

张学福 邓平

我国数据库信息资源建设研究

ABSTRACT In present China, if the problems existed in data bank development are not solved, they will become restrictive factors to influence the economic construction. A probing discussion is made in 6 aspects on problems of data base information resource development. 22 refs

KEY WORDS Data banks Information resources Resource sharing

CLASS NUMBER G356.1

1 前言

随着计算机技术、信息存储技术和通讯技术的迅速发展,传统信息服务业加速了信息电子化的进程,并且逐渐占据主导地位。至1989年底,全球信息服务业市场规模已超过1800亿美元,其中信息电子化水平较高的美国、日本及西欧诸国占97.8%^[1]。

近几年来我国信息电子化水平不断提高。截止1995年底,全国已建成1000多个数据库,但其中隐含着许多不容忽视的问题。

(1) 重视硬件,忽视软件。“七五”期间各部委电子化信息系统建设总投入约200亿元,而用于数据库等信息资源的软件投入只有5~10亿元^[2]。

(2) 部门之间条块分割,各自为政,缺乏分工与协作,使数据库低水平重复发展。

(3) 数据库行业分布与经济发展不相称,大型数据库少,信息服务水平低。截止1992年已建成的806个数据库中10万KB记录量以上规模的数据库不足13%;与国民经济密切相关的经济、商业数据库只占16%^[3],而美国在1990年底同类数据库已占48%^[4]。

(4) 数据库只重视开发,忽视维护,商业服务水平低。我国所建数据库能做到常年更新扩充数据的不足13%,而真正能做到向社会提供商业服务的不足1.3%^[5]。

这种状况已经影响了我国信息服务水平。1992年我国信息服务业产值只占全国GNP的2%,这一比例不及发达国家的1/10^[6]。如不及时改变这种状况,到21世纪,我国数据库信息资源的储备量必然会成为经济发展的制约因素。基于这一点,笔者从数据库信息资源建设的必要性、可能性、总体目标、建设原则及布局模式等方面对其探讨。

2 数据库信息资源建设的必要性

2.1 数据库信息资源建设是我国信息资源战略化的需要

信息作为国家的重要资源和财富,已成为经济建设和社会发展不可缺少的基础。因此,制订信息资源建设发展战略应具有超前性。目前正在全球兴起的“信息高速公路”建设热潮,预示着电子信息网络时代即将来临。虽然我国现有的信息网络与“信息高速公路”的要求相差较远,但随着我国公用通信网的迅速发展,“CERNET”和“三金”工程的逐步

实施,这种差距会越来越小,然而数据库信息资源建设跟不上,必将成为我国经济发展的制约因素。由于数据库建设是一种连续、持久的工程,因此,必须从现在着手数据库信息资源建设。

2.2 数据库信息资源建设是我国图书馆自动化、网络化、资源共享的根本保证

只有建立相应的通用型、标准化的数据库,图书馆才能实现各项工作的自动化,才能借助通信网络,并利用计算机对数据库的资源进行利用,实现资源共享。因此,数据库建设是基础,无此基础,其它也就无从谈起。

2.3 数据库信息资源建设是适应用户信息需求变化的需要

目前,用户对信息的需求不再局限于单一的文献,而是拓展到要了解市场供求、竞争对手等动态信息,即用户对信息由单项需求向广泛性需求转变,且对信息的实效性提出更高要求。满足这种需求仅靠原有的文献服务是不行的,而数据库信息资源可以使用户的这些需求得到满足。

2.4 数据库信息资源建设是改变我国图书馆事业基础薄弱和发展不平衡状况的需要

我国文献资源的密集程度呈现东部—中部—西部逐渐减少的阶梯分布,并且在北京和上海形成两个富集地区。而在县以下广大农村,图书馆及文献资源拥有量极少。为了更好地为中西部地区及广大农村的经济发展提供信息服务,单纯地靠增加图书馆数量、增加藏书来满足经济发展需求既不符合国情,也不是长久之计。随着我国信息网络的发展,通过设置终端来检索和利用数据库信息资源是满足经济发展需求的明智之举。

3 数据库信息资源建设的可行性

3.1 通信网络方面

(1) 邮电通信网和专用网:我国邮电通信网络在规模上和自动化水平方面增长都非

常迅速。各类专用网也有较大发展。目前约有 60% 的专用网已和公用通信网连接^[7],对公用网起到了有益的补充。

(2) 中国公用分组交换数据网(CHINA-PAC)和中国数字数据网(CHINADDN):CHINA-PAC 除覆盖全国 700 多个城市外,目前已与美、日、法等 23 个国家和地区的 44 个分组交换网相连^[8]。CHINADDN 作为我国高速数据通信网,速率可达 2MBPS,现已与美国、香港、日本等地连通^[9]。CHINA-PAC 和 CHINADDN 的开通已形成了我国可供实际使用的中低速信息网络通信平台。

(3) 综合性计算机网络:以支持多种应用为目的的综合性计算机网络是一个国家网络应用水平的标志。1994 年,我国政府批准实施“中国教育和科研计算机网(CERNET)”,其示范工程已通过国家鉴定验收^[10]。目前,CERNET 的 64KBPS DDN 专线主干网已连接清华大学 CERNET 网络中心和 9 个地区网络中心,并于 1995 年与 Internet 网络连通^[11]。

另外,正在兴建的“三金”工程之一的金桥工程将成为我国经济信息网络通信平台。

3.2 计算机方面

目前,全国拥有大中型计算机 9000 台,微机超过 100 万台(指用于信息加工和处理的)。另外,我国微机产量已达每年几十万台,且在小型机、工作站等方面具备了相当的生产能力^[12]。

3.3 数据库制作技术方面

近年来我国已建成《法律条目全文数据库》《素问》《红楼梦》《人民日报全文数据库》^[13]等几十个全文数据库和《多媒体汉英字典》《多媒体动物百科全书》^[14]等多媒体数据库。另外,文摘数据库、索引数据库、书目数据库等制作技术更加成熟。《中国数据库大全》收录的数据库已达 1038 个。说明我国已经完全有能力进行数据库信息资源建设。

3.4 国家政策方面

1995年5月,中共中央、国务院《关于加速科学技术进步的决定》指出:“重视科技信息的有效利用和传播,加强科技图书、资料和数据库的建设,要有计划地建立全国科技信息资源传输的设施,建设连接全国科研机构、高等学校的科教信息网络,实现科技信息共享和交流的现代化。”这说明国家已为数据库信息资源建设创造了有利的社会大环境。

3.5 我国用户利用数据库信息方面

文献[15]对北京等11个城市用户利用数据库信息情况进行了调查,结果表明:我国用户非常熟悉和喜欢利用信息网络检索科研信息和进行文化交流。同时,他们对多媒体信息、电子数据交换(EDI)和多媒体技术越来越感兴趣。

综上所述,我国已具备数据库信息资源建设的各项条件。

4 数据库信息资源建设的总体目标和指导思想

4.1 总体目标

根据我国文献资源建设的总体目标^[16],结合数据库信息资源的特点,可把总体目标定为:通过对数据库信息资源的宏观调控和微观管理,使每个地区乃至整个国家范围内的数据库信息资源的储备和分布形成统一、协调、优化的整体,以求最大限度地满足具有中国特色的社会主义发展的需要。

4.2 指导思想

结合我国国情确定“统筹规划、联合建设、统一标准、联合保障”的方针为数据库信息资源建设的指导思想。

5 数据库信息资源建设的原则及布局模式

5.1 数据库信息资源建设的原则

借鉴文献资源建设原则^{[17],[18]},结合数

据库信息资源本身的特点,数据库信息资源建设应遵循以下原则:

5.1.1 适合国情原则

目前我国正处于社会主义初级阶段,经济比较落后,国力有限,同时,我国地域辽阔,经济发展不平衡,因此在数据库信息资源建设中应充分考虑我国社会主义现代化建设的实际需求并以此为出发点和归宿。

5.1.2 特色化与分工协作,引进与自建相结合原则

我国地域辽阔,文献资源分布极不平衡,在进行数据库信息资源建设时要进行宏观规划,坚持特色化与分工协作原则,避免重复建库和空白学科出现。考虑到我国国情和实际需要,在具体建库时,要坚持自建与引进相结合,重点建设中文信息数据库。

5.1.3 资源共享原则

为达到资源共享的目的,在数据库信息资源建设时应从标准化、通用性、开放性等方面出发。不论谁开发建设的信息库,他人都可通过本地网、国家网或国际网络进行检索和利用。

5.1.4 完备性原则

一方面指在数据库信息的学科建设方面应有较高层次的完备性,每一学科都要有相应的主建单位或单位群,消除空白学科。另一方面指某一学科的建设也要具有完整性;要抛弃“大而全、小而全”的观念,以本学科数据库信息资源建设为目标,确保国家在该学科数据库信息资源的完备性。

5.2 数据库信息资源建设的布局模式

5.2.1 文献信息富集地区重点建学科信息网络资源中心库,其他地区建特色信息中心库

我国文献信息资源在北京和上海形成两个富集地区。近年来我国提供信息资源服务也主要集中在京、沪等地,北京几乎拥有全国数据库总量的50%^[19]。在这些地区建设学科信息资源中心库有可靠的信息源保障,易保证学科建设的完备性。其他地区可根据本地区用户需求特点、文献分布状况,按照全国统

一协调规划,建设本地区特色信息中心库。各地的特色信息中心库与北京、上海等地的中心库通过信息网络互通有无,共同构成我国数据库信息资源保障体系。

5.2.2 学科建设与系统建设相结合模式

这种模式是以学科建设为目的,学科信息网络资源中心与专业系统相结合,进一步提高相应学科资源建设的完备程度。经济、电信、科技、金融等部门利用公用通信网和专用通信网建立了 20 多个全国性信息系统^[20],这为我国数据库信息资源建设奠定了基础。再加上负责学科网络资源中心建设的信息部门在人员素质、硬件设备、信息来源方面都有可靠保证,两者结合可起到互补的作用,从而保证数据库信息资源建设的学科完备,学科资源相对完备。

5.2.3 电子出版社与图书情报部门联合建库模式

鉴于数据库信息的特点,电子出版社可直接把电子出版物采集入库,这样可节省人、财、物力,还可延长电子出版物的使用时间;同时增加信息采集的针对性。

另外,在具体建库时,应注意以下几方面的模式转换:

(1) 信息加工模式由文献为存储单元向存储单元为信息元转化。

目前,我国的数据库大多以文献为存储单元,如文摘数据库、索引数据库、书目数据库等,而以信息元为存储单元的全文数据库只有几十个^[21],为适应数据库技术和用户信息需求向纵深发展的需要,信息处理部门要从以文献为存储单元为主的信息加工模式向以信息元为存储单元、以事实和数据为主体的信息加工模式转化。

(2) 数据库由结构化模式向非结构化模式转化。

随着计算机技术和通讯网络技术的发展,计算机信息系统正在从集中式向分布式发展。作为计算机信息系统基础的各类数据库,只有及时转换模式才能使其中的信息得

以共享。所以在建设及维护数据库时应及时把数据库从集中式系统的结构化模式向分布式系统的非结构化模式转化。

6 问题与对策

6.1 除前言中所提到的问题以外,还有以下几点:

(1) 社会信息机读化程度低,一般为 4% 左右^[22]。

(2) 数据库多以文献为存储单元,不能满足用户需求。

(3) 标准化水平低,数据库通用性差。

(4) 建库人员素质偏低,影响数据库质量。

6.2 对策

6.2.1 加强宏观调控,促使我国数据库信息资源建设协调发展

国家应建立统一的管理机构并组织有关人员对我国数据库信息资源分布状况进一步调查。在调查的基础上进行宏观调控,确定数据库信息学科网络资源建设中心单位(或单位群),特色网络资源建设中心单位(或单位群)。避免重复建设,消灭空白学科,使每一学科的建设达到相对完备程度。

6.2.2 加强以信息元为单位的信息数据库的建设和开发

一方面对已建的数据库进行深层次的开发;另一方面要注重以源数据库(数据数据库、事实数据库和全文数据库)建设为主,增强数据库数据的完整性、可靠性,提高服务水平。

6.2.3 优先在经济发达的地区建设市场经济发展急需的数据库,如商用数据库

北京、珠江三角洲和长江三角洲是我国经济比较发达的几个地区,从这些地区开始进行数据库信息资源建设,不仅有良好的文献基础、经济基础,而且有数据库建设所需的硬件基础和人才基础。再加上这些地区信息需求发展迅速,用户信息意识好,所以容易成

功, 效益也明显。在建设我国市场经济发展急需的商用数据库时, 应多借鉴国外的经验, 特别是美国在商用数据库建设方面的经验, 使我国的中文商用数据库具有世界一流水平, 然后带动其他地区数据库的建设。

6.2.4 加强横向联合, 发展大型数据库, 促进其商品化进程

21世纪是信息时代, 对数据库信息资源的开发、利用程度将直接影响国家的经济建设。鉴于我国大型数据库少的现实, 在国家的统一协调下, 相应学科的主要建设单位应联合起来, 发挥群体优势, 建设相应学科的大型数据库, 同时逐步推进数据库的商品化进程, 最终走以库养库的道路。

6.2.5 大力发展光盘数据库

鉴于目前我国广泛联网及与世界联网还需要一段时间, 同时国内外出版的光盘数据库几乎涉及了各个方面的信息, 我国有关部门已开始了光盘数据库的远程检索, 并积累了一定的经验。因此, 发展光盘数据库是现阶段提高我国文献信息电子化比例的一个重要途径。

6.2.6 加强数据库的标准化建设, 提高数据库质量

数据库的质量主要取决于机读数据、检索语言等的标准化程度。因此在标准化建设方面, 除已有的国家标准外, 尚需建成必要的各国通用的软件标准及完善的信息标准, 使我国数据库建设有标准可依。这样才能使信息分类更规范、更科学严密, 使数据库信息资源更能满足用户需求。

6.2.7 加强人才培养, 提高数据库信息资源建设的人员素质

信息靠人才来采集和管理, 高、精、尖的信息技术靠人才来掌握和使用。培养信息人才是提高我国数据库信息资源建设水平的保证。一方面应加紧对在职人员的培训, 使他们

尽快掌握计算机编目等相关技术; 另一方面在高校中开设相关专业课程, 从高起点上培养信息人才。

参考文献

- 1, 6 时文生等. 世界信息服务业的发展现状与前景. 情报学进展(1994—1995年度评论)
- 2 鄂晓芬等. 科技信息单位面向社会经济服务的发展研究. 情报学报, 1995(6)
- 3, 19, 22 吴慰慈等. 中国信息服务产业的发展态势. 中国图书馆学报, 1995(4)
- 4 杨子竞. 美国数据库服务业的现状与特点. 中国信息导报, 1996(4)
- 5 岳剑波. 越境数据流: 问题、影响、对策. 中国图书馆学报, 1995(5)
- 7, 8, 12, 20 汪冰等. 国家信息基础设施(NII)建设的热潮与我国的进展和对策. 情报学报, 1995(4)
- 9 宴章军, 沈英. 中国信息网络和信息服务的进展. 现代图书情报技术, 1995(2)
- 10 国家教委科技司. 中国教育和科研计算机网(CERNET)简介. 管理信息系统, 1996(1)
- 11 吴阶平等. 中国教育和科研计算机网的建设与发展. 世界电信, 1996(1)
- 13, 14, 21 陆长旭. 中国文献数据库建设的成就. 图书情报工作, 1996(2)
- 15 戚其秀等. “科学教育计算机网络”前期研究. 情报学报, 1995(6)
- 16 肖自力等. 全国文献资源调查与合理布局. 图书馆学通讯, 1990(4)
- 17 沈继武等. 文献资源建设. 武汉: 武汉大学出版社, 1991
- 18 林申清等. 中国文献资源布局模式研究. 图书馆学通讯, 1990(4)

张学福 黑龙江大学信息管理系讲师。
邓平 黑龙江大学图书馆副研究馆员。通讯地址: 哈尔滨市, 邮编: 150080。

(来稿时间: 1996.09.23。编发者: 徐苇)