

图书馆特色数据库建设若干问题研究

李 昕

(江苏省淮安市行政学院 江苏淮安 223001)

摘 要 数字化时代的图书馆,把建设特色数据库作为迎接时代挑战的主要手段之一。特色数据库建立的目的是为了充分发挥图书馆特色化信息资源的功能,为教学科研服务,最大限度地满足用户的信息需求。特色数据库的建立可以增强图书馆在信息社会中的竞争能力。本文从特色数据库的概念和建库原则入手,分析了建设特色数据库的意义,对特色数据库建设过程中的结构设计、元数据设计、质量控制和版权问题作了一些探讨。

关键词 图书馆 特色数据库 信息资源 元数据 版权 建设

中图分类号: G251 **文献标识码**: A **DOI**: 10.3772/j.issn.1674-1544.2009.02.000

21世纪是信息化的时代,数据库建设已成为评价图书馆信息资源建设的重要指标之一。近年来,很多图书馆都加快了数据库建设的步伐,数据库建设总体水平呈明显上升趋势。遵循保证重点、突出特色的资源建设基本原则,图书馆数据库建设也应像传统图书馆一样走特色馆藏建设的道路,建立具有自身特色的数字化信息收藏体系,形成自身鲜明特色,发挥自身特点和优势,为科研和教学提供高层次、高效率的信息服务,只有这样才能在信息资源共享中体现优势和竞争力,在竞争激烈的信息社会中占有一席之地。

1 特色数据库概念及建库原则

1.1 特色数据库的概念

特色数据库是依托馆藏信息资源,针对用户的信息需求,对某一学科或某一专题有利用价值的信息进行收集、分析、评价、处理、存储,并按照一定标准和规范将本馆特色资源数字化,以满足用户个性化需求的信息资源库。

特色数据库建设是图书馆数字资源建设的重

要任务,经过多年的探索,图书馆特色库从书目数据库向内容数据库转变,从馆藏文献资源向网络资源链接,从各馆孤立建设转向多馆联合开发,从自给自足的资源服务方式拓展到网上专题资源的共享服务。当点击图书馆网站时,经常会看到“数字图书馆”、“特色馆藏”、“专题特色库”、“自建数据库”等栏目,能浏览到各式各样的专题数据库。如在CALIS项目的25个特色资源库中,敦煌学数据库、巴蜀文化数据库、蒙古学文献信息特色数据库、东北亚文献数据库、长江资源数据库、东南亚研究与华侨华人研究题录数据库6个数据库的建设与学校所处地理位置有关;钱学森数字图书馆特色库与学校传统有关;其他数据库则与学校的专业特色相关。这种资源分布方式充分发挥了各校的地域特色、专业特色及传统特色,有利于资源的保存和传播^[1]。

1.2 特色数据库建设原则

(1) 计划性原则。数据库建设要进行整体规划,采取合理布局、重点投资、分步实施、从易到难、讲究实效等措施,尽早使数据库投入实施运行。

(2) 实用性原则。数据库建设应从社会发展实际需要和“211工程”文献信息服务系统任务要求出发,有针对性地开发某一专题的信息资源。在格式上要向国外著名的网络信息服务系统学习,以国际、国家的标准为依据,建立兼容性强、权威性高的数据库,面向国内外用户服务。

(3) 完备性原则。数据库所收藏的数据,要保证本专题门类的完整性及文献信息类型与载体的相对完备,并注重收集本专业的电子出版物,包括网络化的文本、图形、声频、数据集和多媒体文献。

(4) 标准化原则。标准化对保证资源和服务的可使用性、互操作性和可持续性具有重要的意义。为了便于信息交换与资源共享,在数据库建设中必须根据国家有关文献著录和标引原则制定、执行统一的实施规范和细则。

(5) 资源共享原则。特色资源库是文献资源保障系统建设中的重要内容。因此必须坚持资源共享的原则。在数据库建设的同时,需制定相应的资源共享方案,允许网上用户使用,并建立高效、方便、经济的文献信息存取和传递系统,以充分发挥数据库的作用。

2 特色数据库建设的意义

网络化、数字化的发展不仅给图书馆的发展提供了新的机遇和更大的发展空间,同时也给图书馆界带来了激烈的竞争。一方面是来自外界的竞争,网络上的每一位用户都可能是高效的信息发布者,谁掌握了资源并使之获得效益,谁就具备了信息服务的竞争力;另一方面是内部的竞争,即并非所有的图书馆都能在网络上占有一席之地,成为竞争中的强者。图书馆进入网络后,如果没有特色数据库供用户访问和使用,它的信息服务很难具有权威性,无法形成品牌效应,大量无人问津的数据最终会成为网络上的冗余数据、垃圾数据。因此,利用馆藏特色,高质量地开发信息资源,创立自己的品牌数据库,是图书馆适应21世纪信息服务的需要,也是图书馆在数字环境下应对挑战的好方法。

(1) 开发馆藏信息资源。特色数据库是根据本—48—
万方数据

馆的馆藏特色、地方特色或者本校的重点专业,集中搜集各类带有指引性的专题文献而建立的数据库。建设特色数据库可以使分散零乱的特色文献资源得以系统化、有序化整理和深层次加工,将传统文献资源转变为数字化资源,并通过网络进行传播使用,使传统文献在网络环境下重新体现其知识价值。

(2) 优化网络资源。互联网的出现和迅速发展,给用户查找资料带来了极大的方便,同时由于互联网上的数字化信息资源通常具有分散性强、变动性大、信息质量良莠不齐的特点,为用户选择、利用信息资源带来了不便。高校图书馆通过建设专题特色数据库,可以使网络上的有关该专题的资源相对集中,同时通过专业人员对专题网络资源进行有效的组织、整序、筛选、重组后,达到了优化网络资源的目的,方便了用户查找资料,节省了用户的时间,最大限度地满足了用户的信息需求。

(3) 开展特色信息服务。特色服务是近年来图书馆在传统服务方式基础上探索的一种新的服务方式。建设特色数据库是高校图书馆开展特色信息服务、变传统低层次服务为主动知识增值服务的体现,是图书馆服务意识和观念的根本转变,也是图书馆今后发展品牌、突出馆藏特色、提高地位的必由之路。

(4) 实现信息资源共享。无论是与发达国家相比,还是与本地区社会经济发展对信息的需求相比,我国网上可用的数据库都十分缺乏,信息资源建设是增加互联网上中文资源的数量、填平数字鸿沟、提高民族文化竞争力的重要举措^[2]。目前,大多数图书馆购买的数据库由于受知识产权的限制,不能在各个图书馆之间共享,而高校图书馆建设自己的特色数据库,只要馆与馆之间有某种协议,就可以达到信息资源共享的目的。

3 特色数据库的相关设计

(1) 结构设计。特色数据库结构的设计是数字资源建设过程的基础。设计数据库结构时要注意框架的设计、检索字段的设计、资源互链的设计、跨库检索的设计和导航的设计^[3]。框架设计揭示数

数据库的整体结构,勾勒出网站的概貌;检索字段的设计要注意根据需要尽可能地设计详细,如设置建库代码、建库时间和点击率字段等,为以后的维护、工作量统计和信息利用率分析打下基础;资源导航根据资源库的特点可按会议、网站、期刊、机构、学者等设计,为读者浏览和检索奠定基础;资源互链的设计关系到实现字段与字段、字段与数据库、数据库与数据库之间的底层链接质量,在字段内容相同或相近的不同数据库中设计互链,可为读者提供一站式检索服务;跨库检索设计涉及在不同数据库中同时实现检索,在为不同数据库设计内容一致的检索字段时最好名称一致,同时制定相应的规范。

(2)元数据规范设计^[4]。元数据规范是描述一类资源的具体对象时所有规则的集合。不同类型的资源会有不同的元数据规范。元数据规范设计是数据标引加工的前提,也是数据库建设的核心部分,规范制定元数据方案是数据库建设的质量保障。目前,CALIS发布的描述性元数据规范有11种,即图片、音频资料、学位论文、拓片、网络资源、期刊论文、家谱资源、会议论文、古籍、电子图书和地方志资源描述元数据规范及其著录规则。一般情况下,各类型资源对象的描述元数据的基本框架都由核心元素、资源类型核心元素和个别元素3部分组成。核心元素在各类资源对象中通用。在元数据的通用性方面,DC(Dublin Core)是最能满足通用性的元数据,因此核心元素复用DC的几个基本元素,支持通用的检索工具。资源类型核心元素根据资源对象特点、参照其他元数据标准制定,在同一类型资源对象中通用。不同的资源类型可有不同的资源类型核心元素设置及其语义定义,要求体现该资源类型的共同特性。个别元素以某种特定的资源对象为基础制定,仅适用于这类对象,不用于交换。应用时仅要求该种对象遵守其定义。

(3)元数据模板及标引字段设计^[5]。文献标引建库时是依据文献资源类型的元数据字段进行加工标引,如果按照数据库子库分类结构去建立元数据标引字段,就会产生诸多问题。由于同一子库的文献可能包含多种文献类型,而不同子库的文献又可能包含相同的文献类,因而,在一个子库

万方数据

里,若将数个文献类型的元数据字段合在一起,模板标引字段就会过长,标引时繁琐不便,效率过低,错误过多,而在一个子库里若按所包含文献类的元数据建立多个元数据模板,又势必会同其他子库相同文献类型的元数据模板重复,造成标引的混乱。实际上,读者在检索文献时,文献的检索只与标引时建立的数据联有关,用户在检索界面看到的子库划分只是导航数据关联的检索途径。因此,我们对资源进行标引时,摒弃了数据库子库概念的束缚,按照不同文献类型元数据建立不同的元数据标引模板,然后按不同元数据模板分配相应的数据项目,对数据进行标引、分类。对于子库与子库间内涵相关的文献建立数据库间的关联检索,即超链。

(4)质量控制。质量控制应贯穿建库的全过程。数据的准确性、完整性和规范性是影响建库质量的最重要因素,也是人为出错因素最高的环节。数据质量是数据库质量的核心,它直接影响数据库的检索性能。为此要控制好3个方面:其一,标词准确性控制。特色数据库元数据的标引虽然事先已制订了相关元数据标准,但主题词、关键词、摘要、分类号、载体形态等相关注册项的正确选择和描述仍存在人为差异的因素,一些特殊文献的标引描述还处于摸索尝试阶段,为此把握标引词的准确性和规范性尤其重要。其二,标引工作规范控制。标引工作要形成规范制度,对文献标引的词表选用、文献标引范围、标引类型、标引方式、标引深度等明确指示,使之有章可循,按章操作;建立严格的校对制度,把好校对关。其三,标引人员的管理和技术控制。要求标引人员要有责任心、有敬业精神和广博的知识以及扎实的标引技术,能及时解决标引过程中出现的疑难问题。

4 版权问题

版权保护一直困扰着数字资源的建设工作,这个问题不仅令数据库的建设步伐放缓,也让一些图书馆在进行数字资源的建设实践中遭遇到法律尴尬。特色数据库信息资源来源涉及以下3个方面:馆藏信息资源、本馆已购买的电子全文数据

库和网络资源。这里仅从特色数据库建设利用信息资源所应涉及的一些版权问题作一个分析了解。

4.1 馆藏信息资源数字化中的版权

“数字化”就是通过计算机技术把一定形式的信息,如文字、图像、声音等输入计算机系统并转换成二进制编码,再进行组织、加工、存储,采用数字传输技术传送,然后根据需要再把这些数字化的信息还原成原来的信息形式。数字图书馆建设特色库中将馆藏文献数字化即通过录入、扫描、手写识别、语音识别等技术手段将馆藏资源进行数字化。对于数字化行为的法律定性,有专家认为馆藏文献数字化是一种复制行为^[6]。从技术角度来说,数字资源的基础性建设较多是对传统文献的数字化。一般的看法是,数字化过程只是对文献资料的形式转换,其间并没有产生新的知识价值,属于复制行为,因此应该参照我国《著作权法》中复制权的有关规定。合法复制主要有“合理使用不适用著作权法保护的文献、超出保护期限的文献等”。

图书馆的馆藏文献包括两个部分:一是公有领域的文献;二是非公有领域的文献。对这两种文献数字化所涉及的版权问题是不同的。根据我国《著作权法》规定,公有领域的作品包括超过保护期限(50年)的作品,如古籍、善本等,不在著作权保护的地域范围内的作品和不适用著作权保护的作品(如法律、法规、时事新闻等),这部分文献的复制权不再受著作权人的支配,也不再受保护,图书馆可以根据需要对其进行数字化而不会造成侵权,但必须尊重作者的署名权、修改权、保护作品的完整权,因为这些权利是永远受到保护的。而对于非公有领域的文献,在对其进行数字化时必须慎重处理,这些文献只有在“合理使用”的前提下才可以不经著作权人的许可,不向其支付报酬,但应当指明作者姓名、作品名称,并且不得侵犯著作权人依照本法享有的其他权利,否则将造成侵权。

4.2 数据库资源采集中的版权

图书馆在进行特色数据库的建设过程中,必然会采集现有已购买数据库的资源,尤其是电子全文数据库,以补充特色数据库的内容。数据库是

一种汇编作品,在对数据信息的汇编过程中实际上是实现了一次知识的增值过程,因此完全可以享有自主的知识产权。数据库可分为指南性数据库和源数据库两种类型。指南性数据库是指引导读者用户到另一信息源获取原文或其他章节,如自建馆藏书目数据库或报刊篇名数据库等;源数据库是指读者用户可通过此类数据库直接获取原始文献及其有关数据,如数值数据库、全文数据库、图像数据库等。

《与贸易有关的知识产权协议》第二条第二款规定:“数字或其他材料的汇编,无论采用机器可读形式还是其他形式,只要其内容的选择或安排构成智力创作,即应予以保护。”也就是说数据库可以作为汇编作品受到版权的保护^[7]。又如欧盟各国在1996年通过的《数据库法律保护的指令》中规定:凡在其内容的选择与编排方面,体现了作者自己的智力创作的数据库,均可依据《数据库法律保护的指令》获得版权保护。世界知识产权组织和世贸组织认为数据库如果符合独创性的标准,就可以受到保护。目前,我国对数据库的法律保护尚无单独、具体的规定。我国著作权法将那些汇集有著作权材料的数据库作为编辑作品加以保护,规定编辑作品的编辑人享有著作权,但行使著作权法时,不得侵犯原作品的著作权。而非独创性数据库则由于不具备独创性而得不到版权法的保护,主要采用反不正当竞争法、侵权法等给予法律救济。

4.3 收集网络资源中的版权

有人认为,网上作品没有著作权^[8]。但根据《世界知识产权组织版权条约》(WCT)第一条第四款的议定声明:“《伯尔尼公约》第九条所规定的复制权及其所允许的例外,完全适用于数字环境,尤其是以数字形式使用作品的情况”。WCT还指出,受保护作品通过网络进行计算机的信息传播时,必然会进入计算机内存,并通过显示屏显示出来,这就构成了“暂时性复制”行为^[9]。如果信息使用者通过网络把自己所需的著作权人的作品下载到本机存储,就构成了永久性复制行为。图书馆特色数据库建设的途径之一就是有针对性地对网上无序的专业信息资源进行搜集、筛选,然后下载到本机上

成了永久性复制行为。这些复制行为都需要征求作品所有人的许可,并根据实际情况支付一定的费用。此外,根据我国《著作权法》第二十二条第二款的规定,适当引用他人作品属于合理使用。如果超出适当引用范围,且信息本身又属于仍在著作权保护期内的作品,即使不以盈利为目的,仍可能构成侵权。因此下载、拷贝网络信息资源极有可能超出适当引用范围,若未经版权人的授权,就可能构成了非法复制甚至侵权。

由于网络信息资源具有丰富性、动态性和时效性等特点,网络上传输的信息都经过了数字化处理,以连串的“0”和“1”组成的代码来表示。图书馆在采集网络信息资源时,面对大量的信息,要确定哪些信息是受著作权保护是比较困难的。那么在特色数据库建设中又该如何避免对网络资源的侵权呢?对于本身就属于免费性质的网络资源,如一些“共享软件”等,图书馆可以对其网站信息加以归类、索引,只要不是出于商业目的,一般都不会产生知识产权纠纷,而对于享有著作权的作品,图书馆可以通过制作目录、索引或连接等间接方式,为读者提供查找全文的线索。制作网络导航条是一个较为妥善的办法,网络知识导航主要通过超文本链接方式,引导读者查阅位于不同服务器上的同类信息资源,大大提高了网络资源的利用效率,而且一般不会被认为是侵权行为。

特色数据库的建设体现的是其馆藏特色,是“以服务求支持,以贡献求发展”的宗旨,是深层次的专业化服务理念 and 具体行为,是为了将图书馆最具特色、最有价值的文献资源以浓缩和精彩的形式提供给读者,以满足网络环境下读者对文献资料多样性和个性化的需求。可以预见,通过特色数据库的建设,图书馆不仅能够更好地实现服务于社会的目的,同时对于提高图书馆的声誉和竞争力也具有重要的意义。

5 结 语

现代信息技术改变了人们生产、存储、描述、分配和获取信息的方式。从历史的角度来看,传统的信息载体和信息传播媒介与信息组织方式正逐渐被以微电子技术为核心的电子载体和电子传播

媒介等新的组织方式所代替。尤其是信息网络技术的广泛应用,使国际互联网成为世界上最大的信息集散地,促进了图书馆文献信息资源向网络信息资源转化。由于网络信息资源是采用数据形式表达的,其内容广泛、无序分散且多为人工语言,难以规范和结构化,而采用数据库方式组织,可以大大提高信息管理效率,降低网络数据传输的负载。1999年,在中国高等文献保障系统(简称CALLS)中心的领导下,由上海交通大学牵头负责的专题特色数据库工作正式启动。目前,该项目已完成的专题特色数据库共有25个,总数据量达280万条。网上访问人数已达几百万人次。2003年,CALLS组织了“十五”全国高校专题特色数据库建设的项目申报、立项评审,并且已经有75个项目获得立项资助,项目研究涉及61所高校。目前大部分项目正在进行中,其中有些项目已经投入使用,例如武汉大学的“长江资源数据库”、北京大学的“古文献资源库”,中国矿业大学的“矿业工程数据库”等^[10]。由此可见,特色专题数据库的使用在实际工作中发挥了巨大的作用。

参考文献

- [1] 魏欣,等. 高校图书馆数字化建设——浅论数字化特色资源的建设[J]. 图书馆工作与研究, 2005(4): 52-53.
- [2] 杨宁莉,等. 高校图书馆特色数字资源建设选题方案研究[J]. 大学图书馆学报, 2003(6): 56-58.
- [3] 朱惠灵,等. 关于高校图书馆特色馆藏建设的若干思考[J]. 农业图书情报学刊, 2004(11): 66-67.
- [4] 上海图书馆. 网络资源元数据规范[Z].
- [5] 吴旭. 数字资源元数据模板、标引与发布系统模块的研究与实现[J]. 情报学报, 2004(6): 709-714.
- [6] 丁小文. 馆藏文献数字化的知识产权风险与对策研究[J]. 图书情报知识, 2003(5): 2-4.
- [7] 王颖洁. 论数字图书馆建设中的知识产权问题[J]. 图书馆学刊, 2005(6): 48-50.
- [8] 李玉春. 图书馆特色数据库建与著作权题[J]. 情报科学, 2003(6): 27-28.
- [9] 吴颖. 数字图书馆资源建设中的法律问题研究[J]. 图书馆工作, 2004(3): 31-32.
- [10] 尚向平. 图书馆特色数据库建设中的问题与对策[J]. 理论界, 2007(10): 241-242.

A Probe into the Construction of Distinctive Database in Libraries

Li Xin

(Huai 'an Administrative College, Huai 'an 223001)

Abstract: Libraries in the digital age construct distinctive databases as one of the main measures to take challenges. The aim of building distinctive databases is to give full play to the functions of libraries in providing information resources, serving teaching and scientific research, and satisfying users' need to the biggest extent. The construction of distinctive databases can enhance libraries' competitive capabilities, consolidating libraries in schools of higher education in the information age. This article first analyzes the concepts and principles of distinctive databases, then elaborates on the meaning of distinctive database construction, and finally talks about the design of database structure, the design of metadata, quality control and copyright problems.

Keywords: library, distinctive database, information resources, metadata, copyright, construction

“建国 60 年科技资源管理理论与实践学术征文”启事

2009 年是新中国成立 60 周年,也是中国科技资源管理事业蓬勃发展的 60 年。60 年来,我国科技资源管理工作经历了从计划经济体制到市场经济体制的重大转变,经历了从科学的春天到自主创新时代的重要飞跃,积累了丰富的经验,取得了重大进展,为我国科技事业的发展提供了重要支撑。为了纪念我国科技资源管理事业 60 年的发展历程,系统总结科技资源管理经验,本刊特举办“建国 60 年科技资源管理理论与实践学术征文”活动。

1. 征文内容

(1) 与科技资源管理有关的重大历史事件及学术争鸣的回顾和分析。

(2) 科技资源管理方式的演变及其与经济、社会、科技发展关系的研究。

(3) 各行业、地区、企业、科研单位科技资源管理的实践经验总结。

(4) 科技资源管理与建设创新型城市、创新型国家之间的关系。

(5) 全球化背景下科技资源管理的新思想、新理论、新方法。

2. 征文篇幅以 5000 ~ 8000 字为宜。文章请按学术论文写作规范提供摘要和关键词。

3. 凡经过评审通过的论文将在《中国科技资源导刊》上刊登。优秀论文作者将给予相应的奖励,并将被邀请参加本刊 2009 年底在北京举办的科技资源管理学术研讨会。

4. 截稿时间 2009 年 9 月 30 日。

5. 来稿请发至《中国科技资源导刊》信箱 zgkjzydk@istic.ac.cn,以“征文 + 篇名”作为邮件的主题词。如果邮寄纸质版,请寄至:北京市三里河路 54 号《中国科技资源导刊》杂志社,邮政编码:100045。联系人:莫妍。电话:010-68514086。信封正面注名“征文”字样。