



普通高等教育“十一五”规划教材

古籍数字化

GUJI SHUZHUA

LILUN YU SHIJIAN

理论与实践

主编 毛建军

2 8 7 6 0
1 6 4 8 3

航空工业出版社

普通高等教育“十一五”规划教材

古籍数字化理论与实践

主 编 毛建军

航空工业出版社

北 京

内容提要

古典文献学以古籍为主要研究对象。随着计算机信息技术的快速发展,一种新的古籍组织和利用方式应运而生,这一新方式即古籍数字化。古籍数字化的出现意味着古典文献学有了新的研究对象,并为其研究提供了全新的研究手段和思维模式,必将给古典文献学研究带来新的研究视野。

本书以古典文献学为中心,试图从理论和实践层面分析古籍数字化的开发与建设问题,以此为古籍数字化基本理论的构建提供可行策略。构建古籍数字化系统理论可以为古籍数字化出版提供基础理论,开拓古籍整理和古典文献学专业的新视野,为文史研究者提供新思维和新工具。

图书在版编目(CIP)数据

古籍数字化理论与实践 / 毛建军主编. —北京: 航空工业出版社, 2009.4

ISBN 978-7-80243-245-1

I. 古… II. 毛… III. 数字技术—应用—古籍整理—研究 IV. G256.1-39

中国版本图书馆 CIP 数据核字 (2009) 第 027793 号

古籍数字化理论与实践 Guji Shuzihua Lilun Yu Shijian

航空工业出版社出版发行

(北京市安定门外小关东里 14 号 100029)

发行部电话: 010-64815615 010-64978486

北京市科星印刷有限责任公司印刷

全国各地新华书店经售

2009 年 4 月第 1 版

2009 年 4 月第 1 次印刷

开本: 787×960

1/16

印张: 10.75

字数: 199 千字

印数: 1—5000

定价: 28.00 元

编者的话

学术上的突破，往往依赖于新材料的发现和新手段的运用。近代新学术的兴起，多与新材料的发现和新手段的运用密切关联。计算机的运用和数字化资源的出现应该说是 21 世纪学术研究新手段运用的标志。计算机不仅在自然科学研究中大显身手，而且在社会科学研究中也发挥着日益重要的作用。

文献的载体就是记录文献的物质材料，是文献信息传递的物质基础和外化。我国传统的文献载体主要有甲骨、金石、简牍、缣帛、纸张、胶片和磁带。20 世纪中后期，随着计算机技术的进展，数字光盘以及数字化储存又成为一种全新的文献载体，出现了大量的数字化资源。从我国古代文献载体的变化与文化发展来看，每一次文献载体的转变必将引发一次新的文化和学术的繁荣。古籍数字化的出现不只是一次外在的文献载体转变，更是一次文化传播和学术研究的新突破。

古典文献学研究重在文献资料的搜集和整理。在计算机技术出现以前，资料的搜集和整理主要依靠手工来完成。但是，随着计算机和网络技术的发展以及古籍数字资源的大量开发，古典文献学研究的方法和手段出现了新的转机。古籍电子索引、古籍全文检索系统的强大检索功能，可以大大节省资料的搜集和整理时间，为研究者提供更大范围的资料基础，缩短研究周期，提高学术研究的效率和精确度。另外，文献整理中的校勘、辑佚以及注释也都可以利用计算机来完成。

中国古典文献学的研究就是在西方文献学的影响下并在版本学、目录学、校勘学以及传统考据学的基础上建立起来的，至今这两种思维模式还贯穿在古典文献学研究之中。以计算机和网络为标志的现代信息技术必将带来古典文献学研究模式的转换。计算机和网络的普及，古籍数据积累和数据交换的加速增长，尤其是以古籍数字化为代表的全新数据模式，将全方位地改变我们的知识获取和思维范式。

最后需要指出的是，古籍数字化还处在早期阶段，技术问题仍然还是古籍数字化的焦点。但是随着古籍数字化资源的丰富和开发经验的积累以及古籍数字化理论的成熟，古籍数字化必将为古典文献学开拓新的研究领域，建立新的学科体系，即古籍电子文献学。古籍电子文献学的基本框架应包括：

古籍数字化资源生产学：研究古籍数字化编辑过程及其规律的科学。古籍数字化是一项艰巨而复杂的系统工程，从组织选题、文字录入、审校标引、版面分析、特征抽取到系统配置都是一个全新的挑战。古籍数字化编辑过程也具

有其独特的规律。因此，研究这些规律将有助于古籍数字化的进一步研发。

古籍数字化资源流通学：研究古籍数字化资源的出版、发行、典藏及其规律的科学。古籍数字资源的出版发行目前很是混乱，除了数字古籍光盘一般由出版发行单位遵照电子出版的原则外，多数网络古籍出版还处于一种无序状态。古籍数字化资源的出版应该具有其自身的运行原则和规律。另外，图书馆和科研院所引进的古籍数字化资源是馆藏资源的一部分，也应该具有数字典藏的基本规律。

古籍数字化资源整序学：研究古籍数字化资源的索引、导航和链接原理及方法的科学。古籍数字化最欢迎的功能就是知识的索引、导航和链接即超文本系统（Hypertext）。超文本系统是数字化古籍的灵魂。数字化古籍的超文本系统一般是指：简繁字关联、联机字典、Text 文本关联以及知识关联。例如文信传文史研究院“汉文史数据库”设计有《人名大典》和《中国历史日历》。所有这些超文本系统都需要古籍整理专家的参与设计，每一个系统数据库都具有其基本的规律和特色。

古籍数字化资源利用学：研究如何有效利用古籍数字化资源的科学。古籍数字化资源的开发最终是要为广大用户使用，因此研究如何有效地利用古籍数字化资源就成为古籍电子文献学中的一个重要的领域。古籍数字化资源利用应包含两个方面，一是图书馆有效服务系统，二是个人对古籍数字化资源的组织和充分利用。

毛建军

2008 年 12 月 1 日

目 录

第一章 古籍数字化概说	1
第一节 古籍数字化概念的形成	1
一、20 世纪 80 年代	1
二、20 世纪 90 年代	3
三、21 世纪初	4
第二节 古籍数字化的基本性质	6
一、古籍数字化的基本要素	6
二、古籍数字化的基本原则	7
三、古籍数字化的意义	7
第三节 古籍数字化的开发层次	8
一、编制古籍电子索引	9
二、建立古籍书目数据库	9
三、古籍原文图像复制	10
四、汇编古籍电子丛书	10
五、古籍标点今译与普及	10
第四节 古籍数字化资源的流通	11
一、古籍光盘的出版	11
二、古籍网络的出版	13
第二章 古籍索引的电子化	15
第一节 古籍索引电子化	15
第二节 古籍索引编制的回顾	18
第三节 古籍索引电子化实践	21
第四节 古籍电子索引的检索	23
一、古籍电子索引的分类	23
二、古籍图谱电子索引	26
三、智能型古籍电子索引	29
第五节 古籍电子索引的规范	29
一、两种《古今图书集成》电子版简介	30
二、两种《古今图书集成》电子版比较	30
三、古籍电子索引的理想模式与规范	31
第三章 古籍书目数据库的规范	34

第一节 古籍书目数据库的概念	34
第二节 中文古籍的计算机编目	35
第三节 古籍书目数据库的著录	38
第四节 古籍书目数据库的规范	40
一、古籍书目数据库的标准	40
二、古籍书目数据库的评价	42
三、古籍联合目录数据库	44
第四章 古籍书目数据库的利用	48
第一节 国外书目数据库的建设	48
第二节 台湾地区书目数据库的建设	52
第三节 大陆地区书目数据库的建设	54
第四节 专题古籍书目数据库	56
一、地方志书目数据库	56
二、家谱书目数据库	58
三、出土文献书目数据库	60
第五章 古籍全文数据库的实现	62
第一节 古籍全文数字化的实践	62
第二节 古籍全文数据库的类型	65
一、古籍全文数据库的存储方式	65
二、古籍全文数据库的类型	65
第三节 古籍全文数字化的实现	67
一、选题确定与版本选择	67
二、数字化的硬件与软件	68
三、古籍全文数字化的基本流程	70
第四节 《中国基本古籍库》	70
一、收录之广与版本之精	70
二、全新便捷的分类方法	72
三、多功能的检索系统	73
四、完善的阅读编辑功能	74
第六章 古籍全文数据库的利用	76
第一节 国外古籍全文数据库的建设	76
第二节 台湾与香港地区古籍全文数据库的建设	80
第三节 大陆地区古籍全文数据库的建设	85
第四节 专题古籍全文数据库	91

目 录

一、地方志全文数据库	91
二、出土文献全文数据库	93
三、中医古籍全文数据库	95
附 录	99
附表一：古籍类光盘数字资源简表	99
附表二：古籍电子索引资源简表	110
附表三：古籍书目数据库一览	116
附表四：古籍全文数据库一览	127
附表五：爱如生数字化技术研究中心古籍数字精品列表	136
附表六：书同文数字化技术有限公司古籍数字精品列表	143
附表七：国学时代文化传播公司古籍数字精品列表	145
附表八：台湾汉珍数位图书公司古籍数字精品列表	147
参考文献	149
后记	163

第一章 古籍数字化概说

第一节 古籍数字化概念的形成

一、20 世纪 80 年代

古籍数字化与计算机信息技术的发展密切相关。1971 年台湾地区国科委马志钦教授提议利用计算机进行中文资料处理研究。不久众多大学和电子研究机构全力投入了这项极具潜力的领域。^①1972 年,台湾清华大学图书馆建立了计算机图书目录,开始了将计算机用于图书编目的工作。^②1984 年台湾地区“中央研究院(中研院)”历史语言研究所(史语所)开始实施开发“史籍自动化系统”。该计划由毛汉光教授倡议,谢清俊、管束东主持开发。^③在台湾,除了单位投资建设中文信息资料库外,部分个人也积极投入到这项极具挑战的工作中。1987 年,台湾师大中文系教授陈郁夫也开始尝试开发《红楼梦》、《水浒传》的全文检索系统。^④

由于大陆地区计算机事业起步较晚,在古籍整理与计算机结合的实践上也稍晚于台湾地区。改革开放后,我国在汉字处理、汉字检索系统及汉字终端设备等方面取得长足进展。计算机技术的发展为古籍整理与计算机结合提供了条件。1983 年,全国语言学学科规划会议提出,古籍整理应与计算机相结合,古籍整理研究手段和方法的现代化是完全可能的。随后,国内相继开始了古籍整理与计算机结合的实践。1983 年秋,江苏省“红楼梦电脑处理”课题启动。参加单位有镇江市科委、镇江市统计局电子计算站和江苏省计算中心。彭昆仑担任课题负责人。^⑤1984 年钱钟书先生提出,古典文献整理和研究应尽早与计算机结合。在钱钟书先生的倡导下,中国社会科学院文学研究所栾贵明、田奕等

① 叶晓珍《电脑摇身,古籍一变——访“中研院”谈古籍全文检索系统》,《国文天地》1990 年第 5 卷第 9 期,第 16~19 页。

② 程光《台湾数字图书馆的发展与启示》,《情报资料工作》2001 年第 6 期,第 28~29 页。

③ 叶晓珍《电脑摇身,古籍一变——访“中研院”谈古籍全文检索系统》,《国文天地》1990 年第 5 卷第 9 期,第 16~19 页。

④ 陈郁夫《“龙泉”初淬——“岭月”中英文全文处理检索系统开发记》,《国文天地》1990 年第 5 卷第 9 期,第 22~23 页。

⑤ 彭昆仑《闯入神奇的知识王国——“电脑红学”研究的甘苦谈》,《国文天地》1990 年第 5 卷第 9 期,第 34~37 页。

组成了研究小组,经过艰难的探索,“全汉字系统”、“《论语》数据库”、“《诸子集成》数据库”、“《全唐诗》检索系统”等古典文献数据库相继完成。^①

在古籍整理与计算机结合的早期实践中,台湾和大陆是主体,但国外也有中文古籍整理与计算机结合的实践活动,这些实践的经验对后期的古籍数字化工作产生了很大影响。如1978年美国人P-J-Ivanhoe运用计算机编制了朱熹《大学章句》索引、朱熹《中庸章句》索引、王阳明《传习录》索引、戴震《孟子字义疏证》索引等。^②1995年初,美国图书馆研究学会通过Internet在北京大学图书馆设立的工作站开始运行,向我国展示了联机环境下中文善本书的检索与编目,为我们提供了宝贵的经验。^③

古籍整理与计算机结合的实践也引起了学术界的思考。1984年8月20日第127期《古籍整理出版情况简报》刊发了栾贵明、李秦的《微电脑与古文獻研究》的文章,文章提出:“微电脑的应用,必将改变古文献的传统研究方式,达到空前的高效率”,“随着微型机数量的增加、功能发展以及分布的扩大,其信息的存储量会愈来愈多,并在一定范围,从一个地区到全国以及世界各地组成网络,形成一个巨大的资料库,所有信息资源便可共享。实现了这个目标,我国几千年来汗牛充栋而又星罗棋布的古文典籍,可尽行收入方寸之地,召之即来。使用微型机对这些古籍进行版本研究、文句校勘、文字订正、字义诠释、篇章会注、作品编年、古语今译,乃至标点、分段等等都将成为现实。”^④文章的作者将“古文献的各种索引”,看做是这个“巨大的资料库”的基石,这种思想在以后很长的一段时间里成为古籍整理工作者的行动指南。^⑤

1988年曹书杰在《古籍整理研究学刊》发表了《古籍整理与电子计算机应用研究的思考》一文,文章论述了古籍整理中的古籍今译、古籍注释、训诂、古籍校勘、古籍辑佚及古籍汇编的计算机辅助等基本问题,并提出了“机整”的概念,即电子计算机参与古籍整理研究领域的工作。文章认为“目前开展的机整工作还未超出‘内容结构式’的范围,多限于一部书的单一研究,尚属电子计算机在古籍文献领域中的低层次应用研究,所以应用计算机整理研究古籍的完整工作程序和方法目前还很难说清楚,还有待于实际工作的深入开展和总

① 田奕《古籍整理与研究的电脑化》,《中国文化》1994年第1期,第85~89页。

② 陈东辉《二十世纪古籍索引编制概述》,《文献》1998年第2期,第69~78页。

③ 张琪玉《古籍索引的一个范例——介绍〈古今图书集成〉电子版的索引数据库》,《图书馆杂志》2000年第5期,第48~49页。

④ 《古籍整理出版情况简报》第127期。

⑤ 陈琦等《武汉图书馆馆藏古籍善本数据库的建设与知识型数据库的实现》,《图书馆论丛》2003年第4期,第28~30页。

结。”^①因此，文章针对古籍整理与计算机应用提出了具体的工作步骤：确定整理对象、下达工作指标、设计程序、标识文献、建立数据库、综合运行。最后作者对存在的现实困难提出了具体建议。

二、20 世纪 90 年代

20 世纪 90 年代，随着计算机用于古籍整理实践的丰富和大量数据库资源的开发，学术界开始出现古籍电脑化、古籍自动化、古籍电子化以及古籍数字化等术语。这些术语或在论文标题中出现，或在行文中出现，但都没有具体内涵的界定。这种现象反映了当时学术界对古籍数字化概念的模糊性。

1990 年台湾《国文天地》（第 5 卷，第 9 期）为推动古籍整理中电脑的运用，特意推出专栏“科技新贵与古籍佳人的结合——中国古籍电脑化”，尽管专栏中并没有一篇文章给“古籍电脑化”一个确切的含义，但却是第一次使用了“古籍电脑化”的术语。1994 年田奕在《古籍整理与研究的电脑化》^②一文中也使用了“古籍电脑化”一词。1995 年台湾元智工学院教师罗凤珠开发出了“唐诗多媒体网络系统”，系统收入约 100 万字资料，包括四个子系统：诗文分析资料库、研究论著资料库、D.I.Y. (Do It Yourself) 个人工作平台和网络教学讨论区。该系统具有作者检索、诗题检索、关键词检索和诗句检索功能。同时，罗凤珠还提出了古籍自动化和文史资料自动化的术语。^③同时，在台湾还出现了“古籍全文资料库”^④、“古籍文献之资讯化”^⑤、“珍藏文献数位化”^⑥等术语。

“古籍数字化”一词最早出现在大陆。1996 年上海图书馆启动重点科研项目——古籍善本全文光盘，计划将馆藏古籍善本全部数字化。上海图书馆研制的中国古籍善本查阅系统，将馆藏善本古籍的全文数据以图像形式录入光盘，并对标引、检索、查阅等功能进行了开发，为古籍善本的数字化作了尝

① 曹书杰《古籍整理与电子计算机应用研究的思考》，《古籍整理研究学刊》1988 年第 1 期，第 44~49 页。

② 田奕《古籍整理与研究的电脑化》，《中国文化》1994 年第 1 期，第 85~89 页。

③ 罗凤珠《携手同行古籍自动化的路：不废江河万古流——D.I.Y. 唐诗多媒体网络系统架构设计》，《国文天地》1995 年第 2 卷第 1 期第 105~111 页。

④ 谢清俊，林晰《“中央研究院”古籍全文资料库的发展概要（上）》，《计算中心通讯》1995 年第 7 期，第 130~136 页。

⑤ 黄沛荣《古籍文献资讯化之现况与检讨》，《国家图书馆馆刊》1997 年第 6 期，第 71~93 页。

⑥ 薛理桂《珍藏文献数位化之发展现况与展望》，《国立“中央图书馆”台湾分馆馆刊》1997 年第 9 期，第 10~21 页。

试。^①在开发建设“古籍善本全文光盘”的过程中,上海图书馆的工作人员刘炜、陈秉仁分别撰文详细介绍了上海图书馆开发建设“古籍善本全文光盘”的过程,并对“古籍影像光盘制作及检索系统”的结构、功能、配置作了全面描述。他们分别在文中提出了“古籍数字化”^②和“古籍善本数字化”^③的术语。20世纪90年代国内各大图书馆和科研院所都积极参与古籍数字化工作。1998年底由上海世纪出版集团上海人民出版社、香港迪志文化出版有限公司推出的《文渊阁四库全书》电子版,无论是从技术上还是从市场上来讲,堪称这一阶段古籍数字化的杰作。面对国内如此丰硕的古籍数字化成果,《中国典籍与文化》开辟了“数字古籍”专栏,专题讨论古籍的数字化问题。其中,陈洪澜在《中国古籍电子化发展趋势及其问题》^④一文中使用了“古籍电子化”的术语。事实上这个阶段“古籍电子化”、“古籍数字化”、“古籍的数字化”是交替使用的。^⑤出现这种情况的原因是由于20世纪90年代“数字化”、“电子化”混用造成的。王冠中认为:数字化是一个制作过程,电子化是一个应用过程。所以古籍数字化就是古籍文献制作成数字成品的过程。而古籍的电子化则是古籍文献数字化后,被使用的过程。因此,综合而言“数字化”这个概念的表述相应要准确一些。^⑥而事实证明,进入21世纪以后,学术界广泛接受了“古籍数字化”这一术语。

三、21世纪初

明确的古籍数字化(电子化)概念的出现是在21世纪初。2000年李运富在《谈古籍电子版的保真原则和整理原则》^⑦一文中首次界定了“古籍电子化”

① 陈秉仁《古籍善本数字化的尝试:中国古籍善本查阅系统述略》,《现代图书情报技术》1998年第1期,第22~25页,第45页。

② 刘炜《上海图书馆古籍数字化的初步尝试》,《图书馆杂志》1997年第4期,第33~34页。

③ 陈秉仁《古籍善本数字化的尝试:中国古籍善本查阅系统述略》,《现代图书情报技术》1998年第1期,第22~25页,第45页。

④ 陈洪澜,《中国古籍电子化发展趋势及其问题》,《中国典籍与文化》1998年第4期,第121~126页。

⑤ 如史睿《论中国古籍的数字化与人文学术研究》,《北京图书馆馆刊》1999年第2期,第28~35页;郑辉《海峡两岸古籍数字化的发展现状及展望》,《信息窗》1999年第3期,第47~49页;孙建越《中华古籍的数字化<文渊阁四库全书>电子版》,《中国电子出版》1999年第4期,第17~18页;周毅《中国古籍电子化的里程碑:写在文渊阁<四库全书>电子版面世之际》,《文汇报》1999年6月26日第6版。

⑥ 王冠中《中文古籍数字化成果与展望》,东北师范大学硕士论文,2005年,第3~4页。

⑦ 李运富,《谈古籍电子版的保真原则和整理原则》,《古籍整理研究学刊》2000年第1期,第1~7页。

的概念:所谓古籍电子化,是指利用现代信息技术,将历来以抄写本、刻铸本、雕版、活字版、套版及铅字印刷等方式所呈现的古代文献,转化为电子媒体的形式。此概念着眼于技术转化角度,缺少古籍数字化目的的界定。彭江岸对此作了补充:“古籍数字化就是利用数字技术将古籍的有关信息转换成数字信息,存贮在计算机上,从而达到使用和保护古籍目的。”^①然而需要指出的是,古籍数字化事实上是一项系统的工作过程,以上两个概念缺少关于这方面的描述。2001年乔红霞又对古籍数字化作了进一步的归纳:“我国的古籍数据库建设即古籍数字化工作伴随着我国数字图书馆发展的步伐,经历了认识、探索、研制三个阶段。在这三个阶段的探索中人们把古籍的数字化归纳为这样一个概念即利用多媒体技术、数据库技术、数据压缩技术、光盘存储技术、网络传输技术等手段把馆藏印刷型文献、缩微型文献、音像型文献等传统介质文献转化为数字化、电子化的光盘或网络信息的工作。”^②但这个概念又略显复杂,因此,张雪梅又作了简化:“古籍数字化就是采用计算机技术,对古籍文献进行加工、处理,制成古籍文献书目数据库和古籍全文数据库,用以揭示古籍文献中所蕴涵的极其丰富的信息资源,从而达到使用和保护古籍的目的。”^③不过,张雪梅的定义又缺少对古籍数字化动态过程的描述。其后刘琳、吴洪泽在《古籍整理学》第九章“古籍整理手段的现代化”中对古籍数字化作了动态描述:“所谓古籍数字化,就是将古代典籍中以文字符号记录的信息输入计算机,从而实现了整理、存储、传输、检索等手段的计算机化。”^④段泽勇、李弘毅也认为古籍数字化是一个动态过程:“古籍数字化则就是把人们几千年来常用的文字符号转化成能为计算机识别的数字符号的一种过程。”^⑤

综上,我们可给古籍数字化作如下界定:古籍数字化就是从利用和保护古籍的目的出发,采用计算机技术,将常见的语言文字或图形符号转化为能被计算机识别的数字符号,从而制成古籍电子索引、古籍书目数据库和古籍全文数据库,用以揭示古籍文献信息资源的一项系统工作。需要指出的是,古籍数字化还处在理论和实践的探索之中,对古籍数字化概念存在不同的认识应是古籍数字化发展中所必须经历的过程。

① 彭江岸《论古籍的数字化》,《河南图书馆学刊》2000年第2期,第63~65页。

② 乔红霞《关于古籍全文数据库建设工作的思考》,《河南图书馆学刊》2001年第4期,第58~60页。

③ 张雪梅《古籍数字化与文献信息资源共享》,《天津工业大学学报(社科版)》2002年第3期,第85~86页。

④ 刘琳,吴洪泽《古籍整理学》,四川大学出版社,2003年版,第335页。

⑤ 段泽勇,李弘毅《古籍数字化的回顾与展望》,《图书馆理论与实践》2004年第2期,第37~39页。

第二节 古籍数字化的基本性质

关于古籍数字化的基本性质,学术界目前已经达成共识。国家图书馆古籍特藏部史睿研究员^①、中国社科院郑永晓教授^②、武汉大学信息管理系李明杰博士^③均认为古籍数字化属于古籍整理的范畴,古籍数字化是计算机用于古籍整理工作的新实践。“古籍数字化是保护和传承古文献最有效的形式,也是古籍整理工作的必然趋势。”^④

古籍数字化是对已存古籍的再现和加工,属于古籍整理的范畴,是古籍整理的一部分。古籍数字化的最终结果是将古籍资源制成古籍电子索引、古籍书目数据库和古籍全文数据库,从而达到利用和保护古籍的目的。古籍数字化不仅涉及传统古籍整理的知识和技能,诸如版本、目录和校勘等,而且还涵盖当代最先进的科学技术即信息技术,古籍数字化的研究也具有古籍整理学和数字技术相结合的跨学科特征。古籍数字化是21世纪古籍整理的主流,代表着未来古籍整理的发展方向。

一、古籍数字化的基本要素

古籍数字化是一项复杂而庞大的系统工程,需要具备以下几个基本要素。

(1) 加工对象:古籍。古籍数字化的首要问题是选定要进行数字化加工的对象即古籍。古籍的选择最关键的是确定一个优秀的版本,这样可以为读者或研究人员提供更为可靠的材料。

(2) 加工工具:软件。强大的软件支持是古籍数字化的根本保证。近几年,许多软件开发公司积极参与古籍数字化工作,开发了很多古籍数字化的程序。如中易公司开发的《中易汉神e》应用平台系统、北大方正集团公司开发的“DABSI-方正德赛古籍数字化系统”。

(3) 著录标准:机读目录格式。统一的、行之有效的机读目录格式是建立古籍数据库首先要解决的问题。20世纪80年代颁布、90年代重新修订的《中国文献编目规则·古籍著录规则》在题名、编著人、卷册页码、刻工、文献形

① 史睿《论中国古籍的数字化与人文学术研究》,《北京图书馆馆刊》1999年第2期,第28~35页。

② 郑永晓《古籍整理与古籍的数字化》,《古籍整理出版情况简报》2000年第5期。

③ 李明杰《中文古籍数字化基本理论问题刍议》,《图书馆论坛》2005年第5期,第97~100页。

④ 曹亦冰《从中国大陆当代古籍整理的现状看其类别、方式方法及走向》,《古籍整理研究学刊》2005年第1期,第1~7页。

态等条目上的著录都未完善,无法真正实现资源共享。

(4) 实施者:古籍整理专家和数字化技术专家。在古籍数字化过程中,古籍是内容、是核心,数字化是技术、是形式。内容决定形式,形式服务于内容,因此,古籍数字化必须坚持古籍和技术专家的协调配合,才能达到内容与形式的完美结合。^①

5. 管理协调者:统一管理机构。古籍数字化必须要求成立一个统一规划的部门,负责古籍数字化的选题协调、出版规划以及审核工作,避免重复开发。

二、古籍数字化的基本原则

古籍数字化是一项系统工作,必须要在一定的基本原则的指导下才能健康、有效地运作。为此可提出以下几个原则作为参考。

(1) 古籍数字化必须遵守古籍整理的基本原则。古籍数字化加工的对象是古籍,属于古籍整理的范畴,因此必须首先遵循古籍整理的基本原则。

(2) 古籍数字化必须遵守保护与利用相结合的原则。通过数字化可以实现古籍文物价值和史料价值的分离。通过古籍的数字化,一方面,可以大大提高古籍的利用效率,符合古籍利用的要求;另一方面,数字化古籍避免了对古籍原本的直接使用,符合古籍保存的要求。

(3) 古籍数字化必须遵守统一规划、系统开发的原则。古籍数字化建设需要各部门相互协调,相互配合,统一规划,统一指挥,避免重复浪费。成立古籍数字化领导小组,是实现古籍数字化资源互补的有效保证。

(4) 古籍数字化必须坚持统一著录标准的原则。编制标准的古籍目录是建设古籍机读目录和古籍数字化资源共享的基础。“因为只有经过深入标引和严格规范控制的数据库才能产生再生资源,而再生资源经过有效的排序和筛选,才能实现知识发现。”^②

三、古籍数字化的意义

古籍具有两个特殊的价值:一是史料价值,一是文物价值。史料价值属精神遗产,文物价值属于物质遗产。然而,由于古籍又具有非再生的特性,很多古籍由于缺少保护,已不能阅读。另外,部分古籍收藏单位将大量的古籍当做文物收藏起来,研究人员很难借阅得到,更有一些古籍至今还不被学术界所知。这就造成了古籍的利用与保护的矛盾。古籍数字化以后,这个矛盾便可迎刃而

① 史睿《数字化条件下古籍整理的基本问题(论纲)》,《文津流觞》第8期。

② 史睿《数字化条件下古籍整理的基本问题(论纲)》,《文津流觞》第8期。

解。“多年的实践证明,传统方式已无法圆满解决图书馆长期存在的古籍善本书保管与利用的矛盾,现代化技术的介入使我们看到了解决矛盾的契机。”^①古籍数字化只是对原古籍的再现和整理,它将原古籍的文物价值和文化价值剥离开来,一方面,实现了古籍的永久保存,另一方面也实现了古籍的文化价值,研究人员可以在不见原著的基础上进行科学研究。因此,“古籍数字化是缓解日益严重的古籍使用与保存这一特定矛盾的有效解决之道。”^②古籍数字化的意义不只在古籍的利用意义上,而且还表现在古籍的保护上,可以说古籍的数字化是一件“一劳永逸”的工程,因为,经过数字化的古籍不仅可以永久保存,更可以永久利用,而且其传播中国文化的速度更快,范围更广。

许逸民先生在谈到古籍数据库的优越性时说:“一是容量巨大,一是检索便捷多样,一是便于永久保存。这是古籍数据库的三大优势,其中不论哪一条都是书本无法望其项背的。”^③传统的古籍整理都是通过手工翻阅进行的,很多学者为此“皓首穷经”。但是,有了数字化古籍后,这种情况可能会彻底改变。更为重要的是,传统的手工翻阅在文献资料收集上可能存在难以避免的疏漏,从而会遗漏一些重要的信息资源。数字化古籍在文献检索方面具有速度快、准确度高的优势。例如,我们想查找《四库全书》中有关李白的资料,如果用手工翻阅可能需要一年的时间,而使用《四库全书》电子检索版,只需输入“李白、太白、青莲居士”等信息,就可在几秒中内检索出 11,709 条相关信息。“这种大容量、高速度的任意检索与模糊检索方式,十分有利于积极开发古籍中蕴藏的特有信息资源,使人们可以在最短的时间内查遍所有有关的古籍目录以及后人的研究成果,不仅可以省却我们无数爬梳资料的劳动,而且能很容易地得出我们所需要的分析结果。”^④另外,古籍数字化在文献辨伪、辑佚、编制索引等开发古籍中蕴藏的信息资源方面也具有手工检索所不具备的优势。

第三节 古籍数字化的开发层次

古籍数字化是古籍整理工作的数字化体现,那么古籍数字化需要进行哪些具体工作,即古籍数字化的开发层次问题也就显得十分重要。顾力仁^⑤研究员结合台湾地区“中文古籍善本书联合目录计划”将典籍数字化的工作归为书目

① 李为实《关于古籍数字化的思考》,《四川图书馆学报》2002年第3期,第75~78页。

② 彭江岸《论古籍的数字化》,《河南图书馆学刊》2000年第2期,第63~65页。

③ 许逸民《关于制作古籍数据库的几点想法》,《古籍整理出版情况简报》2003年第11期。

④ 李为实《关于古籍数字化的思考》,《四川图书馆学刊》2002年第3期,第75~78页。

⑤ 顾力仁《图书馆古籍文献整理》,《海峡两岸图书馆事业研讨会论文集》,台北:国立政治大学,1996年,第367~371页。

建档、影像处理、网络传输以及光碟系统等四个方面。卢锦堂^①教授在其著作中指出,台湾地区图书馆在中国古籍整理电脑化的过程中所采取的具体方式包括古籍机读编目、古籍全文资料库、善本全文影像系统以及古籍精华多媒体光碟。李国新^②教授在谈到古籍资源数字化面临的主要任务时指出,下一步古籍数字化的主要工作是大规模总结性古籍检索工具的数字化、总结性古籍书目数据库的建立、类书资源的数字化、普及性作品的多媒体化及计算机古籍整理通用系统的研制开发。楼宇烈教授曾从一名使用者的角度谈到,古籍数字化应建立三个层次的古文献资料库,即“版本库”、“专题库”和“普及库”。^③这些观点为古籍数字化工作指明了方向。我们认为古籍数字化具体应包括编制古籍电子索引、建立古籍书目数据库、古籍原文图像复制、汇编古籍电子丛书及古籍标点今译普及等五个层次的工作。

一、编制古籍电子索引

计算机用于古籍索引编制首先表现在中文文献数据的处理上。20 世纪 80 年代初到 90 年代初,古籍索引编制电子化进入了全新的发展阶段。计算机技术的成熟和汉字输入技术的完善使得古籍索引编制逐步向专题数据库和可检索型数据库方向发展。计算机技术的发展为古籍索引的电子化提供了条件。1984 年中国社会科学院文学研究所开始建设“《论语》数据库”、“《诸子集成》数据库”、“《全唐诗》检索系统”等数据库。这些数据库具有 7 种索引形式:单字索引、人名索引、地名索引、援引著作索引、专有名词索引、补遗索引、衍文索引。20 世纪 90 年代以来,古籍全文检索系统的开发与建设使得这一理想得到实现。北京大学的“《全唐诗》检索系统”、南京师范大学的“《全宋词》检索系统”等都是古籍电子索引的代表。

二、建立古籍书目数据库

古籍书目数据库建设是古籍数字化工作的重要内容。参与古籍书目数据库建设的单位主要是公共图书馆、高等院校和科研院所图书馆。古籍书目数据库建设的服务目的主要是反映馆藏古籍收藏情况、服务于教学和科研。古籍书目

① 卢锦堂《台湾地区图书馆中国古籍整理电脑化论述》,《21 世纪中文图书馆学术会议》,香港:香港中文大学,1999 年,第 4~8 页。

② 李国新《中国古籍资源数字化的进展与任务》,《大学图书馆学报》2002 年第 1 期,第 21~26 页,第 41 页。

③ 楼宇烈教授认为:“版本库”主要是为了保护古籍;“专题库”主要是为专业研究提供古籍资料;“普及库”主要为一般读者提供最基本的古籍读本。(楼宇烈《汉文化资料库不同层次的实用要求》,《古籍整理出版情况简报》2003 年第 8 期)。