

方志资源知识组织方法研究

赵嘉朱

(中国社会科学院图书馆, 北京 100732)

摘要: 多年来,我国方志数字化工作取得长足的进展。本文对传统方志数字化现状进行梳理和分析,并结合大数据时代和社会需求,提出方志文献库和专题库建设具体建议,最后基于本体知识组织理论,探讨方志专题库的知识组织方式,以提高我国方志资源的组织、管理和应用水平。

关键词: 本体论; 知识组织; 方志; 数字化

中图分类号: G250

DOI: 10.3772/j.issn.1673-2286.2016.4.009

方志是富有特色的重要知识资源。汉唐以来,志书编修代代相因,从未断绝。据不完全统计,目前尚存于世的历代志书超过8 200种,约占现存中华古籍总量的1/12。一般说,这些志书被学界统称为旧方志或旧志,以与中华人民共和国成立后编纂的方志相区别。后者一般称为新方志,或新志。随着改革开放步伐的加快,我国科学文化事业蓬勃发展。新方志编纂的扩展也出现了不断加速的趋势。迄今,新志总数已超过7万种,远超历代旧志总和。新方志品种齐全,数量众多,系统完备,方志文献的“功用”已经不再仅仅局限于“资治”“教化”和“存史”三个方面,而是成了中国政治经济发展必不可少的国情和地情资源宝库,能够促进历史学、社会学、法学、地理学、生物学等学科的发展,具有非常重要的意义^[1]。

基于本体论的知识组织理论近年来影响力不断扩大,为方志资源的知识组织提供了新方法。随着信息技术和网络的发展,社会知识总量以指数速度不断膨胀。知识在促进社会全方位发展的同时,本身也需要进行有序化揭示和服务。以本体论为代表的知识组织的研究和应用,取得了多方面的丰硕成果。因此,以本体论为指导进行知识组织,有助于推动方志数字化建设向纵深发展。

1 方志数字化建设主体分析

方志数字化工作于20世纪90年代末加速开展。目

前,不仅北京、上海、广州等大城市取得了显著成绩,而且很多中小城市的成果也很突出。同时,不仅东部和南部经济发达地区方志数字化日益扩展,而且西北、西南等经济相对落后地区方志数字化的整体面貌也日新月异。全国各地推动方志数字化的主要建设力量,大体可以分为以下几类。

1.1 公共图书馆

其包括国家级和省市县三级图书馆。公共图书馆拥有的地方志资源十分丰富。如国家图书馆以及北京、上海、天津、陕西、湖北、浙江、安徽、四川等省级图书馆,由于其历史悠久,资金比较充足,因而方志数字化进展较快。此外,一些地区级图书馆和市县图书馆也在积极推进方志数字化。江苏省成绩较为突出,省内昆山图书馆、苏州图书馆、常熟图书馆、吴江图书馆等地方图书馆都对馆藏地方志的纸质资源进行了数字化处理,并以全文影像提供网络共享服务,方便读者浏览或下载。

1.2 大学和科研机构图书馆

这是方志数字化的重要组成部分,包括北京大学、北京师范大学、复旦大学、武汉大学、中山大学、中国科学技术信息研究所、中国社会科学院、中国科学院等各类大学和科研机构的众多图书馆。这些大学和科研

机构不仅历史悠久,资源丰富,而且人力财力较为雄厚,方志数字化进展迅速,成绩突出。

1.3 电子信息资源开发公司

如万方、超星、同方等公司技术先进,背后有充足的资金支持,积极和各类方志收藏机构建立固定的多方面合作关系,能够迅速地完成任务大型电子信息项目开发,因而也成为方志数字化近年发展的重要推动力量。

1.4 省市方志机构

近十余年,一些省市的方志机构在当地政府的积极支持下,不断努力推进省(市)情信息网建设。目前,已有20多个省、市、自治区正在开展省(市)情数据库和网络建设。其中不少省、自治区或直辖市已经建成并且开放了网络数据库,如广东、山东、黑龙江、吉林、福建、安徽、浙江、内蒙古、四川等。省(市)情信息网是一种新的方志资源数字化形式,其网络和数据库的建立和维护,一般由当地方志机构主持。省(市)情信息网的内容一般包括历史、文化、政治、经济、军事、科技、社会生活等方面,信息规模一般以亿字计算。

2 方志数字化知识组织模式与方法

方志数字化资源建设,可以付诸应用的产品涉及诸多方面,目前主要集中在三大类,即全文影像库、全文文本库和书目库。

2.1 全文影像库

全文影像库是数字方志的基础库。它采用彩色数字扫描技术对方志进行处理,能够以图像方式再现方志原貌。全文影像库能够忠实地长期保存原始档案,同时能够快速传播,便于使用。

2.2 全文文本库

其形成可以采取两种方法:一是直接将方志纸质资源通过人工录入形成可以编辑的数字文件;二是对方志全文影像文件进行机器识别从而形成可以编辑的数字文件。全文文本库的优点包括三个方面:①实现对

方志内容的全文检索。检索,即寻找可用的内容,是方志数字化应用的重要组成部分,在目前的技术条件下,只有文本库才可以应用各种数字手段进行快速查找,如按字或词实现快速查询。②支持内容标引。只有可以编辑的文本才可以进行标引,从而实现与其他数字产品的关联查询。③方便版本校勘。只有经过识别的文本,才可以采用相关技术对不同版本方志进行校勘。

2.3 书目库

它是将方志资源目录使用多种方法单独提出,独立建库。数字文件目录与纸质文件目录具有同等效用,而前者在使用、传播等方面都较后者具有明显优势。

3 方志数字资源专题构建

国内方志数字资源的建设已经取得相当大的成绩,但仍存在一些值得思考的问题。除了相关各界经常提到的地区分布不平衡外,主要表现在方志产品种类开发不平衡。一些重要的产品基本处于未开发状态。无论从理论或构建公共文化服务体系的实际需要看,还有另外两大类应当受到关注。

3.1 相关文献库

该库主要收入各类方志的相关评论,如人物研究、名胜古迹考证、事件述评等。它是加深方志信息理解的重要源泉。作为数字方志的一个关联库,它可以起到扩大视野、拓展知识的作用。迄今为止,相关文献库的开发基本处于空白状态。

3.2 专题库

它是指经过整理合并将方志资源中某一类内容综合成一个独立的专门库,以满足特定对象对特定内容的需求。专题库可以极大地提高检索率,缩短有用资料的查询时间。然而,专题库的构建以及检索都需要多向交叉的网络技术支持,而传统的信息组织方式不能满足这方面的要求。由于这一原因,专题库的建设长期处于停滞状态,很多方面尚属空白。而完整的专题库体系则至少应涵盖以下几个方面。

(1) 地名库。该库内容包括旧地名、新地名、新旧

地名沿革、与其他相关地名的关系以及相互的影响和变化情况。

(2) 人物库。该库应收录方志中人物、选举、职官等类目中含有传记资料的人物。人物库的主要内容包括人物的本名、室名别号、更名、性别、生卒年、主要活动时代、民族、籍贯以及主要成就、贡献等。

(3) 艺文库。该库收录艺术、文化方面的实物、作品、器物等。

(4) 遗迹库。该库主要记载一个地区的名胜古迹，是方志的重要内容之一。而这些名胜古迹的废置、变化，也从一个侧面反映出当地人文的发展轨迹。遗迹库应包括方志中“八景”“十景”等所有历史文化内容。

(5) 图集库。图，是方志重要组成部分。很多情况下，用文字难以说清楚的问题，图则可以口诵其词、目明其像。缺乏图像载体，描绘一个地区自然与社会历史全貌，难以十全十美。图库建设，势在必行。其内容应包括疆域、分野、山川、江河、湖泊、城池、公署、坊都、学宫等各个方面。

(6) 自然灾害库。它包括方志中涉及的如地震、洪灾、旱灾、虫灾、雪灾、风灾等所有各类灾害的发生、发展以及应对措施等各方面的信息，对于今天的防灾救灾具有特殊重要的意义。

综上，加强专题库建设，势在必行。方志数字化欲求进一步发展，必须在构建分门别类的综合性专题数据库方面有所突破，而不能再局限于或满足于单种方志的数字化。方志的最大优势在于以地区为中心汇集资料系统。同一地区不同时期的同类资料，集合起来，就实现了内容体系的比较，便于相关类别的研究。

4 方志专题库知识本体揭示与应用

方志数字化领域日益认识到当前流行的以主题标引和分类标引为基础的信息组织深度的不足，无法满足专题库对内容标引的建设需要。基于本体论的知识组织为破解相关文献库和专题库技术难题指明了方向^[2]。相关文献库和专题库，特别是后者，是方志资源中某类知识的集合。它可以实现知识的有效获取，直接满足用户的知识需求。

4.1 本体原型

本体论原是一个哲学概念，指关于存在及其本质

和规律的学说，后来被数字技术学界借用。本体论强调特定领域中的本质概念，也强调这些本质概念之间的关联。当前学界一般认为，本体论是对概念化的精确描述，本体论的最终目标是精确地表示那些隐含的信息，使它们可被软件系统使用和共享。

本体是一个关于某些主题的、层次清晰的规范说明，它是一个已经得到公认的形式化的知识表示体系。作为一种有效表示概念层次结构和语义的理论和方法，本体目前已经被广泛应用于计算机科学和信息管理领域，并且被成功应用于构建新的智能组织和检索系统。由于本体具有良好的概念层次和对逻辑推理的支持，基于本体的知识组织和检索是基于知识的、语义的匹配，在查全率和查准率方面能够获得较好的结果^[3]。近年来，本体理论的发展和逐步成熟为基于复杂检索技术的方志数字化的进一步发展带来了新的活力。

4.2 本体论知识组织的实现

成型的本体论组织是一个动态系统，其概念、关系和软件均可根据发展的需要而不断更新。其建立过程大体如下。

本体语义关系包括同义关系、反义关系、属中关系、交叉关系、全异关系等。这与分类法和主题表有一定类似，但远为简洁和灵活，可以根据需要增减。具体步骤如下：

(1) 确立目标，划定范围。这是建立本体系统的第一步。一般说，领域越大，应用越广，所建系统就越庞大、越复杂，难度也就越高。

(2) 本体分析。即定义本体所有术语的意义及其之间的关系，该步骤需要有关领域的各方面专家参与，探讨越深入，所建本体就有可能越完善。

(3) 本体表示。一般用语义模型表示本体。

(4) 本体系统检验。一个好的本体系统应当在清晰、一致、完整、可扩展性等方面均能符合要求。清晰性指本体中的术语无歧义的定义；一致性指术语之间关系逻辑上应一致；完整性指本体中的概念及其关系应是完整的，应尽量包含该领域内所有概念；可扩展性则指本体应该能够随着该领域的发展而不断加入新的概念。

4.3 基于本体论的知识检索与应用

基于本体的知识组织在构建和检索方面的优势，使

其能够满足以方志专题库建设为主要标志的方志数字化的进一步发展的多方面复杂要求,从而趋向于成为在当前技术环境下方志数字化发展的近期方向。

5 结语

中国的方志文化源远流长,光辉灿烂,是世界文化的一部分。进入21世纪后,方志文化再创辉煌,它将肩负着“修志问道,以启未来”的伟大使命,为实现中华民族伟大复兴提供资政辅治之参考,为后世留下堪存堪鉴之记述。通过本体理论,可以进一步完善方志数字化资源的建

设、组织与服务方式,推动我国方志资源建设的发展。

参考文献

- [1] 仓修良.方志学通论(增订本)[M].上海:华东师范大学出版社,2014:339-385.
- [2] 何芸,黄立新.基于本体的大学图书馆地方特色资源库建设——以天津近代商业文化文献资源库为例[J].图书馆工作与研究,2013(9):39-41.
- [3] 宋博.论本体论在智能信息检索中的作用[J].山东图书馆学刊,2013(3):78-80,93.

作者简介

赵嘉朱,女,研究馆员,研究方向:方志资源建设,E-mail: zhaojz@cass.org.cn。

Study on the Method of Local Resources of Knowledge Organization

ZHAO JiaZhu

(Library of the Chinese Academy of Social Sciences, Beijing 100732, China)

Abstract: Over the years, our country has made great progress in China's digital local chronicles work. With the era of big data and the social demand, this paper investigates and analyzes the present situation of traditional local chronicles digitization and puts forward some specific proposals to the local literature database and thematic database construction. Finally, based on the theory of ontology knowledge organization, explores the knowledge organization method of local special library to improve our level of local chronicles resources organization, management and application.

Keywords: Ontology; Knowledge Organization; Local Chronicles; Digitization

(收稿日期: 2016-02-29)

校正

2016年第2期“基于霍尔模型的情报工程知识体系构建”一文,作者之一唐晓波教授的个人简介“武汉大学信息资源研究中心主任,教授”校正为“武汉大学信息资源研究中心教授”。