

●杨晏平

建立图书馆古籍文献数据库刍议

ABSTRACT There is a great amount of ancient books in the library demanding prompt arrangement and development. The only feasible way to solve the problem is to set up a data base for literature resources of ancient books and to accomplish automation and network system of the management work for Chinese ancient books in libraries of the whole country. 9 refs.

SUBJECT TERMS Library collections-Studies Data base of literature resources-Ancient books

CLASS NUMBER G356.1

计算机技术、现代通讯技术、现代文献制作技术正步入图书馆。图书馆自动化系统不断健全,自动化程度日益提高。作为图书馆自动化工作基础与核心的图书情报数据库技术得到了迅速发展,各类型多用途的图书情报数据库不断涌现,如馆藏数据库、全文文献数据库、连续出版物数据库、文摘数据库以及引进磁带文献数据库等。但是古籍文献数据库的建立尚未引起足够的重视。

1 建立古籍文献数据库的必要性

古籍,原则上指书写或印刷于1911年以前,反映中国古代文化,具有古典装帧形式的书籍。

我国历史悠久,千百年流传下来的古籍浩如烟海,据不完全统计,种数约达12万左右(不含每种古籍的不同版本)。其中,全国公共图书馆收藏善本古籍200多万册,普通古籍2600多万册。这些数字似乎无法与“知识更新”、“信息爆炸”时代产生的文献相比,加之这些古籍分散于众多图书馆中,就更显得

微乎其微。但这些古籍是中华民族灿烂文化与历史的象征,是人类精神财富与智力资源的宝库,对社会的过去与将来具有特殊意义。可以说,历经沧桑留存至今的每一种古籍都是国之瑰宝,是任何其他文献不可替代的。这些古籍在整个图书馆藏书体系中也是不可缺少的重要组成部分。

由于编排体例、文字内容、装帧形式与普通图书不同,古籍的编目著录有一定难度,又因条件所限,大多数图书馆只能沿用手工方式进行加工处理。结果速度缓慢,质量不高,致使许多古籍难以提供借阅。在北京图书馆为例,因所藏普通古籍60%是简编或未编书,为将这部分馆藏公诸于世,从1983年起合流编目,12年来,耗费了大量人力,仅完成总量的12%左右,照此计算,还要八九十年方能大功告成。据了解各馆情况普遍不够乐观。陈旧、落后的手工编目形式远远落后于现代图书情报事业的需要,非改不可。

多年来古籍一直以传统的卡片目录为检索工具,体系庞大,使用维护颇为不便。在北京图书馆普通古籍合流编目过程中,目录环

节投入的人力往往与编目人员不相上下。据估算,待全部合流完毕,将产生书名、著者、分类三种目录卡片 25 万余张,如果再考虑丛书子目分析、参照、互见、续片等,至少可达 80 多万张卡片,以每个目录屉存放 800 张卡片计算,需 1000 个目录屉方能容纳,而这部分古籍藏书仅占全馆藏书的 11.9% 左右,可以想见,合流编目完成以后,全部馆藏的目录体系会有怎样惊人的扩展,不但存放空间难以解决,庞大的目录体系也会造成管理与检索的诸多不便。况且,其间周转时间长,程序多,质量无法保证,更不必说还要耗费大量资金与人力。这种现状很难满足图书馆古籍工作现代化发展的要求。

我国的古籍不仅国内图书馆有藏,还有相当多的珍善本古籍或稀见之本流散于国外,收藏分散,不便于查找利用。为了给读者节省时间、精力,让更多的人了解、利用古籍,也为了保存、弘扬中华民族的优秀文化,大量的古籍亟待系统整理,丰富的文献资源亟待深入开发,唯一可行的办法是实现全国图书馆古籍管理工作的自动化、网络化。

图书馆古籍管理引进自动化机制,建立古籍文献数据库已势在必行。当前,建立古籍文献数据库的有利因素是:古籍数量、种类有限,今后不会再有大增大减,便于集中力量解决问题;不利因素是:有关古籍的研究一直比较薄弱,没有经验可以照搬,只有靠努力探索。

2 建立古籍文献数据库的条件

建立古籍文献数据库是个系统工程。与普通图书相比,面临的困难、需要解决的实际问题更为复杂,所需的人财物各方面条件也有更高的要求,因此不可能急于求成。应持科学态度,结合国情、馆情,努力创造条件,从基础工作做起,循序渐进。

2.1 领导重视

建立古籍文献数据库需要一定的人力配备和物质保障,需要馆内外多方面的协调配合,这是一个较长的时间过程,会出现许多意想不到的情况,因此首先要得到领导的重视。领导应具有远见卓识,克服传统思想的束缚。领导决策与管理水平是实现图书情报工作自动化的关键。各级图书馆领导应将古籍文献数据库的建立提到议事日程上并予以切实的支持,实地组织有关问题的研究解决。万事开头难,重要的是行动,决不能持等待观望的态度。

2.2 人员培训

现代图书情报工作不断地向专业化、知识化、多学科、多层次转变,在古籍处理工作中要改变传统模式,引进自动化技术,就必须培养造就一支新型的古籍工作队伍,通过岗位培训、参观学习等多种方式,使他们既具备图书情报知识、古籍及相关学科知识、熟练的古籍加工处理能力,又能主动了解有关领域的国内外发展动态与最新成果,并掌握计算机等现代工具的基本知识与操作技能。

2.3 资金与设备的保证

巧妇难为无米之炊。建立古籍文献数据库需要一定的资金与设备做后盾。各馆可根据本馆的财力物力选择最佳方案,购置相应的软硬件设备,确定建库的初步规模与方式。

2.4 加强古籍保护措施

数据库的建立需要较长的时间,其间如果由于保护不善致使古籍流失、损毁,数据库也就无从建立。以北京图书馆普通古籍为例,新馆建成后善本古籍进入了设施良好的新型书库,而多达 200 万册的普通古籍只能存放于旧馆的大书库中。建于 30 年代的书库已十分陈旧,无防护设施,顶层漏雨,存放在普通开放型铁书架上的古籍布满灰尘,或因受潮、日晒而变形,书页风化、糟朽霉烂,或虫蛀严重、字迹漫漶,任其发展下去,将会造成无可弥补的损失。据了解,由于经费短缺,设备落后,空间紧张,各图书馆的古籍保护条件普遍

较差。加强古籍资源的保护已刻不容缓,应尽快采取有效措施,利用先进的科学方法与现代技术手段,使尽可能多的古籍延长寿命,以保证所建古籍文献数据库的相对完整与系统。

3 建立古籍文献数据库的几个具体问题

3.1 根据馆情,确定建库的总体规划与合理布局

建立古籍文献数据库对大多数图书馆来说是新的尝试,应首先进行周密调研,认真分析,立足于本馆实际,确定建库的总目标和具体实施步骤,不可脱离实际,盲目上马,追求大而全,小而全。由于古籍收藏比较分散,在建库之初应尽可能与全国各主要古籍收藏部门取得共识,协调一致,共同确定建库的总体规划与合理布局,使所建数据库系统具有通用性,为达到资源共享创造条件,也可避免人财物的重复浪费。

目前,各种类型、不同规模、多种层次、实用性强的数据库层出不穷,不妨借鉴他人经验,少走或不走弯路。

一个完备的图书馆古籍文献数据库应包括文献数据库和业务工作数据库。前者由书目数据库和文献型数据库组成。后者由采访、编目、流通、典藏、用户、统计、索引编制几部分组成。

可根据本馆资金能力,设备情况,服务对象,藏书规模,人员水平等条件,本着先易后难,先简后繁,以点带面的原则,从小处着手,为建库做好充分的前期准备工作。一般宜先从建立书目数据库开始。

3.2 数据准备

(1) 数据的标准化和规范化处理。建立数据库的首要环节是要对数据进行规范处理。普通图书由于出版规范化,具有固定的著录来源,从而提供了统一可靠的数据。古籍则不然,必须经过一番考订,对全书进行分析、

鉴别、筛选,才能确定需要的内容。对相关著录项目用标准化、规范化加以限定,进行必要的处理,在建立古籍文献数据库中具有重要的特殊意义。

我国已经有了《古籍著录规则》(GB3792.7—87),并参照《国际标准书目著录(古书)》I ISBD(A)加以修订,其目的是解决汉字古籍的著录标准化,为我国实现汉字古籍整理、加工的自动化以及沟通国际间的信息交流、传递提供准则和依据。但这一著录规则也只是在文献著录项目、顺序和标识符等方面有章可循,还应制定相应的规范档,编写适于古籍的分类表、主题词表等。只有采用标准与规范,使数据严格有序,才能保证古籍文献数据库的高质量。推行贯彻古籍著录标准与规范应该自上而下,采取措施,加强力度,使古籍工作尽早纳入标准化、规范化轨道。

(2) 准确无误地输入数据。数据输入既是直接关系到数据库能否正常使用的重要环节,又是十分繁杂琐细的基础工作。具体方法是按照专门设计的工作单中编目人员所记录的标引结果逐项输入计算机。为确保准确,应由经过培训的兼具古籍著录知识的熟练的计算机操作技术人员承担,可能的话,由编目人员直接上机。录入数据后经专门人员审查无误,正式存储,形成数据库。

从某种意义上说,数据的标准化处理与录入远比建立数据库本身更为重要,这是一项长期而又艰巨的工作,来不得丝毫马虎,如果输入的数据杂乱无章,甚至谬误百出,所建数据库将无法使用。

3.3 研制、开发适于古籍的应用软件

在建立古籍文献数据库的过程中,必须根据古籍特点、数据内容、用户需求编制应用软件,程序设计须由计算机专业人员与古籍工作人员共同协商,使数据库的使用达到预期效果。

在程序设计时,选择适用的字库很重要。

应根据古籍文字特点在汉字数量、造字功能上多下功夫,以满足繁体汉字输出、异体字转化等需要。但仅仅做到“简体录入”,“繁体输出”也不行,如“松”字用于地名,古今都一样,输出繁体,就会出现“鬆花江”的错误;又如“發”与“髮”同为“发”的繁体,也不能一概而论。应根据实际情况,在选择汉字时,使其具有自动分析判断功能。

4 开发与建立古籍文献数据库的前景

4.1 图书馆古籍处理各环节工作自动化

利用古籍文献数据库的基本数据,视各环节的特殊需要增删修改,可生成不同用途的数据库,在此基础上形成一系列古籍处理自动化系统。由于计算机运算与处理数据的速度极快,且一种书只需著录输入一次,即可产生多种数据,因此提高了效率,也将工作人员从繁琐的手工操作中解放出来。

4.2 改进古籍文献检索方式,增加检索途径

传统的古籍检索方式,查找不便,费时费力。古籍文献数据库既提供了不同角度的多个检索点,诸如篇名、卷数、著者、分类、主题、专题、专科以至书中的个别内容等,与手工查检相比,提高速度近40倍,还具有极高的查准率、查全率。

4.3 自动生成二次、三次文献,实现古籍检索情报化、计算机化

长期以来,读者阅读古籍文献是直接阅读原文,使用不便,又不利于古籍的保护,在古籍资源的开发方面也受到限制。把静态的馆藏资源变成情报、信息流,增加揭示古籍的深度与广度,形成二次、三次文献,会逐渐成为古籍开发工作的主流。地方志、人物传记、医药医学等类古籍,可作为首先开发的重点。

4.4 实现目录机读化,改善图书馆环境

传统的古籍著录以纸制卡片为主要载体,建立一套卡片目录体系需要大量人力投入,效率不高,生产周期长,质量不易保证,体

系过于庞大,不便使用,易于破损。使用机读目录从根本上克服了手编卡片目录的不足,而且还具有手编目录达不到的种种优越性:信息存储量大;处理速度快;能短时间内查遍数据库中的全部数据,查全查准率高;输出方式多;输出的目录种类多,可以根据不同需求随时转换成卡片目录、书本目录、缩微品或磁带、磁盘等形式;所存储的书目信息容易删除、改写或重新组织,目录体系容易维护。建立古籍文献数据库,机读目录取代传统的卡片目录后,图书馆不需要很多目录柜,取而代之的是优雅宽敞的环境,读者和工作人员不再有逐张翻阅卡片的烦恼。

4.5 采用光盘-微机系统,提高存储与检索效益

光盘技术自70年代后期间世以来,以其体积小,重量轻,容量大,费用低等优越性而受到用户青睐,建立古籍文献数据库需要存储与检索大量的数据,更适于采用光盘与微机相结合的光盘系统。

光盘按用途分为:只读光盘、只写一次光盘、可读与可写光盘以及激光视频盘。其最大特点是所谓海量存储,容量约为同尺寸软盘的500倍,一页按2000字符计,一张5.25英寸550MB的只读光盘能存27万页文献。

光盘还具有许多优点:传递速度比软盘高2倍多;新一代的可擦写光盘能够在盘上反复读写与更新数据,其次数高达 $10^6 \sim 10^7$ 次;应用范围广,可存储正文、数字、图象、音乐等多种形式的信息;易于操作使用,寿命长,耐冷热,耐指纹,不易碎裂,不受磁场干扰,数据不易被破坏。

建立古籍文献数据库采用光盘技术可产生特殊效用:

(1) 提供了保存古籍的经济实用的方式。可将古籍全文存储于光盘,也可将原存于缩微胶片上的文献转换到光盘上,不但有效地从外观到内容上保护了古文献的全貌,而且解决了永久性存储问题。这是因为一次写

光盘不可改写,用于保存文献万无一失。并且光盘使用寿命10到100年,还可随时快速地完成盘片复制而不会产生任何误差,其保存珍善古籍效果大大优于缩微胶片。

(2) 改进服务方式,提高古籍利用率。由于古籍的特殊价值,大多数图书馆对借阅手续、范围有很多限制。有了光盘可以实现多用户同时取用,无需再为古籍丢失、损坏而担心,也不会因没有复本而拒借。可供读者存取的既有书目数据、文摘索引,也有古籍的全貌,读者想要复制其中的内容也只是举手之劳。

(3) 可以加快古籍及相关资料的出版。针对用户的咨询请求,由计算机自动编辑处理,可在一二天内完成。

(4) 在尚不具备联机检索的条件下,光盘—微机系统不用联网,也可达到资源共享效果。如美国OCLC编目光盘数据库中,2张光盘存储了古籍图书和有用的非书型资料编目记录约120万条。采用这一系统,各馆将所藏古籍的全部有关记录(含全文)存入光盘,至多几十张光盘就能容纳,只需通过互相购买光盘就能共享对方的古籍资源。

4.6 联机检索,资源共享是建立古籍文献数据库的最终目标

国际联机检索,共享图书情报资源是图书情报工作的理想境界,也是建立古籍文献数据库的最终目标。现存古籍资源有限,又极为分散,一个图书馆的经济实力再雄厚,都不可能将存世古籍收藏齐全,因此,实现古籍联机检索更具有现实意义。

要按照统一的要求与严格标准先建地区、系统的图书馆古籍文献数据库网络,待水

到渠成,再向更大范围扩展。美国研究图书馆组织(RLG)所进行的中文善本书项目(CHRB)已经开了先例。它借鉴现代编目方法,1989年起将中文善本书记录输入到RLG的计算机网络(RLIN)中,并在实践中形成了《RLG中文善本书编目规则》,于1993年6月提交美国图书馆协会编目审查委员会(CC/DA)审议,拟作为美国的国家标准使用。这表明古籍文献数据库与国际联网的可行,也增加了发展我国自己的古籍文献数据库、实现网络化的紧迫感。

参考文献

- 1 陆长旭. 中国的文献数据库. 图书情报工作, 1994, (4): 50~54
- 2 宓永迪. 应重视数目数据标准化. 图书情报工作, 1994, (6): 49~50
- 3 郑莹. 图书情报管理现代化实施条件之我见. 见: 当代图书情报管理的理论与实践. 北京: 兵器工业出版社, 1993: 279~282
- 4 纪昭民. 文献数据处理标准化是建立文献信息网络的关键之一. 上海高校图书情报学刊, 1994, (3): 5
- 5 马自卫等. 图书情报自动化. 北京: 邮电出版社, 1993: 295
- 6 曹淑文. 现代编目方法与中文善本特点相结合. 中国图书馆学报, 1994, 20(6): 54~55, 39
- 7 杨学梅, 张大尧, 崔明. 信息高速公路、金桥与图书馆. 中国图书馆学报, 1994, 20(6): 44~48
- 8 张力, 魏京学. 试论高校图书馆微型计算机应用模式. 高校图书馆计算机应用经验交流论文集, 32~40
- 9 许绥文. 文献处理自动化发展的新趋势. 见: 北京图书馆同人文选·第二辑, 北京: 书目文献出版社, 1992: 123~127

杨晏平: 通讯地址: 北京文津街7号, 邮编: 100034.

(来稿时间: 1995—05—24. 编发者: 徐苇.)

(上接77页) 心和决心,在邓小平同志建设有中国特色社会主义理论和党的基本路线指引下,努力工作,以新的成绩迎接和进入21世纪。

鲍振西 原为文化部图书馆司副司长。已离休。
李哲民 原为文化部图书馆司研究馆员。已退休。

白国庆 文化部计财司经营处处长。

(来稿时间: 1995—09—20. 编发者: 刘喜中.)