

刘 刚

建设“数字方志” 传承华夏文明

摘要 “数字方志”是一项采用数字化方式对地方志进行整理、加工,并提供给读者利用的工程。国家图书馆的数字方志,将建设包括全文影像库、全文文本库、书目库及各种专题库在内的大型数字方志资源库。其用户服务系统具有多样检索、原文再现、多屏比较、版本校勘、关联查询、个性编辑、繁简转换、版权保护、电子商务等多种功能。数字方志工程在保存和传播中华优秀文化、更好地发挥地方志资源的作用等方面,具有重要意义。参考文献 1。

关键词 地方志 数字图书馆 资源建设 信息服务

分类号 G250.74

ABSTRACT “Digital Local Chronicles” is a project to digitally process local chronicles for reader’s access. In this paper, the author introduces contents and functions of digital local chronicles developed by the National Library of China, and thinks that it has special significance for the inheritance and dissemination of Chinese culture. 1 ref.

KEY WORDS Local chronicles. Digital library. Resource development. Information service.

CLASS NUMBER G250.74

1 项目的提出

地方志为我国所特有的一种文献形式,其内容丰富,被誉为一个地区的“百科全书”,在经济建设、文化发展等方面起着重要作用。1985年出版的《中国地方志联合目录》收书 8264 种,1996年出版的《中国地方志总目提要》收录 8577 种^[1]。在进入新世纪之际,如何进一步挖掘方志资源,以促进社会的发展,是一个十分重要的课题。而数字图书馆的建设为发挥其作用提供了一个良好的机遇。

国家图书馆所藏方志冠于海内外各藏书单位之首。笔者以《中国地方志联合目录》为依据进行统计,结果表明,按编纂年代统计,国家图书馆所藏占其全部的 70%左右;按不同版本统计,国家图书馆占其总数的 60%左右。丰富的馆藏为国家图书馆开展信息服务工作奠定了坚实的基础。现在,国家图书馆正积极开展中国数字图书馆工程和全国文化信息资源共享工程的建设。采用数字化方式对方志进行整理,不仅可以极大地充实上述两个工程的建设内容,而且无论对于研究者还是一般用户,在利用方志上都可以达到事半功倍的效果。

为此,国家图书馆开始建设 1949 年以前的“数字化中国地方志资源库”(以下简称“数字方志”),并在 2002 年 7 月 8~11 日的“北京国际数字化公众信息服务与技术展览会”上展示了演示系统,得到与会者

的重视与好评。

2 项目的内容

2.1 目标

首先以馆藏方志为加工对象,在对其进行数字化处理的基础上,采用数字图书馆方式进行组织与整理,形成以方志全文影像库和全文文本库为底层,以经过规范化加工的书目、地名、人物、作品、景点、事件、插图等专题库为检索层,并实现与其他数字图书馆资源库的关联检索与跨库连接,通过友好的用户界面提供读者使用,形成内容丰富、检索便利的馆藏数字方志资源库。其次,在条件成熟时,通过共建共享方式,逐步完成全国其他机构所藏方志的制作工作,使现存于世的方志得到有效保存和广泛利用。

2.2 资源建设

“数字方志”的资源建设内容由全文影像库、全文文本库、书目库、专题库、相关文献库五部分构成。

(1)全文影像库:全文影像库是数字方志的基础库。它采用数字扫描技术对方志进行处理,形成图像资源库。这样,读者可以通过计算机屏幕欣赏到方志原貌。为客观再现方志原貌,要采用彩色数字扫描方式进行扫描,通过加工形成一个完整的筒子页展现给读者。扫描分为档案格式和网络传播格式。档案格式图像要长期保存,因此采用 TIFF 标准;考虑到目前互联网的带宽,TIFF 格式文件难于

快速传递,因此,要选择压缩率高、还原效果好的格式在 TIF 基础上进行转换,同时对转换后的图像进行加密处理以保护版权。

(2)全文文本库:全文文本库是在对方志全文影像文件进行 OCR 或采取人工录入的方式处理方志内容而形成的资源库。制作全文文本库的目的有三:一是实现对方志内容的全文检索。作为一种检索手段,全文检索在方志的利用过程中将发挥重要作用,它可以实现在较大范围内的检索,实现按字或词的随意查询。二是支持内容标引,通过对方志全文文本进行标引,实现与其他数字图书馆资源库的关联查询。三是实现版本校勘,通过采用相关技术和方法,可以在此基础上实现对同一种方志不同版本进行校勘的功能,这对于研究、整理方志是一个非常重要的手段。

(3)书目库:该库是在图书馆制作的 CNMARC 上再加上卷目名称,是图书馆检索的重要途径,也是数字方志的主要检索方式。在利用 MARC 的同时,将方志卷目提供用户使用,可以方便用户的分类查询。但是,方志卷目名称有多样化的特点,随着时代的发展,其名称也为之变化。因此,需要对其进行必要的规范处理,将内容基本相同的类目予以集中,提供用户使用。如可将官宦、耄耋、技艺等归纳在“人物”名下。同时,继续保留方志中的客观类目表,使规范化的卷目与客观卷目配合使用。

(4)专题库:专题库是根据方志特点,选择其中可以独立形成检索点的内容进行规范处理后所形成。它与全文检索、书目检索配套使用。设立专题库主要是针对目前古籍全文检索存在明显的不足,即检索范围过于扩大,导致检索后的无用信息增多,需要用户对命中结果逐个浏览后方可决定取舍,从而降低了用户检索效率的问题而提出的。只有使全文检索与专题检索互相补充,才能达到事半功倍的效果。此次设计制作以下几个专题库:

地名库。地名库首先制作方志书名所涉及的地名,然后再逐步扩大到方志内容所涉及的地名。根据图书馆界的名称规范标准,参照数字图书馆元数据的有关要求,地名库的基本内容包括原书地名、现地名、历史上其他名称、与其他地名的参照、隶属关系、变迁情况等。这样,无论从哪个途径进行检索,均可将该地区历史上曾经使用过的名称一并查询,提交用户使用。同时,通过连接,可以浏览数字方志中的全文影像和全文文本内容,还可以与其他数字图书馆相关资源库实现连接。该库可与数字方志配套使用,也可以单独使用,并可成为图书馆界地

名规范库的组成部分。

人物库。该库将收录方志中人物、选举、职官等类目中有传记资料的人物。根据图书馆界有关制作名称规范的标准,参照数字图书馆元数据的有关要求,人物库的主要内容包括一个人物的本名、别称、别号、室名、性别、生卒年、主要活动时代、民族、籍贯,以及根据该人物的主要贡献等确定的主题等项目。这也是研究、开发地方资源的重要工作。另外,制作人物库可以达到规范检索的效果,使用户任意检索一个人物的相关名称就可以得到所需要的准确内容,大大提高检索效率。该库可与数字方志配套使用,也可以单独使用,并可成为图书馆界人物名称规范库的组成部分。

作品库。该库的制作范围是方志中的艺文志、经籍志、金石志等类目中有具体内容的作品,对于书目则暂不收录。作品名称库的主要内容是一部作品的名称、作者、体裁、分类等。它们不仅要与全文影像库和全文文本库实现连接,还要与其他相关的数字图书馆资源库(如后人的评论资料)连接。通过查询该库可以挖掘方志中的各种作品,其作用与人物库基本相同。该库可与数字方志配套使用,也可以单独使用。

景点库。记载一个地区的名胜古迹是方志的重要内容,而这些名胜古迹的变迁,也从一个侧面反映出当地的发展轨迹。景点库是将方志中“八景”、“十二景”等关于当地著名景点的内容进行著录,提供用户使用。主要著录项目包括景点名称、地点、主题等。它与全文影像库和全文文本库进行连接,与其他相关的数字图书馆资源库(如旅游资源库等)进行连接。

插图库。插图是方志的重要组成部分,一般每部方志都配有各类插图,主要有舆地图、景观图、营造图、风物图、器物图、服饰图等,这些插图是研究当地社会发展、建筑风格、民俗风貌等的重要史料。插图库的主要内容为插图名称、主题等。它与全文影像库连接,与其他相关的数字图书馆资源库(如器物中的兵器可与军事资源库进行连接)连接。

事件库。大事记是方志记载一地历史上所发生重大事件的栏目,如地震、洪灾、虫灾、抵抗外来侵略等。通过检索事件库,可以了解一地历史上所发生的重大事件,也可以了解一类大事的发生地点与时间等情况,对于研究当地的社会发展有着积极作用。事件库内容包括事件名称、地点、主题等。它与全文影像库和全文文本库连接,与其他相关的数字图书馆资源库连接。

(5)相关文献库:该库主要收入后人对一部方志

的评论,对一位人物的研究,对一个景点的考证,对一起事件的评论等等。它作为数字方志的一个辅助库,可以起到扩大视野、拓展知识范围的作用。

2.3 用户服务系统

用户服务系统是实现数字方志功能,满足用户需要的重要条件。应用系统除具备翻页查看、历史记录查询、打印输出、局部放大缩小的基本功能外,还要具备:

(1)多样检索功能:实现书目(含卷目)、地名、人物、作品、景点、插图、事件、全文的单项与复合检索。适应不同用户的检索需要。

(2)原文再现功能:显示彩色扫描的古籍方志原文,使用户能够赏心悦目地浏览。

(3)多屏比较功能:可以对检索命中的记录,调出不同版本的全文影像进行多屏幕比较,使用户可以在计算机屏幕前实现以往需要翻阅大量文献方可完成的研究。

(4)版本校勘功能:可对同一种方志不同版本的全文文本文件进行自动校勘,并显示出其差异。这将大大提高有关版本研究的效率和准确性。

(5)关联查询功能:不仅实现人物、地名、作品、书名等专题检索与连接,还可以实现与其他资源库的跨库无缝连接。如检索一个人物时,不仅可以浏览该人物在数字方志的图像文件和文本文件,还可以通过相关的连接,查阅其他数字图书馆资源库有关该人物的内容。

(6)个性编辑功能:用户可根据自己的需要,对图像文件和文本文件的内容进行编辑、标点、注释等,并可将从选中的内容复制到自己的文件中。

(7)繁简转换功能:“数字方志”的汉字应建立在UNICODE基础上,并要实现整体汉字与简体汉字的自动转换,既通过输入简体汉字,得到整体汉字,反之亦然。

(8)版权保护功能:利用水印、数字签名等相关技术,实现在互联网环境下对“数字方志”资源库的版权保护。

(9)电子商务功能:在互联网环境下实现有偿使用。

3 项目的意义

3.1 “数字方志”是保存、传播中华特色文化的需要

由于历史久远,现存1949年以前编纂的方志多已破旧。为长久留存,各藏书单位一般都将原件列为专藏,很少示人,多采取拍摄缩微胶卷、影印出版

等形式,将复制品提供给读者。但这些形式的文献流通范围有限,且为黑白图像而不能客观反映方志原貌,更由于缺乏相关的规范索引,大大限制了读者的利用。采用数字化的方式整理方志不仅可以使原件得到妥善保存,而且可以通过编制相关的专题库,为读者提供功能强大的检索平台,更由于数字方志与其他数字图书馆资源库实现了关联查询,丰富了用户的检索途径,方便了用户的利用。同时,采用彩色扫描方式的方志图像,可以通过互联网向世人传播、展示中国五千年的悠久历史,弘扬中华文化。

3.2 “数字方志”对于促进经济建设与社会发展有重要作用

方志蕴藏的丰富内容具有重要研究和应用价值,以往已经发挥了重要作用。但由于采用手工作业,耗时费力,效率低下,使得人们深入挖掘整理方志内容的愿望不能实现。而“数字方志”将可以在一定程度上解决这些问题。读者可以从不同角度检索、利用方志资源,满足不同需要。这对于经济建设与社会发展将起到积极的促进作用。

3.3 “数字方志”是中国数字图书馆工程和全国文化信息资源共享工程的重要组成部分

中国数字图书馆工程与全国文化信息资源共享工程的实施,将极大地推动图书馆现代化建设与资源的共享。两个工程资源建设的主要任务之一是整合各类特色文化信息资源,提供读者使用。而“数字方志”无疑是其中极具价值的资源。

3.4 “数字方志”可以强化科研、培养人才

“数字方志”涉及很多领域,需要有研究成果做支撑。如果没有相应的人员进行整理、标引工作,就很难达到理想的效果。因此,数字方志的建设过程也是一个进行科研和培养人才的过程。通过建设“数字方志”,可以推动古文献整理及其标引规范等方面的研究,培养出一批专业人才。

参考文献

1 金恩辉. 中国地方志总目提要. 台北:汉美图书有限公司,1996

刘刚 国家图书馆分馆副馆长,副研究馆员。通讯地址:北京市文津街7号。邮编100034。

(来稿时间:2002-09-01)