

● 郭志高

图书馆文献信息数据库的构建

ABSTRACT In building the library document information data base, a viewpoint that the information has its price must be established so as to cultivate the information market with effort; the protection of intellectual property rights should be put emphasis on; the input and fosterage from the government must be tried to gain; the planning guidance should be strengthened and the structure of library professional work should be well-regu-lated 4 refs

KEY WORDS Libraries document information Data base construction

CLASS NUMBER G356 1

随着我国综合经济实力的不断提高和计算机通讯技术的快速发展,图书馆用现代信息设备来处理文献信息已成为现实。图书馆计算机系统和网络发展的生命力主要在于信息资源的数量和质量,以及能为读者提供什么样的有使用价值的信息。图书馆在实现自动化管理的过程中数据库的建设至关重要。

1 数据库建设是基础工作

在支撑信息社会的计算机、通讯和信息3大支柱产业中,信息是核心。人们十分形象地把信息网络建设比作道路建设,把硬件设备比作运输工具。一个信息系统能够真正运转并发挥效益,仅有“路”和“车”是远远不够的,关键问题是要有“货”可运,货就是信息资源。在信息资源中,数据库又起着举足轻重的作用。

数据库是将信息按照一定的格式,一定的标准,加工处理成为有序的数据储存在计算机中。图书馆自动化系统的硬件、软件和通讯网络建设都可以通过引进、花钱购置来实现。而

文献信息数据库,特别是适合我国和本地区经济建设以及政治、科学、文化、教育事业发展的文献信息数据库,只能依靠各馆自身的力量,按一定的标准和规范,依次录入计算机。由此而建立起来的各种各样的数据库,提供给广大读者使用,为社会提供公共信息服务。这是一项长期的工作,是一个积累的过程,虽然它是一项投入大、技术性强、周期长、见效慢的基础工作,但若没有这项工作,图书馆自动化系统就会失去存在的意义。图书馆作为国家信息工程系统的重要组成部分,搞好文献信息数据库的建设有着责无旁贷的义务。

2 存在的问题

数据库的发展大致分为3个阶段:60年代初至70年代初为第1阶段,数据库的内容多以科技文献的题录和文摘为主;70年代中期至80年代中期为第2阶段,以商业、金融、事务为主要内容;80年代后期至今为第3阶段,以面向社会、家庭为主要内容^[1]。据1995年10月结束的我国以数据库为重点的信息

资源调查数据表明,我国已建成的数据库多为文献型的科技和工程数据库,以二次信息为主^[2]。创建于 80 年代末的我国公共图书馆自动化系统在 90 年代中期发展较快。在数据库建设中,具代表性并形成一定规模的有:北京图书馆的中国国家书目机读数据库、上海图书馆的社会科学期刊篇名数据库和广东中山图书馆的事实、全文数据库。但从我国目前的图书馆总体来看,普遍存在着规模小、类型单一、发展不平衡等现象,相当于国外 70 年代初期的水平,其主要问题有:

(1) 资金投入严重不足。目前在图书馆自动化系统建设中,一般只有硬件设备的专项投入,数据库的建设难以立项。各馆主要从常年事业经费中挤出一部分来支持数据库的建设和维护,而对于经费严重短缺的图书馆则只能是望库兴叹。重硬件建设、轻资源开发和数据库建设的现象在图书馆界极为普遍,由此造成数据库与硬件建设投入比例严重失调。数据库建设滞后是影响图书馆文献信息系统协调发展的根本原因。

(2) 数据库类型单一,规模小。图书馆数据库的建设应建立在自身文献优势的基础上。但目前建立的数据库类型主要是馆藏书目和二次文献信息。全文、图形图像、事实信息数据库不多。大多数馆数据库规模小,基本没有联网的价值。

(3) 标准化程度不高,影响到数据库的利用。主要表现在:格式不规范,一些数据不能按照 CNMARC 格式处理,造成数据接受和二次开发的困难;检索命令不规范,同一系统中有不同的检索命令,让使用者不适应;标引不规范,在同一数据库的文献信息,有的用主题词,有的用关键词或词意标引,前后不一致,造成文献信息检索误检、漏检,效率不高。

书目数据库建立中还大量存在着低水平重复建库,机读数据时差大,利用率低,联机编目建库费用高等问题。这些都困扰着图书馆数据库的发展。

3 建设我国图书馆数据库的构想

目前,大量的信息不仅存在于图书馆收藏的书刊等传统出版物中,还存在于由政府机构、新闻单位、出版界、公司、企业、社会团体等编印的以法规、文件、简报、快报、调查统计数字、图片、表格、产品广告等形式出现的文献中,给图书馆数据库的建设造成了新的难点:一是数量大,图书馆不可能抽出许多专业人员去标引和录入;二是难以有序化,对一些文献信息根本没有相应的词表和分类法可用来处理;三是内容表现形式多样化,除文字外,还有图形和图像等。这就要求我们围绕文献信息的表达方式、内容和适用范围三个角度,采用分布式、多媒体等数据库技术来建设与读者用户需求相适应的数据库。

在文献信息的表达方式上,我们必须在坚持建设好图书馆书目、题录、文摘等二次文献数据库的同时,有重点地逐步建立起全文、数值、事实、图文、多媒体数据库。

在文献信息的内容方面,我们应在建好科研学术信息数据库的同时,向经济、管理、教育、娱乐等领域扩展,满足不同类型、不同层次读者用户的不同需求。

在文献信息的适用范围方面,我们要在建好综合性数据库的基础上,根据国家和地方政治、经济建设的发展实际,加强对专业性、专题性数据库的建设,满足不同用户群在分析、研究、决策时对文献信息获取的需求。

图书馆数据库是通过文献信息的堆积形成的,建库是一项知识密集型的工作。要提高数据库的建设速度,仅靠人工录入是不够的,还需要采用自动扫描和 OCR (光字符阅读)、自动核对、自动标引等先进的信息技术和设备,从根本上打开文献信息的入口,解决数据库录入加工整理的瓶颈问题。利用全文检索软件,提高文献信息的加工处理和利用的能力。采用主流数据库技术,解决数据库管理平台的问题,加大数值

数据库的建设,采用文件管理与数据库管理相结合的软件产品,实现图文、多媒体数据库的建设。当前,数据库系统的一些新用途逐渐得到开发,如数据开采(Data mining)、数据仓库(Data Warehousing)和数据中心库(Data repository)等也应引起充分的注意^[3]。这些都将对图书馆数据库的建设产生积极的影响。

4 对策与措施

目前我国的社会信息服务部门群雄并起,并以其信息量大、收费低廉、服务周到受到广大用户的欢迎。对此图书馆界应有紧迫感和危机感。要充分利用图书馆丰富的文献信息资源优势,把建库、管库、联库作为文献信息开发和利用的重点方向和突破口,采取有力对策和措施,抢占有利的位置。

(1) 树立信息有价的观念,提高对建库的认识。要树立“三分技术、七分管理、十二分数据”的指导思想,彻底改变重硬轻软、重建设轻维护的观念。以技术驱动为充分条件,市场驱动为必要条件建立起读者用户需求规范化和实效性强的数据库,使图书馆开发的文献信息得到充分的有效的利用。

(2) 争取政府的投入和扶植。有文献认为由于我国信息资源开发率相当低,以至于能提供公开服务的机读信息量比国外要低三个数量级^[4]。鉴于国外数据库的发展经验,我国各级政府对尚在发展初期的图书馆公益性数据库应给予必要的资金投入。对于面向商业和市场的数据库,政府拨给一定的起动资金,或通过一定的政策给予无息或低息贷款进行扶持,使之逐步地走上以库养库的道路。

(3) 重视对数据库知识产权的保护,承认数据库开发者的智力劳动,并建立相应的法律法规,有利于数据库建设的标准化。

(4) 加强规划引导。希望文化部将此列入工作的重要方面,在与全国其他系统、行业协调的基础上,制定图书馆数据库发展规划,

集中人力、物力和财力实行重点支持和突破,使图书馆数据库建设避免低起点和重复,以便尽快形成产业规模。

(5) 造就高素质的信息工作队伍。主要包括两个方面:一是培养人才,应针对数据库建设的多层次、多样化工作需求,通过举办不同类型的培训班、进修以及兄弟馆之间开展合作研究等方式,培养具有战略眼光与规划能力的信息管理人才和学术带头人,培养硬软件专家和运用计算机技术的骨干人才;二是加强职业道德教育,制定适当的倾斜政策留住人才,保持队伍的稳定和发展。

(6) 调整好业务工作结构。按自动化工作的规律和要求,调整业务工作结构,成立专门的数据制作业务部门,将人员和设备相对集中,进行合理的配置,使之形成规模,走集约化发展的道路是每一个数据库数据保持连续性和时效性,达到高标准和高质量的要求之必然。

(7) 努力培养信息市场。鉴于当前国民收入水平较低,信息用户量不足,层次也还不高,要发展信息市场,一方面有赖于经济的持续增长,另一方面可以通过加强宣传力度等方式,激发用户的信息需求,吸引他们利用文献信息。这样不仅有利于促进科研、生产的发展,有利于提高全民科学文化素质,同时还可成为推动数据库建设的直接动力。

参考文献

- 1 陈通宝 我国数据库业的现状、困惑与出路 科技日报, 1996- 02- 08
- 2 全国信息资源调查领导小组办公室 全国信息资源调查分析报告 中国信息导报, 1996(2)
- 3 王珊等编译 数据库研究: 进入 21 世纪的机遇和成就(下). 计算机世界: 专家论坛, 1996- 05- 27
- 4 徐如镜 发展信息产业事关国家命运前途 计算机世界: 信息论坛, 1995- 11- 29

郭志高 广西桂林图书馆副馆长。通讯地址: 桂林市安新小区。邮编 541002。

(来稿日期: 1996 9 2。编发者: 徐苇)