

我国出土简帛文献电子索引建设研究述评

□ 上海海洋大学人文学院 楼 兰

摘要 本文结合我国出土简帛文献电子索引建设的具体实践,对其建设成果进行了梳理和总结。纵观历史,出土简帛文献电子索引的建设工作已经取得了不俗的成绩。但与简帛整理的丰富成果相比,电子索引建设工作仍显滞后,还存在一些共性问题,即收录材料完备性不够、资源重复建设、检索特色和检索功能有待进一步提升。

关键词 简帛 帛书 索引 电子索引

中图分类号 G271 文献标识码 A 收稿日期:2015-2-25

出土简帛文献是研究中国先秦、秦汉乃至魏晋时期的第一手材料。随着计算机技术的发展和普及,出土简帛文献的研究方法和手段得以拓展丰富,同时也为出土简帛文献电子索引的建设开辟了一片更加广阔的新天地。

一、出土简帛文献电子索引的概念

出土简帛文献电子索引是指利用计算机技术对已出土的简牍帛书编制文献索引。其内涵包括两个方面:一是用计算机编制的各种纸质形态的索引,二是用计算机表达的索引系统^{[1]28}。它涵盖面广,内涵丰富。从材料范围来看,包括现已出土的从战国到魏晋时期的全部简帛文献材料。据不完全统计,包括已经正式发表和尚未公开发表的简帛总数超过24万枚^{[2]47}。它们不仅数量庞大,而且涉及面广。已出土的简帛文献总体上可分为著作和文书两类。从内容上看,依《汉书·艺文志》的分类,已出土的简帛著作包括六艺、诸子、诗赋、兵书、算术、方技等^{[3]67-68}。而文书类则包括法律文书、占卜记录、遣策等等。因此,出土简帛文献电子索引涵盖政治、经济、军事、历史、社会生活等诸多领域,它为历史学、考古学、古文字学、文献学等多个学科提供便利的检索资料,并极大方便学者了解、掌握和利用古籍文献,有助于学者进行更深更广的研究。

二、出土简帛文献电子索引建设的回顾

1. 纸质形态的出土简帛文献电子索引。最早可以追溯到1995年由日本学者大庭脩先生编制的《居延汉简索引》。该书包括地名

索引、人名索引、官职索引、语句考证索引等,基本上涵盖了居延汉简的重要字词,是迄今所见有关居延汉简的第一个比较全面的索引,为我们检索居延汉简提供了方便^{[4]27}。但很遗憾,该书不是逐字索引,而且在人名条目的设立、编排等方面还有待完善。

2010年12月和2013年11月,西南大学张显成教授分别出版了《秦简逐字索引》和《楚简帛逐字索引》两部专著。《秦简逐字索引》一书将已全部刊布的秦简分五批编为逐字索引,具体是《睡虎地秦简逐字索引》《放马滩秦简逐字索引》《周家台秦简逐字索引》《龙岗秦简逐字索引》和《散见秦简逐字索引》。每一部分索引含检字表、正文和释文三部分,书末为总检字表。《楚简帛逐字索引》一书收录了《包山楚简》《郭店楚简》《望山楚简》《九店楚简》《曾侯乙墓竹简》《新蔡楚简》等多种楚简文献。该索引将上述楚简文献逐一进行校核,然后编制成释文逐字索引。使用者可以通过检索某字,得到楚地各种涉及此字的简帛材料。

除此之外,华东师范大学刘志基教授编著的《中国出土简帛文献引得综录·郭店楚简卷》于2012年11月出版。该书以公布的《郭店楚简》为材料范围,由释文、引得和检字三大部分构成。其中,释文和引得两部分都以楷字释文和简帛文真实原形对照呈现。该书的特色还在于以传统文本与网络检索资料互补的形式呈现。该书信息检索之《郭店楚简电子诂林》《郭店楚简电子类书》都在华东师范大学中国文字研究与应用中心创制的“文字网”关

要影响因素在于:(1)控制词表规模过于庞大,且控制词字符过多,通常在10个字符以上。(2)无效匹配现象严重。主要原因在于图书馆文献有多种语言表述的关键词,一旦对不同语言的控制词没有进行区分,这样在查找一种语言的文献时,系统无法进行正确的区分,常显示两种语言的文献。(3)编程语言效率过于低下。图书馆快速自动信息标引系统中标引工具选择采用的是PHP系统。这一系统虽然具有优势,但在数据处理效率方面则低于Web软件系统。(4)算法不科学。本系统读一条控制词与一条文献,并在处理后将此文献写回数据库,如此重复进行下一条控制词及文献的读取。很显然,这种数据重复读取的过程中,算法有不合理的地方。信息标引系统工作时间复杂度相当于 $m \times n$ (其中,m表示控制词条数,n则表示被标引文献记录条数),值得注意的是,计算机读取次数的增多将导致磁盘数据工作负荷加重,造成I/O吞吐量太大,降低了标引效率。

在上述信息标引技术影响因素中,无效匹配来源于重复标记。也就是说,后两点是信息处理速度缓慢的影响因素。因此,采用更合理的编程语言,并逐渐改进算法,可有效地降低系统数据

的吞吐量。因此,应优化数据处理方式,实现一次读取。实践证明,内存计算技术的引入,可提高其信息标引的处理效率。

近年来,内存计算技术被广泛应用于电子商务、企业管理等领域,其卓越的数据处理能力得到人们的认同。内存计算技术应用于图书馆管理工作大大提高了文献知识库的快速自动信息标引处理效率。但该技术对计算机硬件和软件具有较高要求,信息标引技术依然存在一定的问题,其中算法不科学和编程语言效率较低是其主要影响因素^{[5]55}。基于此,技术人员应对计算机数据库进行重新设计,满足庞大的数据业务的需求,使内存技术在图书馆的快速自动信息标引中起到积极作用,使图书馆信息处理业务能够处于先进行列,提高其处理效率。

参考文献

- [1]赵衍.基于内存计算的计算机快速自动信息标引技术[J].图书馆学研究,2013(19).
 - [2]郭嘉凯.内存计算技术:大数据应用引擎[J].软件和信息服务业,2013(7).
 - [3]奚晓音.基于内存计算技术的电信行业数据仓库系统构建探讨[J].电信科学,2012(2).
 - [4]罗保山.SAP HANA在信息处理领域应用的探讨[J].科技传播,2012(24).
 - [5]袁培森.基于内存计算的大规模图数据管理研究[J].华东师范大学学报,2014(5).
- ★作者曾金发为江西财经职业学院信息工程系副教授。

联书系》上呈现,这突破了传统文本检索的局限,提供了更为便捷的检索手段。

2.用计算机表达的出土简帛文献电子索引。(1)单机安装版。1984年,香港中文大学中国文化研究所编制了《汉达古籍资料库·出土竹简帛书文献第1辑》,由中文大学出版社出版。该资料库共收九种简帛文献,分别是《马王堆汉墓帛书(一)(三)(四)》、《银雀山汉墓竹简(一)》《居延汉简释文合校》《散见简牍合辑》、《武威汉简》《武威汉代医简》和《睡虎地秦墓竹简》。该资料库以电子形式出版,系统包括古文献资料库、文献用字统计及有关资料、电脑检索程序以及特定造字软件(补充倚天中文电脑系统未能提供之字形)^{[5]216}。安装“汉达古籍索引资料库”后,使用者就可同时通过电脑显示简帛图片和相关释文,释文可根据需要以直排或横排方式显示。检索方式便捷多样,支持单字、字符串及句式检索,方便查阅和研究,检索结果还可直接打印或存档。

2003年,华东师范大学刘志基教授开发了《战国楚文字数字化处理系统》,由上海教育出版社出版。该系统收录了战国楚系简帛文、铜器铭文、玺印文和货币文等四种文献。其中,楚系简帛文献包括《包山楚简》《长沙子弹库战国楚帛书研究》《郭店楚简》、《九店楚简》《曾侯乙墓竹简》《战国楚竹简汇编》《上海博物馆藏战国楚竹书(一)(二)》共八种材料,合计56689字。安装该系统后,使用者可以利用楚文字输入法,进行任意字、句的全文检索,任意字的频率数据检索和文献附着材料的各种属性检索,检索功能强大。

(2)网络检索版。相对单机版成果,网络版的出土简帛文献检索系统数量更多,形式也更多样。有以文献本体为主要检索对象的电子索引。台北“中央”研究院历史语言研究所现有“简帛金石资料库”和“汉代简牍数位典藏”两大数据库。“简帛金石资料库”收录的简帛材料以秦汉时期为主,战国时期的较少。具体包括《睡虎地秦墓竹简》《居延新简》《马王堆汉墓帛书》《江陵张家山汉简》《罗布淖尔汉简》等40余种简帛文献整理专书。该资料库采用台湾GAIS搜索引擎,使用台湾BIG5内码,分全文和书目两个分库,可实现释文、编号、所在图书页码等检索方式。“汉代简牍数位典藏”则收录了该所收藏的汉简,其中以《居延汉简》为主,另外还包括《罗布淖尔汉简》《敦煌小方盘城汉简》和《武威刺麻湾汉简》。该数据库可通过检索简号、品名、释文和遗址这四种方式来查阅相关汉简的基本数据、释文以及彩色照、红外线照等数字影像。此外,北京时代瀚堂科技有限公司开发了“瀚堂典藏古籍数据库”,是目前唯一采用国际通用的超大字符集进行加工校勘的古籍数据库。其中的简帛文献数据库,可实现快速全文检索和连续图文对照阅读,并支持复制、打印功能。值得一提的是,国内各大博物馆还开通了馆藏精品数据库查询,如国家博物馆馆藏精品查询,故宫博物院馆藏精品查询均可进行简帛图谱的检索查询^{[6]24}。

对单一简帛文献制作的电子索引,有香港中文大学图书馆研制的“郭店楚简资料库”“走马楼三国吴简·嘉禾吏民田家僮资料库”和台北“中央”研究院历史语言研究所研制的“居延汉简补编图像检索系统”。“郭店楚简资料库”是1999年与张光裕教授合作共同研制出来的。使用者可通过网络对张光裕教授的《郭店楚简研究·第一卷·文字编》进行全文检索,可从竹简编号、篇目名称或有关字词等方面进行检索,也可检索《郭店楚简》主要参考论著目

录。“走马楼三国吴简·嘉禾吏民田家僮资料库”是与张光裕教授、黎明钊教授共同制作的。使用者可通过出土编号、地名、人名、年代编号等内容作全文检索。“居延汉简补编图像检索系统”则把台北“中央”研究院收藏未发表1100多枚居延汉简加以考释,并以红外线结合电脑,处理成较清晰的图像,建立了图文检索系统。

还有以字词为主要检索对象的索引。武汉大学简帛研究中心和香港大学中文系合作开发了“中国古代简帛字形、辞例资料库”。其中,楚简帛书包括《郭店楚简》《清华大学藏简》《望山楚简》《上海博物馆藏战国楚竹书(一)~(五)》等十多种;秦汉简牍分别是《里耶秦简》和《张家山汉简》。该资料库于2007年11月在网络上启用,使用者可通过输入单字或单个偏旁实现字形检索。还有华东师范大学中国文字研究与应用中心开发的“战国楚简帛文字形体检索系统”。该系统是在2003年出版的《战国楚文字数字化处理系统》软件光盘的基础上审核、增补完成的,增加了新出版的《上海博物馆藏战国楚竹书(三)~(五)》以及《新蔡楚简》的材料。由于楚简帛字形纷繁复杂,用字通假混乱,该系统以现有考释研究为基础,目前只提供GBK编码范围内通用字头的检索。

三、结论与思考

纵观出土简帛文献电子索引的编制历史,电子索引的数量不在少数。但与简帛整理的丰富成果相比,检索编制工作相对滞后。上文提及的各种电子索引成果,虽然建成时间不一,收录材料范围不等,检索对象也有不同侧重,但总体而言,还存在一些共性问题。

1.收录材料完备性不够。目前,学术界消化现有简帛整理成果和编制电子索引的速度远远跟不上简帛出土的速度。20世纪70年代以来出土的一些资料直到目前尚有未纳入电子索引者,而新的大批资料又相继涌现。比如就秦简而言,张显成教授的《秦简逐字索引》所收秦简材料非常丰富,但近年发现的岳麓秦简尚未囊括其中。岳麓秦简于2007年由香港购回和捐赠。目前《岳麓书院藏秦简(一)~(三)》已由上海辞书出版社出版。这批竹简内容丰富,对于研究中国数学史、法律史以及秦代的历史、地理、习俗等都有重要的文献价值^{[7]5}。此外,三国两晋时期的简帛材料收录存在缺憾。在上文提及的电子索引中仅有“走马楼三国吴简·嘉禾吏民田家僮资料库”,其他诸如楼兰尼雅汉晋木简等还未制作电子索引。

2.检索特色有待加强,检索功能急需提升。经过百余年的发展,简帛学已成为一个综合性的多学科研究领域。使用出土简帛文献电子索引的研究者们,已不满足于一般的逐字全文检索,而是希望能借助先进的技术手段,完成更丰富多样的学术研究功能。在这一方面,华东师范大学刘志基教授做了尝试。其在网站推出的《郭店楚简电子诂林》《郭店楚简电子类书》检索系统,与图书《中国出土简帛文献引得综录》之《郭店楚简卷》关联使用,这突破了传统文本的篇幅限制,为科研成果补充了大量的背景资料,有利于引导古文字学者进行更深层次的研究。然而,限于目前出土简帛文献电子索引开发的深度不够,类似这样的索引还很少。此外,电子索引的检索功能还需进一步提升。比如香港中文大学图书馆研制的“郭店楚简资料库”,该网络索引系统对于郭店楚简的资料检索较为便利,但集外字尚只能以图片形式显示,不能不说是一种局限。当然,这是与电脑古文字处理的瓶颈问题相联系的。

3.资源重复建设,缺乏整体规划协调。出土文献在全国各地的图书馆、博物馆、档案馆、科研单位均有收藏,由于各地区和部门信

无线网络技术条件下的移动图书馆服务模式与发展路径

□ 西安铁路职业技术学院 胥大男

摘要 本文通过对移动图书馆及其服务模式的概念进行阐述,在结合其服务模式应用现状的基础上,对无线网络技术条件下移动图书馆服务模式的发展路径展开了深入研究。

关键词 无线网络技术 移动图书馆 服务模式

中图分类号 G271 文献标识码 A 收稿日期:2014-12-12

对无线网络技术,特别是移动通信技术在图书馆及其服务模式中的应用进行研究一方面可以进一步扩大移动图书馆的服务范围和服务质量;另一方面,对于各类信息在社会生产、生活领域的传播与利用效率的提高也具有重要的促进作用。通过对移动图书馆服务模式的应用现状进行分析,并结合移动图书馆的特点及其服务模式的概念,进而为其服务模式的创新与发展提出了合理的意见和建议。

一、移动图书馆简述

1.移动图书馆及其服务模式的概念。所谓移动图书馆是指以先进的无线网络技术为依托,并以国际互联网技术和多媒体技术为支撑,在打破时间与空间限制的基础上,通过借助手机、掌上电脑等移动设备从而以便捷的方式进行图书信息的查询、获取和浏览的一项新型图书馆信息服务^[1]。而移动图书馆的服务模式是指随着无线网络技术的不断发展和完善,移动图书馆为满足用户信息需求而采取的不同类型的移动服务方式。

2.移动图书馆的特点。就现阶段而言,移动图书馆主要具备如下几方面特点:(1)移动性。所谓移动性是指移动图书馆各项服务的使用不受空间的限制,从而使用户通过手机终端,借助无线网络便可以对其所需信息进行全面浏览。(2)全天候。全天候是指移动图书馆的使用并不受时间的限制,以24小时不间断地服务模式为满足用户对各类信息的需求,此特点也是移动图书馆区别于传统图书馆的主要特点之一。传统形式下的图书馆,大都遵循早九晚五的经营模式,其营业时间有限,这便使得营业时间外的用户对图书资料的相关需求无法得到满足。而移动图书馆则以其全天候服务的特点,通过借助用户的移动设备终端,通过无线网络

技术传递信息,进而实时满足用户对图书资源的需求。(3)个性化。所谓个性化是指用户可以根据其相关需求为其自身定制信息服务板块^[2]。移动图书馆以个性化的特点为用户提供符合其自身需求的服务。例如,将社会各个生产、生活领域的图书进行划分,从而使用户可以根据自身需要有针对性地查找到相关服务,从而满足用户对于某一信息的需求。

二、移动图书馆服务模式的应用现状

1.移动图书馆无线门户网站。作为图书馆Web门户网站的延伸,移动图书馆无线门户网站将移动通信网络同Internet进行有机地结合,进而使用户可以使用终端(手机、笔记本电脑、平板电脑)通过Web网站获取图书馆相关的服务^[3]。就现阶段而言,移动图书馆的无线门户网站主要为用户提供的图书馆服务内容包括了馆藏目录检索、图书馆新闻公告、用户借阅信息的查询、账号或借阅卡挂失、预约信息与服务时间查询、图书馆馆藏分布查询以及借阅规则与培训课程预订和新书通报等^[4]。虽然近年来,无线网络和移动信息技术的发展较为迅速,3G网络的普及使得无线网络环境得到了良好改善,移动图书馆的各项信息服务得到了大幅提升,但4G网络的成功推行又对移动图书馆的服务质量和速度提出了更高的要求。目前,4G网络仍未在我国全面普及,因此移动图书馆的无线门户网站仍需在现有的基础上进行改进。

2.短信服务。作为移动图书馆最早开展的服务项目之一,短信服务以其简单、快捷地服务方式为用户提供着图书馆的各种服务。所谓短信服务是指用户通过手机终端向移动图书馆发送指令后而获取的相关服务,短信的内容以短小精炼的数字或文字为主,服务在体现其自身及时性的同时,也较好地体现出了其与用

息互不沟通,出土文献数据库的建设出现了各自为战的局面^[5]。出土简帛文献电子索引的建设也是如此。以郭店楚简为例,郭店楚简于1993年10月在湖北荆门郭店村出土,竹简内容包含儒道两方面著作。自其出土以来,便引起了学术界的强烈关注。为方便学术研究,国内图书馆和科研机构也为其相继开发和建设了大小不一的电子索引。从上文可知,现在包含郭店楚简这一材料的电子索引就有四五个之多。这种重复开发建设,造成了人力、物力和财力的浪费。不仅如此,由于各个索引之间的功能区别并不明显,使用者在选择使用电子索引上也存在困境。目前,还没有相关部门联合出土简帛文献电子索引的开发单位共同沟通协商,进行有效管理和合理调控。

出土简帛文献电子索引的建设工作是简帛研究的必然趋势,也是古籍数字化建设的重要组成部分。出土简帛文献电子索引的建设是一项系统工程,仅凭单个科研机构或研发企业的力量很难

完成。基于此,政府有关部门应该在出土简帛文献电子索引的建设中发挥重要的推动作用,加强全局规划,统一协调电子索引的编制工作,并联合多方专家学者不断吸收考释成果,完善信息检索技术,建立具有多学科特色的电子索引。相信随着更多简帛材料的出土和计算机技术的不断提高,出土简帛文献电子索引将不断丰富其内涵和功能,为学术研究和科学创新提供良好的平台。

参考文献

- [1]毛建军.两种《古今图书集成》电子版的比较[J].图书馆理论与实践,2008(3).
 - [2]谢桂华.二十世纪简帛的发现与研究[J].历史研究,2003(6).
 - [3]刘钊.出土简帛的分类及其在历史文献学上的意义[J].厦门大学学报,2003(6).
 - [4]刘乐贤.读居延汉简人名索引[M]//简牍学研究(第二辑).兰州:甘肃人民出版社,2002.
 - [5]陈东辉.台港地区所编古籍索引综述[J].辞书研究,2005(3).
 - [6]毛建军.简帛文献数字化述论[J].兰台世界,2011(8).
 - [7]程薇.秦简绽放:十多批重要秦简的发现与影响[J].中国文化遗产,2013(2).
 - [8]梁松涛.浅析我国出土文献数据库建设的制约因素及对策[J].图书馆工作与研究,2008(4).
- ★本文为国家社科基金重大项目成果,课题号:10&ZD118。
★作者楼兰为上海海洋大学人文学院副教授,文学博士。