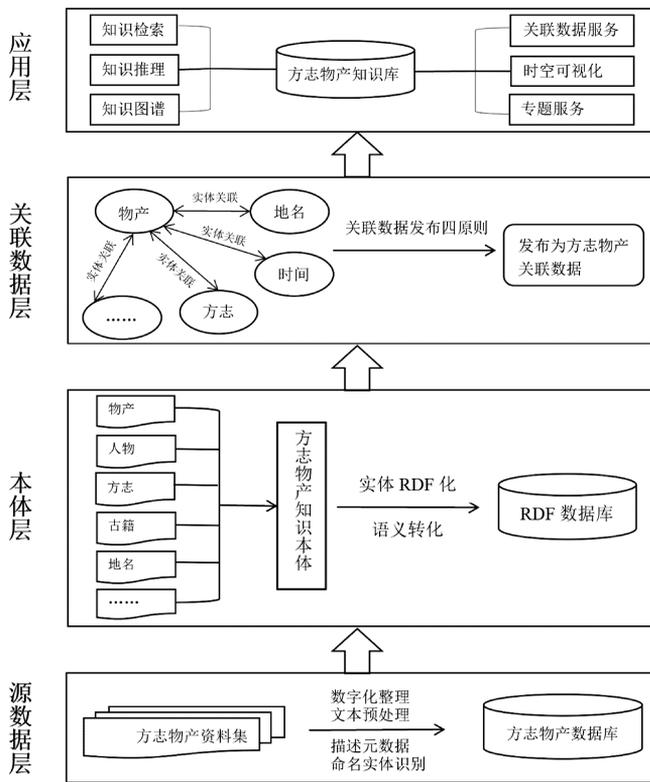


图 1 基于关联数据的方志物产语义知识组织框架

源数据层以《方志物产》数字化文本资料为基础,同时结合领域相关文献资料对其进行再扩充,尽可能将方志物产领域所涉及的多层级的知识吸收并完善,再通过深度学习与人工校对的方式提取物产、方志、时间、地理位置、人物、引书等一系列命名实体数据,形成关系型数据库文件实现文本数据的多层级结构化。在本体层中需设计方志物产知识本体,然后基于知识本体模型对源数据层进行知识实体抽取并转换为 RDF 数据格式,存储于三元组数据库。关联数据层主要任务是实体关联以及关联数据发布。在应用层可基于关联数据与语义技术框架构建方志物产知识库系统,面向特定领域的数字人文研究需求,为相关学科提供知识服务,开展《方志物产》资料深度利用研究。



4 方志物产知识本体模型设计与构建

本体可以看作是一种知识的抽象模型,其经典定义为“共享概念模型的明确的形式化规范说明”^[23],常用来描述领域知识,用于共享、交流和重用。近年来,本体作为重要的知识组织方法论,被学者广泛运用于知识工程、人工智能、数字图书馆等领域。本文在已有本体理论与国内外相关本体模型基础上,按照领域本体设计思路与构建步骤,尝试构建了方志物产知识本体模型,对方志物产领域知识中的概念及关系进行描述与揭示。

《方志物产》作为方志专题汇编资料,具有独特的体例格式,其按行政区划详细记录了各地旧方志中的物产史料,对摘抄的方志也编制了索引,因此《方志物产》、物产记录、来源方志均为核心知识元素;方志具有纂修者以及大量时空信息,《方志物产》也有责任者,故人物信息、地理信息、时间信息也是

重要的知识元素;此外,考虑到该资料未来进一步的延展,如将其他非方志类古籍中的物产信息也进行辑录并增补入内,因而本研究加入了古籍知识元素。这些知识元素之间存在记载关系、地理分布关系、类属关系、编纂关系等多种语义关系。方志物产知识本体模型参考借鉴了 FOAF、GeoNames、上海图书馆等本体词表以及都柏林核心元数据(Dublin Core Metadata,DC),共有六大核心类(Class):《方志物产》(Fangzhiwuchan)、物产(Produce)、方志(Fangzhi)、人物(Person)、地名(Place)、古籍(Ancient Books),每一个类具有相应的数据属性(Data Properties),实体之间的关系通过对象属性(Object Properties)来描述,其模型如图 2 所示。前缀(fzwc)为方志物产知识本体的命名空间,定义为 <http://www.fzwc.online/ontology/>。

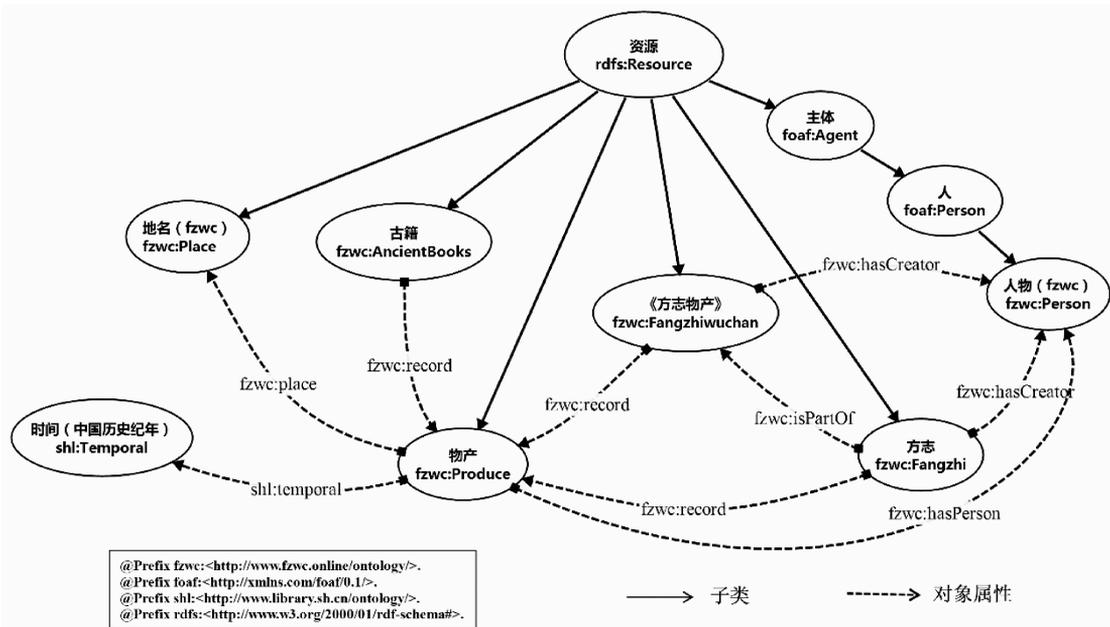


图 2 方志物产知识本体模型

物产是领域知识中最为重要的实体,物产类是根据《方志物产》所引方志中物产知识书写模式来制定的。《方志物产》物产知识书写模式大致有四类行文格式:(1)先叙述物产名,后对物产作解释与说明;(2)只有物产名,没有描述性文字;(3)叙述物产时,描述性文字或有或无;(4)物产出现在一段叙述性文字中,需经过人工阅读后提取物产,示例见表 1。物产描述信息中有大量的物产别名、产地、人物、引书等实体信息,对于物产的知识发现有重要的作用,因

此可作为属性归入物产类。古人纂修方志时,物产大多会标有分类,有些分类较为细致如“稻之属”“兽之属”等;有些分类较为粗略如“土产”“特产”;还有些物产只用“物产”之名笼统概括,我们保留各来源方志的分类方法并设置为“分类标签”属性,另再新建一套适用于全局的物产分类体系,设为“专家分类”属性。物产的时空属性颇为重要,许多人文学者以物产所载志书的成书年代作为物产分布与变迁的时间依据,以志书中物产来源地区作为地理信息的

参照,由此也可进一步结合地理信息系统(Geographic Information System, GIS)进行可视化展示,故设置了“物产记载时间(西历)”数据属性以及古代地名映射词表作为地理对象属性,同时链接到上海图书馆的“中国历史纪年本体词表”,获取对

应朝代与年号信息。此外,《方志物产》、方志以及古籍均记载了大量物产,通过建立关系可以将这几大核心类关联起来;物产与物产之间也有同名异物、同物异名还有同名异属等关系。物产类属性如图3所示。

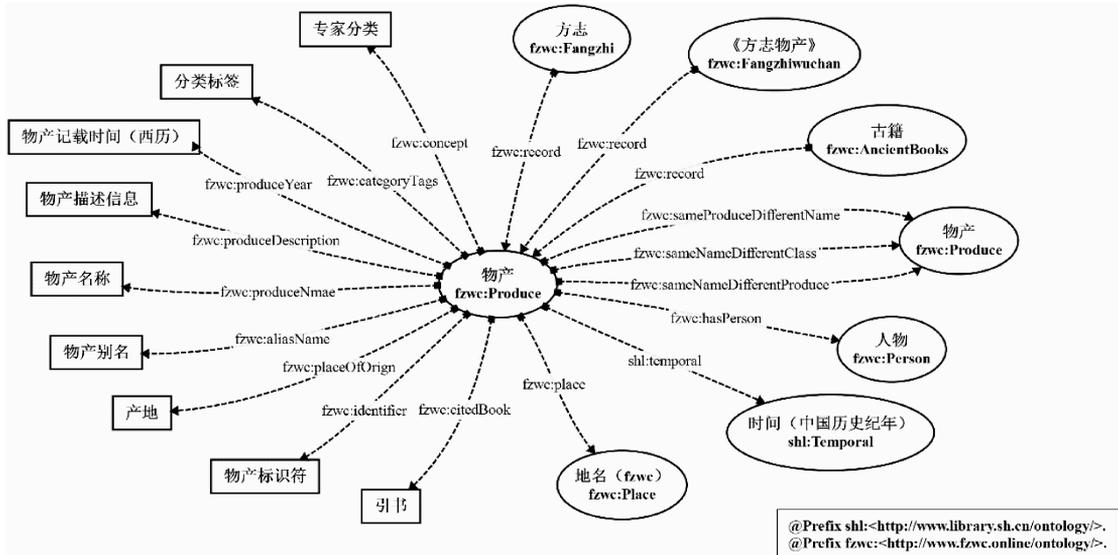


图3 物产类及其属性

表1 物产四种知识书写模式

知识书写模式	示例
模式一	土锦(以木绵花纺成绵线染为五采织以花文土人以为之衣)……(摘自明景泰六年《云南图经志书》)
模式二	葡萄 木瓜 枇杷 柿 荸荠 西瓜 茨菇 櫻桃 石榴 菱角 松实 香圆 梨 杏 桃 李……(摘自清康熙五十五年《楚雄府志》)
模式三	开化府 马金囊(即紫槟榔嚼之饮水味甘除热可治疮毒) 姜黄 攀枝花 苏木 桃榔(可为面济食一名董棕粉)……(摘自清乾隆元年《云南通志》)
模式四	野兽近边多虎豹丽江永昌景东永北产犀象牦牛大理普洱产鹿楚雄产龙足鹿永昌有白鹿武定产野牛普洱东川产岩羊临安产风兽吸风永昌产貂食铁猩猩出永昌白面猿出广西州松根豹出丽江……(摘自清光绪三十四年《云南地志》)

对属性内容进行了增补;责任者与责任方式属性统一归至人物类;方志与物产、人物、地名、方志目录等实体具有关联关系。方志作为物产的“源流”,对于学者进行物产相关知识考证有很重要的参考价值,《方志物产》为手工摘抄难免存在讹差,将不同版本的方志信息、馆藏信息乃至全文信息整合进来可为学者提供更多线索进行“循证实践”,因此我们特地加入了电子书 URL、馆藏地、来源机构数据库、方志目录等属性,同时后续还将基于关联数据与华东师范大学图书馆“数字方志集成平台”进行知识聚合,使得方志信息更为丰富、立体。

方志物产领域知识中的实体之间存在大量的关联关系,可通过这些关系进行知识推理发现潜在的知识。这些关系主要包括物产之间的关系,物产与方志、古籍、《方志物产》之间的关系,物产与人物、时间、地名等实体的关系,《方志物产》与方志、人物的关系以及方志与人物、地名以及方志目录的关系。这些关系以物产、方志、《方志物产》为中心,可深入挖掘蕴含其中的知识元素之间的复杂联系,从而形成有机的方志物产知识世界。

为了揭示《方志物产》这套珍贵方志专题资料的语义层级的知识并与其他外部知识进行有机关联、

《方志物产》摘抄来源方志约 7000 余种,其成书时间从宋至民国,每种方志对应一个版本,相应的方志题名与成书年代均有记录。本研究基于都柏林核心元数据标准与地方志元数据规范^[24],添加了方志类的相关数据属性,如摘要、卷次、分类号、类型、纂修日期、刻板日期、重印增补日期等,结合相关资料



体、属性、值”三元组来描述领域资源。通过方志物产知识本体以及其他外部本体词表,可对领域资源中的各类实体对象进行规范化描述。本文采用 RDB2RDF 直接映射的方法^[26]将关系型数据库中的数据转为 RDF 格式,即数据库中每一行代表的实体可通过表名和主键值一起构成的 UUID 来标识,UUID 为 16 位的内部唯一标识符,由 26 个字母和 1 至 9 数字随机组合生成。关系数据库中的六张数据表正好对应本体中的六个核心类及其数据属性,对象属性另外构建语义映射表,最终生成 RDF/XML 格式文件,存储于 Triple Store 数据库中。

5.3 实体关联

与外部开发数据集进行语义链接是关联数据技术的特点,可以进一步扩展方志物产领域知识的广度,促进知识的聚合与发现。本文选择上海图书馆的开放数据平台“中国历史纪年表”以及华东师范大学图书馆的“数字方志集成平台”的关联数据项目进行数据匹配与关联。通过与上海图书馆的“中国历史纪年表”的关联,可获取物产实体记载年份对应的朝代与年号信息;与数字方志集成平台关联可以获得《方志物产》中来源方志对应的其他版本、责任者、馆藏地等更多信息。

5.3.1 物产实体与上海图书馆的开放数据平台“中国历史纪年表”关联

上海图书馆的开放数据平台“中国历史纪年表”包括了从夏朝至今的朝代、年号、帝王、帝王姓名、开始年份及终止年份等属性,是一份较为完整的中国历史纪年词表。《方志物产》在对物产进行记载时均标注了来源志书的成书年份,这些西历年份可通过上海图书馆的“中国历史纪年表”转换成对应的朝代与年号,使领域知识中的时间维度信息更为丰富。本文先提取物产实体的“物产记载年份”属性,将这些西历年份通过 SPARQL 语句在线检索上海图书馆“中国历史纪年表”中对应条目,将检索得到的资源 URI 通过设置对象属性关系与物产实体进行关联。例如表 2 所示,以某物产的记载年份“1455”为关键词进行查询,得到上海图书馆规范词表中的 URI 以及其中的朝代与年号属性信息为“明景泰”。此外,还可通过与检出的起始年份计算得出 1455 年为明景泰六年,由此所有的物产实体均可获得中国历史纪年信息以及对应的西历年份。

表 2 物产记载年份与上海图书馆的“中国历史纪年表”匹配

```

PREFIX shl: <http://www.library.sh.cn/ontology#>
SELECT ? s ? d ? r
WHERE {
    ? s shl:dynasty ? d.
    ? s shl:reignTitle ? r.
    ? s shl:beginYear ? b.
    ? s shl:endYear ? e.
    Filter(? b<1455 && ? e>1455).
}
    
```

5.3.2 方志实体与华东师范大学数字方志集成平台关联

华东师范大学图书馆的“数字方志集成平台”采用美国国会图书馆的 BIBFRAME 书目数据模型将华东师范大学图书馆、北京师范大学图书馆、上海师范大学图书馆、超星、CADAL、中国方志库、翰堂典籍库以及方正电子书等机构的 55037 条方志元数据记录进行了重新整合,并对外提供关联数据的共享与利用。BIBFRAME 的核心模型为“作品(work)—实例(instance)—单件(item)”,作品为实体的抽象概念,实例是作品的出版形态,而单件决定了实例的获取方式^[27]。在《方志物产》中,物产辑录的来源方志版本是唯一的,即每种方志只录入了时间信息,其他方志著录信息则是缺失的。对于人文学者而言,获得方志原文可进一步深入比对、考证《方志物产》所载物产史料,因此方志的版本、责任者以及馆藏地或网上电子书资源等均为重要的线索信息,将方志物产领域知识中的方志实体与华东师范大学图书馆的方志作品进行关联,可获取到不同方志版本以及相关馆藏地信息,从而实现方志资源信息的知识再造。

首先以《方志物产》中的来源方志题名为检索词,通过华东师范大学图书馆“数字方志集成平台”的 SPARQL 端点获取对应名称的方志作品实体 URI;考虑到该平台采用 nodeID 来标记中间过程的空节点属性,可先根据作品实体 URI 取出非空节点的属性,再通过上一级非空节点取出位于空节点下一层级的属性值,并将查询结果进行连接;与方志物产领域知识的方志实体合并诸如方志题名、责任者信息及地理信息等具有相同含义的属性,将其他各自独有属性保留,最终《方志物产》云南卷共有 39 个方志实体可以与华东师范大学图书馆的方志实体进行关联。

5.4 关联数据存储与发布

云南地区方志物产领域知识经过数据预处理、



实体转换之后,共有物产实体 67062 个,方志实体 232 个,古籍实体 198 个,人物实体 567 个,古云南地名实体 317 个。考虑到方志物产领域知识的数据规模以及未来的可扩展性,本研究采用 OpenLink Virtuoso^① 这一典型 Triple Store 作为数据库管理软件,并基于 Virtuoso 数据库,遵循关联数据发布四原则,通过配置服务器对外进行方志物产关联数据的发布。

关联数据发布后需全面向用户展示关联数据集内各类实体的属性及其关系,我们采用 LodView^② 工具将服务器上的 RDF 数据转换为 HTML 网页进行展示,例如物产实例“牛蒡子”经过 LodView 发布后的详细页面如图 5 所示,通过“fzwc:record of”属性链接跳转到该物产的来源方志(如图 6 所示)、来源《方志物产》及来源其他古籍的详细页面。

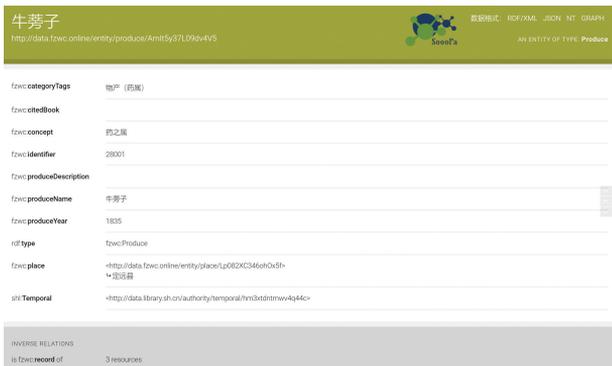


图 5 物产“牛蒡子”详细页面



图 6 来源方志《定远县志》详细页面

5.5 基于关联数据的方志物产知识库应用

方志物产知识库是面向农史领域和社会公众的数字人文研究基础设施建设的重要组成部分,相关学者可借助知识库运用农学、动物学、植物学、生态学、社会学、人类学等多种学科理论与研究方法,对

特定物产与人类社会之间复杂的互动关系展开研究;兼顾政企与公众需求,知识库还可促进科普知识推广、扩展旅游资源开发以及提升农业文化遗产保护等。方志物产知识库构建过程包括数字化、数据化、知识化、平台化等四个步骤^[28]。本文在方志物产资料数字化及数据化基础上,通过构建方志物产知识本体实现该资料的语义化知识组织,并基于关联数据与语义开发框架构建知识库系统,使其成为可为领域用户提供知识服务的平台。

目前,已开发完成的方志物产知识库系统可提供知识检索、多维知识聚合、关联数据可视化、GIS 时空展现等功能。如图 7 所示,通过检索进入物产“甜菜”详细页面,可获取该物产来源方志、来源其他古籍等书目信息,来源方志对其分类的标签信息以及方志与其他古籍对该物产的详细描述信息。另外,我们采用 RESTFUL API 技术与“搜韵网”^③以及“中文通用知识图谱”^④相连接,获取与该物产相关的诗句以及中文三大百科(百度百科、互动百科、维基百科)中该物产的词条数据。基于知识库系统及云南地区方志物产数据集,可进一步梳理、挖掘明至民国时期物产的名称、种类、数量、分布、记载起始、消长变化、逸闻趣事等状况,并探究由此产生的社会、经济、文化影响与时代特征。

物产的分布、变迁以及知识书写方式的转变是农史领域学者较为关注的学术研究热点,我们在知识库系统中基于 GIS 技术,结合“中华文明之时空基础架构”的明代历史地图(1582 年)、“中国历史地理信息系统(China Historical Geographic Information System, CHGIS)”的清代历史地图(1820 年)及近代历史地图(1911 年)等历史地图,尝试对上述研究问题进行阐释。例如,图 8 展现了“甜菜”以时间轴为序,各种方志记载的情况包括方志名称、年号、物产记载出处以及分类标签;同时结合地名关联数据中的经纬度信息在地图上进行分布展示。通过这种时空展现的方法,可以让学者便捷地追踪到某种物产最早出现于哪种方志;通过物产分类标签的变化来归纳总结各地区地方性知识书写模式的特点,并分析其背后的深层次原因。如晚清时期出现“西学东

① 网址: <https://virtuoso.openlinksw.com>。

② 网址: <http://lodview.it>。

③ 网址: <https://sou-yun.cn/>。

④ 网址: <http://zhishi.me/>。

渐”的思潮,方志编纂者对物产分类的方法受西学影响较大,物产分类标签中也开始出现“显花部”“哺乳门”等现代分类概念,由此可从知识地理学、量化史

学等视角展开分析,探讨西方科学知识与本土地方性知识的关系,以及这两种类型的知识体系是如何各自解构与建构的。



图7 方志物产知识库物产“甜菜”详细页面



图8 物产“甜菜”的时空展现

6 结语与展望

我国自古以来重视修志用志活动,新中国成立后许多机构开始对旧志资料加以分类汇编,取得了不少重要成果,这些珍贵的资料也为相关领域研究提供了丰富且详尽的史料。在当今全球化数字人文研究背景下,对这些资料进行数字化加工整理与知识组织,使其成为可在线浏览、数据共享及知识发现的知识服务平台,成为相关领域学者和社会公众的迫切需求。本文以中华农业文明研究院特藏资源

《方志物产》为例,在深入剖析内部知识元素与语义特征关系的基础上,提出基于关联数据的方志物产语义化知识组织框架,并从领域知识本体构建、关联数据构建与发布以及基于知识库的应用等方面对该知识组织框架进行实证研究。采用本体与关联数据技术可以进一步利用、共享旧方志中的物产信息,与其他领域知识进行知识聚合,使史料形态与内容全面升级,有可能发现隐藏其中单纯通过人工阅读不能发现的隐性关联知识,再通过专家多轮考证后,可



为相关学科研究提供新观点和新论据。

目前仅整理了古代云南地区的相关数据,后续有待将其他省份数据陆续添加入库完善,从而构建得到具有时间上的历时性、空间上的地域性、类别上的多样性、规模上的海量性以及语义知识点上的多层次性、细致性和广泛性的知识库,可进一步结合外部异构知识对全国范围内特定物产的分布情况、不同地区物产的丰富程度以及物产随时间变迁的消长情况、特定物产在时空框架下的变迁路线等学界热点问题展开深入研究。同时,不断增长的方志物产关联数据集将以开放数据的形式向外界提供数据消费、共享与重用,也期待与更多机构携手共建具有中国特色的数字人文研究基础设施。

本研究综合运用情报学、历史文献学、信息科学与技术等学科的知识,对中国传统历史文献地方志中的物产史料进行语义化知识组织,生成并发布关联数据,取得了较好的实证效果,既拓宽了方志汇编资料文献知识发现的边界,又丰富了特藏文献知识服务的场景,同时也实现了真正意义上的跨学科交叉融合。随着未来研究的进一步深入,客观上也将加强图书馆学、情报学与文献学与其他学科之间的联动,而这一种学科关联研究范式,既丰富了本学科的研究方法,也将推动数字人文研究方法的应用与发展。

参考文献

- 《方志百科全书》编纂委员会.方志百科全书[M].北京:方志出版社,2017.
- 包平,李昕升,卢勇.方志物产史料的价值、利用与展望——以《方志物产》为中心[J].中国农史,2018,37(3):117-126.
- 张舜徽.中国文献学[M].上海:上海古籍出版社,2009.
- 来新夏.中国地方志的史料价值及其利用[J].国家图书馆学刊,2005(1):5-8.
- 李振宏.论互联网时代的历史学[J].史学月刊,2016(11):97-113,137.
- 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.
- 鲁丹,李欣.数字人文环境下异构方志元数据整合策略[J].图书馆论坛,2019,39(4):158-165.
- 北京师范大学新闻网.北京师范大学“985工程”重大项目《中国地方志分类史料丛刊》(1663册)出版[EB/OL].[2020-04-11].<http://news.bnu.edu.cn/zx/ttgz/108126.htm>.
- 衡中青.地方志知识组织及内容挖掘研究[D].南京:南京农业大学,2007.
- 朱锁玲.命名实体识别在方志内容挖掘中的应用研究[D].南京:南京农业大学,2011.
- 李娜,包平.基于《方志物产》的物产分类体系智能化研究——以

- 《方志物产》山西分卷为例[J].中国农史,2016,35(4):31-38,143.
- 李娜,包平.方志类古籍中物产名与别名关系的可视化——基于社会网络分析技术视角[J].图书馆论坛,2017,37(12):108-114.
- Berners-Lee T. Linked Data[EB/OL].[2020-04-12].<https://www.w3.org/DesignIssues/LinkedData.html>.
- Summers E, Isaac A, Redding C, et al. LCSH, SKOS 和关联数据[J].现代图书情报技术,2009(3):8-14.
- 王燕红.国内基于关联数据的资源整合研究述评[J].图书馆理论与实践,2018(6):36-40.
- 周育彬.基于关联数据的数字博物馆语义融合研究与实现[D].北京化工大学,2014.
- 祝帆帆,高劲松,梁艳琪.馆藏文物资源关联数据的创建与发布——以中国十大绘画为例[J].图书馆理论与实践,2018(4):96-101.
- 曾子明,周知,蒋琳.基于关联数据的数字人文视觉资源知识组织研究[J].情报资料工作,2018(6):6-12.
- 侯西龙,谈国新,庄文杰,等.基于关联数据的非物质文化遗产知识管理研究[J].中国图书馆学报,2019,45(2):88-108.
- 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.
- 万国鼎.方志体例偶识[J].金陵学报,1935,5(2):48-57.
- 王雅戈,侯汉清.近代索引研究的先驱万国鼎——纪念万国鼎先生诞辰110周年[J].大学图书馆学报,2008(4):106-110.
- Studer R, Benjamins V R, Fensel D. Knowledge engineering: principles and methods[J]. Data & Knowledge Engineering, 1998,25(1-2):161-197.
- 徐周亚,王惠,肖禹,等.地方志元数据规范[DB/OL].[2020-04-12].<http://www.chinalibs.net/ArticleInfo.aspx?id=348733>.
- 徐晨飞,叶海影,包平.基于深度学习的方志物产资料实体自动识别模型构建研究[J].数据分析与知识发现,2020,4(8):86-97.
- W3C. A direct mapping of relational data to RDF[EB/OL].[2020-04-12].<http://www.w3.org/DesignIssues/LinkedData.html>.<https://www.w3.org/TR/rdb-direct-mapping/>.
- Library of Congress. Overview of the BIBFRAME 2.0 model[EB/OL].[2020-04-11].<https://www.loc.gov/bibframe/docs/bibframe2-model.html>.
- 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.

作者单位:徐晨飞、姜霖,南通大学经济与管理学院,江苏南通,226019

徐晨飞、包平,南京农业大学中华农业文明研究院,江苏南京,210095

张惠敏,南京农业大学信息管理学院,江苏南京,210095

收稿日期:2020年7月1日

修回日期:2020年11月3日

(责任编辑:关志英)

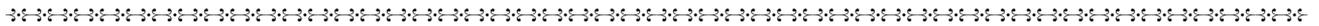


Research on Linked Data Based Semantic Knowledge Management of Produce Literature Materials from Local Chronicles

Xu Chenfei Bao Ping Zhang Huimin Jiang Lin

Abstract: The historical literature materials of produce from old local chronicles is an essential carrier for inheriting the 5000-year history of Chinese agricultural civilization, which has a large amount of data and high historical value. Knowledge organization of the materials in the environment of Semantic Web is a prerequisite of in-depth utilization and knowledge discovery based on the produce literature materials from local chronicles. Taking *Local Chronicle: Produce*, the unique collection of the organization as the research object, this paper firstly proposes a semantic knowledge organization framework based on linked data, and then designs and constructs an ontology model of local chronicle to standardize the description of concepts, attributes, and relationships. Moreover, it discusses the construction and publishing process of linked data sets of *Local Chronicle: Produce* based on linked data technology from data processing, RDF data generation, entity association, linked data storage, and publishing. Finally, taking Yunnan volume of *Local Chronicle: Produce* as an example, it constructs a knowledge base of *Local Chronicle: Produce* based on linked data for providing various knowledge services for domain users and the public. This study can activate the multi-dimensional knowledge contained in the historical literature materials of produce from old local chronicles, and also provide a reference for the semantic knowledge organization of other local chronicles compilation historical materials. Meanwhile, it enriches the application scenarios of digital humanities research in China to a certain extent.

Keywords: *Local chronicle: produce*; Semantic; Linked Data; Ontology; Knowledge Organization; Digital Humanities



(接第 62 页)

Investigation and Analysis on the Core Competencies of Electronic Resource Librarians in China

Ye Lan Zhao Ya Wang Ying Zhang Wen

Abstract: Based on the existing framework and standards on core competencies of librarians and in combination with the electronic resource service practices, this paper puts forward the core competencies framework of electronic resource librarians. It carried on surveys on the status of core competencies of electronic resource librarians in China to find out that the core competencies framework of the e-resource librarians established is basically consistent with the actual survey situation. The survey also found that at present, electronic resource librarians in China remain an overall low level of the six core competencies which are electronic resource lifecycle management, technology, research and assessment, effective communication, trends and professional development, and personal qualities. The average scores of the six core competencies are between “somewhat familiar” and “basically familiar” without any core competency yet reached the level of “familiar” or “very familiar”.

Keywords: Electronic Resources Librarian; Core Competencies; NASIG