

数字图书馆建设中的古籍数字化探讨

庄舒蓉

(福州职业技术学院 图书馆,福建 福州 350108)

摘 要: 数字图书馆利用高科技为古籍文化解决了传播不易的问题,即古籍数字化。通过介绍数字化古籍的影像处理、全文资料库建置、机读编目制作、索引编制及网络古籍资源检索等方式,分析了图书馆古籍数字化过程中存在的影像标准化、古籍应用系统、古籍合作编目、Metadata 及网络古籍资源的利用等方面的问题,阐明了古籍数字化及建立有效的管理机制的重要意义。

关键词: 数字图书馆;古籍数字化;古籍整理

中图分类号: F507.454

文献标识码: A

文章编号: 1671-931X (2012) 05-0089-03

一、引言

古籍即中国古代书籍,它蕴含着中华民族丰富多彩的文化遗产,民族文化有赖于它而传承、延续。然而古籍是经过长久积累形成、保存下来的,它的存在形式、包含内容与现代文化有着很大的差异,造成了使用上的障碍。中国文化博大精深,由于历史原因古籍散布世界各地,其中大多保存在图书馆中,作为古籍收藏重地,图书馆肩负着特殊的使命,而它所藏的古籍具有自身的特点,一是有着特殊的价值;二是不易保存;三是内容、形式多样化;四是生产数量相对固定。

作为电脑技术和网络技术发展下的产物,数字图书馆具有信息储存、整理及多方共享等多项功能。随着社会各界对历史文化的关注和程度越来越高,古典文献数字化也成为人们关注的焦点。受到科学技术的冲击,图书馆在作业形式发生变化的同时,古籍整理也有了相当大的改变,建设数字图书馆古籍是不可或缺的,应根据中国传统文化的特征,解决其时代性所带来的问题。

因此,古籍的整理、组织与利用逐渐成为数字图

书馆时代图书馆的重任。而古籍作为中国文化的经典,网络环境能够充分发挥自身优势通过电子媒体将其加以传播:(1)电子版古籍可反复复制,知识共享范围广泛;(2)在网络环境中,电子版古籍不存在运输和分配上的问题,可以一日千里;(3)电子版古籍易于整理,经过参照比对后能产生新信息;(4)电子版古籍便于使用。综上所述,电子版古籍便是将古籍现代化,便于人们传阅、学习。古籍是中华民族文化的遗产,古籍电子化是在网络上反映民族文化的不可缺少的方式,古籍电子化标志着古籍整理进入了一个新的里程。

二、古籍的数字化方式

作为古籍的“藏身地”,图书馆在保护好古籍的同时,还应尽力为使用者提供古籍的内容知识。获取古籍知识应先整理出其线索,古籍的内容和线索是古籍整理工作的两个重点对象,这一原则同样适用于数字图书馆中的古籍整理。古籍影像化和全文资料库呈现是数字图书馆提供古籍内容的两种方式,前者可以还原古籍,便于浏览、使用;后者将古籍全文转为电子方式,如编制古籍的诠释资料、机读编目及

收稿日期:2012-06-09

作者简介:庄舒蓉(1981-),女,福建泉州人,福州职业技术学院图书馆助理馆员,研究方向:图书馆学。

编制索引,它们的共同之处在于都是从网络上获取古籍资源的重要途径。因此,图书馆古籍数字化及其组织与检索的途径包括影像处理、建置全文资料库、机读编目、编制索引、网络古籍资源的组织与检索等。

(一)“如出一辙”的影像处理技术

通过将古籍或文献扫描输入电脑进行数字化,将珍藏在图书馆的古籍以最真实的形式传承,并且做到最大程度保持原貌,这一方法便是影像处理技术。1995 年,美国的国家数字化图书馆将利于学习美国文化及历史的原始资料数字化,获得明显成效,其数字化的对象除了古籍,还包括图书、文献、图片、录音资料、电影片等等,这个计划带动了世界各国数字图书馆对文化和历史资产的重视,相继有许多保护文化遗产的计划及图书馆纷纷以珍藏文献为数字化的主题,如联合国世界文化遗产数字化计划、大英图书馆的数字化计划以及中央图书馆的古籍善本数字化等等。

(二)“全盘把握”建置全文资料库

古籍影像化虽能以最真实的数字形式复制、保存古籍文献,但也存在着不足:一是若原稿的文字甚小或虫蛀严重,即使影像化后仍无法判断细小的注记文字;二是古籍的影像化便于浏览、传递,但不易进行编辑、查找。如果古籍的全文实现了电子文本化就可以克服上面的问题,建立古籍全文资料库可将古籍文档录入电脑,使之数据化,而后通过文字辨识软件进行影像的处理、古籍文字的识别和分析、信息的检索及构建古籍资料库,便于古籍文献查询与阅读。

古籍全文资料库虽然解决了检索问题,但对于古籍各版本之间的对应、注疏和正文间之间的参照等问题仍无法处理,针对涉及到古籍的内容结构和外在形式的特殊问题,中央研究院开发出的中文文献处理系统,可以对古籍进行超文件处理,解决这一问题。

(三)“提纲挈领”制作机读式编目

在数字图书馆的建设进程中,机读目录的编制是必需的,中国及北美都已编制古籍机读目录。1988 年 12 月,共有包括台湾大学图书馆在内的超过 8 万笔书目记录,被放在国家图书馆的全国图书信息网络上的善本古籍联合目录内。1980 年,北美地区美国研究图书馆协会的中国古籍善本国际联合目录计划,欲用机读形式编制中国清嘉庆以前的印本及抄稿本古籍文献,此计划称为中文善本书计划,目前这一计划由美国普林斯顿大学东亚研究所负责研究;1991 年,包括中国大陆 5 所,美加地区 13 所,共 18 所学术机构的图书馆参与提供书目资料;1998 年,这一目录在已具有 12,000 笔书目资料的情况下,每年稳步增加 2,500 笔;1999 年,国际电脑图书馆中心使用中日韩文自动化系统建立资料库,有 10,698 笔十八世纪以前的出版品。我国进行古籍机读编目的图书馆除了上述参与目录计划辽宁省图书馆、湖北省图书馆、北京大学、复旦大学、

中国科学院,还包括 2000 年北京图书馆按照机读格式著录宋元版古籍,录入古籍善本、新善本书目数据 3 万多条,南京图书馆相继建置了南京馆藏方志目录、南京馆藏书目、中国古籍总目等 3 个古籍书目数据库。

(四)“由点到面”编制索引

作为图书馆整理资源、服务读者的方式之一,编制索引长久以来倍受关注,如今编制索引已发展为利用电脑编制机读书目,并成为馆际、甚至国际间的合作系统。因为索引在这方面的开发较为缓慢,过去检索方式主要是人工逐字检索,但在实践中人们发现全文检索系统更加全面、快捷,以往的逐字索引已被全文检索系统代替,早期人工编制索引相关的标引信息都能通过电脑在短时间内检索到。如中央研究院建立了有史以来容量最大的汉籍电子文献,内容包括台湾史料、二十五史、大正经及其他典籍,其中的台湾方志资料库包括 46 中志书,可谓门类齐全,无论细到人名或地名都能在这一资料库中检索到。以往对于查阅方志、类书中的资料以及检索古籍中的篇名较为困难,而目前借助电脑也可以快速解决。

(五)网络古籍资源的利用

Metadata 是对封装品资料属性的描述,它是通过网络进行古籍资源组织与检索的主要方式,Metadata 译为资料描述格式、诠释资料、超资料或元资料,它的主要功能便是搜寻资料。Metadata 的著录功能便是使资料的持有者通过 Metadata 认知、分辨、组织及运用资料。换句话说,网络电子资源可通过 Metadata 进行搜寻、记录、评价与选取。在目前的 Metadata 中,都柏林核心集及电子文献编码记录受到广泛的重视。

三、数字化过程中亟待解决的难题

以上所提的影像处理、建置全文资料库、机读编目、编制索引以及研制古籍的诠释资料 Metadata 是图书馆在数字化环境下整理古籍的几个努力方向及发展的趋势。早期由于古籍的使用率不高,所以在图书馆自动化的顺序往往被排在后面,以至于不如其他馆藏一样受到重视并有许多实际的成果和研究心得。然而发展的进度虽然不如其他馆藏,但从另一个角度来看,古籍能积极吸收过去图书馆自动化的经验,并利用目前已开发的技术来处理,不但可以在已建立的基础上吸收经验,加以应用,而且可以研发出更先进的技术,作为其它馆藏的参考。目前在利用资讯技术处理古籍虽有所成就,也面临着困难和挑战。

(一)影像处理标准化问题

制作图像不同的对象所需要的图像效果可能不一样,那么就应该根据不同的需求来制作不同的解析度,而制作出来的图片所呈现出来的效果也应因人而异进行修整,这其中包括黑白及彩色图像储存格式上的不同。古籍影像资料库建置过程中需要用

到多种技术进行操作,其中包括中文文字辨识系统;储存媒体;文件影像分析处理技术,其中中文文字辨识技术为最大技术瓶颈所在,也是系统自动化处理的最大障碍。未来应针对古籍的特点和需求,考虑文件影像的分析和处理,研发可以辨读各种文字和字体的新技术,开发出失真率低、压缩比高的影像技术。此外,预期中将来各单位所制作古籍影像的资料量相当大,所以有必要在储存格式及媒介等相关问题上先行设定一个标准,这样更有助于以后的管理。

(二)古籍文字应用系统的缺陷

目前已开发的较为完善的影像文件辨识、检索与管理系统的优点在于:能逼真的重现原文件;检索内容详尽;系统自动对输入电脑的文件进行一系列操作。这一系统将影像处理、文件分析、文字辨识等技术应用于其中,但目前尚无文字辨识系统能有效辨识手写或木刻的古籍,仅识别铅字、活字版或电脑印刷文件,因此这一功能有待于进一步开发,希望将来能改良出一套理想方案。

(三)古籍合作编目的标准制订

大陆、台湾及北美等三个地区内的图书馆在不同的时间、空间,依据不同的建档规范及作业模式,所建立的书目资料难以一致,但也各有其区域性的特点:大陆地区所订定的规范较为切合实际,目前虽然书目数量不足,但将来不论是独立发展或与其他地区合作,其发展空间都无限广阔,另一方面因为书目品质追求精良,且经费受限因此在量上无法大幅增长;台湾地区目前在数量上虽然最多,但在编目规范的标准化以及著录内容的精确上都有待增强。既然中国古籍不再继续增长,而古籍编目又牵涉到版本的考证问题,所以合作编目应该是上述三地区彼此间以及各自内部间建立书目数据库的可行方式。合作建立中国古籍书目资料库应考虑到收录范围及类型、著录规则、分类及标题方式、建档格式、中文内码等因素,这些问题的解决有赖共同制订统一的标准才能得以实施。

(四)Metadata的改善

在数字化图书馆建设中,古籍文献是其重要内容,但信息检索的重要性已经远远超过了文件资料的数字化或影像化。如上述提到的描述珍藏品资料属性的Metadata,一般是在封资料的检索和运用时使用。因此珍藏文物要数字化,除了文物、文献本身须数位化以外,还须建立书目信息并做合适的信息组织,以便提供有效的检索方法。古籍既然是国内外图书馆珍藏文献中最主要的藏品,倘若在电子图书馆的环境中来探讨古籍整理与利用,Metadata的发展及其和古籍的关系是不容忽视的。

四、结语

图书馆在数字化与古籍整理的过程中扮演着枢纽

的地位,不仅因为图书馆是保存者,还肩负着规划、执行和协调的角色。数字图书馆古籍整理,在面临常规挑战的同时还承受着新的冲击和时代性需求,包括:(1)更新观念。数字化图书馆环境下进行古籍整理,除了应用新技术对古籍加以保存,更应注重深层次的技术上的进一步开发和利用,使信息得以最大限度的使用。这一观念上的更新应结合读者需求,有计划有目的有条不紊地进行;(2)增加内容。古籍数字化过程毕竟是将古籍融入网络科技的过程,在这过程中增加相应内容的数据以便于古籍数字化是不可避免的,如扫描与编辑图像、相关信息连接等;(3)加深文献揭示程度。在古籍整理过程中文献揭示程度不能仅停留在表面上,应深入到文献的细节中;(4)改变工作环境及方式。借助现代化工具完成。

将来各图书馆对古籍的数位整理势必处于分散状态,所以若能建立一个有效的管理机制,将有助于馆际间的整合与分工。在这方面,美国的国家数字化图书馆提供了一个参考模式,该计划是以为期五年的遗产的历史性档案来数字化,透过电子图书馆的建立,协助研究美国的历史与文化。此计划的范畴,包括:馆藏的选择和数字化、提供更方便有效的检索方式、设计适合各式各样使用者的电子图书馆架构等层面。国内目前正在推动国家典藏数字化计划,其内容是将各类重要文物资产加以数字化,以充实网络的内涵。在构想中,是由各单位将其重要馆藏分别数字化,但是建立一个共同的管理机制,以达到信息的共享。届时在执行时,可以根据实际的需求,并参考美国国家数字化图书馆计划,来建立一个有效的管理机制。

数字图书馆带来的古籍整理上的变化,是一种时代的需求,这种变化不是单纯的技术转变,更是强调观念的变化。数字时代将促进古籍以高科技的形式提供的服务内容更丰富,形式更加多样化,以及更深层次的的网络文化信息。在数字化时代,古籍所赋予的时代意义有待于我们潜心研究,而古籍的整理方式也应受到进一步关注与理解。

参考文献:

- [1] 毛建军.古籍数字化理论与实践[M].北京:航空工业出版社,2009:162.
- [2] 潘树广.古籍索引概论[M].北京:书目文献出版社,1984:253.
- [3] 王纯.古籍数字化之趋势[J].图书馆理论与实践,2000,(3):50-51.
- [4] 李郎达.Metadata初探[J].情报科学,2001,(6):606-607.
- [5] 陈爽.网络文史资料库一览[J].文史知识,2002,(6):109-116.
- [6] 康尔琴.中国古籍书目数据库建设浅议[J].图书馆界,2003,(2):14-16.

[责任编辑:张磊]

(下转第94页)

（上接第 91 页）

The Exploration on Digitalization of Ancient Books

ZHUANG Shu-rong

(Fuzhou Polytechnic, Fuzhou350108, China)

Abstract: Digital library solves the difficulty of spreading ancient culture with the new technology. That is to digitalize ancient books. The article presents the image processing of ancient books, the construction of full text database, computer cataloguing, index making and searching of online ancient books resource. Then it analyzes the problems in image standardization, the application system of ancient books, cooperative cataloguing of ancient books, Metadata and use of online ancient books resource. To sum up, it emphasizes on the importance of digitalizing ancient books and establishing an effective management mechanism.

Key words: digital library; ancient books digitization; ancient books