

●谢琳惠

我国数据库产业的现状、问题及对策

摘 要 我国的数据库产业存在着规模小,上网数据少、商品化程度低,缺乏市场意识、重复建库严重、全文数据库少、标准化程度低,兼容性差,重视硬件建设,软件投入不足,政府监管不力、维护不及时,待建栏目多等问题。发展对策是:数据库经营的商业化、市场化;数据库产业的法律化、规范化;数据库经费的正常化、制度化;数据库企业的现代化、品牌化;数据库标准的科学化,统一化。参考文献4。

关键词 数据库 数据库产业 市场化 规范化 制度化 品牌化 科学化

分类号 G250.74

ABSTRACT The author analyzes some problems in Chinese database industry, such as small size, few online, low commercialization, lack of market awareness, high duplication rate, few full-text databases, low standardization, poor compatibility, lack of investments in software development, lack of government control and ineffective maintenance. Then, the author proposes some recommendations to solve these problems. 4 refs.

KEY WORDS Database. Database industry. Marketization. Standardization. Systemization.

Branding. Scientificity.

CLASS NUMBER G250.74

数据库产业是按照一定的信息需求把信息加工成为计算机可读介质并提供服务的信息服务行业,是信息产业的重要组成部分。数据库产业包括数据库、数据库产生者、数据库提供者3个部分。早在20世纪60年代后期,世界上一些发达国家(如美国等)就开始出现了数据库产业。我国的数据库产业起步比西欧晚10年,比美国晚20年,但由于政府的重视和支持,我国的数据库产业迅速崛起,已经发展成为重要的第三产业。

1 我国数据库产业发展的现状

我国的数据库产业从引进到自建大体上经历了引进、起步和发展3个阶段。

1975~1979年,主要是引进、学习、借鉴国外数据库的理论及其成果。如1975年由国防科技信息中心和北京市科协共同组建的北京文献服务处引进了美国政府研究报告GRA数据库并进行了试验性服务;又如1978年机械情报研究所引进了英国两种磁带库。

1979年~1992年,主要是我国中小型数据库起步建设阶段。如1979年由国家科技部重点扶植建立的中国药学数据库、1980年化工部信息硬件司建立的中国化学化工文献库等;又如1986年4月由全国340所高校参编的《中国高等学校学报论文文摘(CUJA)磁带》。该文摘磁带共存储了1万多条记录,并在北京文献服务处UNIVAC1100/10计算机上建立了CUJA的两个数据库。1986年国家海洋局首次引进两种CD-ROM光盘。1987年北京图书馆引进ERIC光盘。1988年又建立了中国产品信息数据库。至1988年底,我国自建的中西文数据库共有260多个。从1979年到1992年底,在国家科委科技信息司登记注册的全国已建成的数据库共计806个,记录信息5000万条。同时从国外通过各种方式共引进50多种数据库磁带,70多种CD-ROM光盘,共有3000多万条记录。

1992年至今是我国数据库产业形成和发展的重要时期。诸

如:1992年建成了中文科技期刊数据库、中国专利公报数据库;1993年生产的中国化学文献数据库、中国机电工业10万企业及产品商情数据库;1994年生产的中国生物医学文献数据库;1995年北京图书馆建立的中国国家书目数据库、中国企事业单位咨询数据库;1996年建成的中国人民大学书报资料中心文献资料全文数据库、中国百家报刊精选数据库,等等。特别是1993年2月18日我国第一家数据库专业公司——北京万方数据股份有限公司宣告成立,并于1993年8月在因特网上推出万方数据库系统,这标志着我国的数据库产业已开始市场化。据1996年11月出版的《中国数据库大全》统计,截至1995年11月,国务院64个部门和27个省(市)、自治区共建数据库1038个。它们多为中央及地方的信息机构、研究单位和图书馆所建,而且大部分集中在北京。仅北京市就占全国数据库总产量的40%,其次是上海、广东、江苏、浙江、湖南、湖北、山东、四川等。如1989年重庆维普资讯有限公司推出的VIP数据库系统,1993年上海图书馆承建的《全国报刊索引数据库》系列、1995年正式立项建设的CNKI工程等,都是我国著名的大型数据库群。近几年我国数字图书馆的建设方兴未艾,如1997年开始建设的CALIS联合书目数据库,2000年1月正式开通的在线超星数字图书馆系统,2000年4月正式运营的中国数字图书馆书目数据库,2000年5月正式开通的书生之家数字图书馆,2002年12月开始建设的CNKI-DL数字图书馆等。

2 我国数据库产业存在的问题

我国缺乏统一的管理机构,各部门、各地方过分强调地方性、专业性,因而造成目前多数数据库容量小,有的只有几千条记录,几十兆字节。我国目前数据库总容量约为世界数据库总容量的1%,产值也只有世界数据库总产值的0.1%,上网的数据库更少,还不足因特网上的1%。虽然20世纪90年代之后我国数据库产业的发展进入了快车道,但与发达国家相比,仍然缺乏大型的数据库,特别是具有特色的数据库不多。

由于我国处于市场经济的初级阶段,许多企业缺乏信息意识和竞争意识。致使一些数据库仅供内部使用,就是投放市场的数据库,也由于缺乏大企业的支持而难以维系。目前我国建成的数据库能够提供商品化服务的仅占10%,年产值不足1亿元人民币;多数数据库还没有商品化,信息资源共享性差,利用率低,甚至只建不用,成为死库。

我国目前还没有设立数据库规划和协调部门,缺乏宏观规划和总体布局,各自为政,自立项目,自行上马,盲目建库,重复建设严重,图书馆界更为严重。另外,多家重复引进数据库,浪费了国家的大量外汇。如MEDLINE就有44家引进。这种低水平重复建设,浪费了人力物力。

全国几乎各级图书馆都在建书目数据库,而且都是自建自用,向社会开放的少,全文数据库特别是专题全文数据库更少,极大地影响了读者对图书馆的利用。前文所说1995年前的1038个数据库中,书目数据库就有100多个,累计记录量在100多万条;其中一次信息只占30%,二次信息就占70%。

我国数据库标准化工作还未跟上数据库建设的需要,已建成的数据库都使用自己的标准,兼容性差。有些数据库大同小异,如重庆维普公司的“中文科技期刊篇名数据库”与万方公司的“中文科技文献数据库”重复率很高,特别是管理软件和检索界面的不规范,严重影响了资源共享。

据统计,在过去5年时间里,我国各部门共投入信息系统建设的费用达到200多亿元人民币,而投入到数据库建设的只有5亿~10亿元人民币,各种软件的研发经费不足。

我国从20世纪80年代进入信息化建设的高潮,政府投入大批资金用于各级各部门的数据库建设。由于经费是上级下拨,任务是上级布置,很多单位应付了事,经费用完,鉴定会开过,热闹一阵子后,由于缺乏后续资金和专人负责,许多数据库难以维护,由活库变成了死库。特别是政府部门和一些国有企业的数据库此现象更为严重。

有的数据库有名无实,有的半年甚至多年不更新。还有的数据库长期设置的待建栏目迟迟不建,有些死链不删除,浪费读者的时间。

3 我国数据库产业发展的对策

(1)在数据库产业发展的初级阶段,国家的投入支持非常必要。但今后,我国的数据库产业应走商业化的道路,这也是必由之路。如万方数据库股份有限公司,维普资讯公司等。当然,对于推向市场上的数据库企业,政府在税收、监管等方面应给予优惠待遇。培育市场,公平竞争,优胜劣汰。

(2)我国数据库的生产、经营、使用、安全及知识产权保护等方面还缺乏配套的法律法规,数据库在生产过程中也缺乏一定的技术规范。这是我国发展数据库产业亟待解决的问题。

(3)我国数据库产业目前出现的粗制滥造、重复浪费、无序竞争、价格居高现象,主要原因是缺乏政府监管,法规不健全。因此,建立国家级的数据库管理机构实属必要,它可隶属于各级信息产业部门。数据库管理部门的主要任务是对数据库产业发展作出规划,制定政策,对数据库建设进行协调,对数据库生产企业进行登记、统计、监管、指导等,打破部门所有、条块分割的局面,把数据库推向市场,发挥其应有的作用。

(4)凡是由各级政府部门经营的数据库,其经费应该纳入政府事业经费。否则由各级政府建设或支持的数据库就会因一哄而起,又因经费、人员等因素而昙花一现。如我国的“三上网”(即政府上网、企业上网、家庭上网)工程,有些是走过场、图形式,浪费了巨大的人力、财力。

(5)我国的数据库企业,除了需要改善外部环境,必须加强企业内部管理,创名牌产品。有些数据库生产企业不是

在提高产品质量、降低成本、技术创新、满足用户等方面下工夫,而是靠价格垄断制胜,造成价格高,销量少,用户买不起,商家卖不掉,影响了数据库市场的形成和发育。数据库企业要想有出路,应当在数据压缩、检索噪音控制、智能搜索引擎等方面搞自主创新,从而争得更多用户群和市场份额。

(6)数据库建设是一项复杂而又经常性的工作,没有统一的科学标准,难以实现共享。从内容服务考虑,制订数据库建设的标准可考虑以下内容:①收录信息的完备性标准。网络信息无限膨胀,任何数据库都不可能囊括无余。因此,数据库无论大小都应有恰当的地位,有一定的收录信息范围。在一定的范围内要尽可能地收录完备,特别是不遗漏有价值的信息,当然不能收录垃圾信息和冗余信息。②数据加工的彻底性标准。信息内容表达的复杂、元数据置标的模糊和文献出版的不规范,都会导致元数据检索结果的不确定性。因此应严格按照机读目录格式进行深加工,全部采用数字化工艺进行文本处理。采用扫描方式开发图像型数据库的方法并不可取。③数据库更新的及时性标准。从信息采集、数据库制作到网上服务,需要一定周期。但如果长达半年或一年才更新一次,则会被用户视为呆库、死库。目前由出版单位将出版物的电子文档直接传给数据库制造商,大大缩短了数据更新的周期。④使用版权的合法性。与其他有形产品不同,数据库产品属于知识产品,受到法律保护。如果数据库开发单位不依法解决版权使用问题,就有可能带来法律纠纷。因此一定要解决好版权的内部控制制度和合作制度,严格管理。⑤数据库的检索手段和检索方法要便捷。后台工作不论如何复杂,也要使前台工作简单化,尽可能不用复杂的标记符号和太多的组配方式。既要减轻用户的操作负担,又要让用户快速查到数据库的信息。

(7)我国数据库产业要想实现跨越式发展,赶超欧美发达国家和地区,需要发挥社会制度的优势,搞强强联合,联合建库,组建现代化的数据库产业集群,参与国际竞争,打入国际市场。国家和地区的数据库企业进行合作,吸收先进的科学技术,创造自己的国产品牌。我国已于2001年年底正式加入世贸组织,我国经济将以更快的速度融入世界经济。国外大批企业陆续进入中国市场,他们渴望了解到更多的中国信息;我国大批企业陆续走向世界,需要了解更多的外界信息,特别是竞争对手的信息。市场的需要就是商品发展的强大动力,数据库作为商品也是如此。当前由政府推动的电子政务、电子商务和“三上网”工程,特别是由政府重点投资的信息高速公路(如中国的十大“金字工程”建设),都极大改善了我国数据库产业的生长和发展环境,从我国政府对数据库产业的发展给予的高度重视来看,中国的数据库产业还是大有希望的。

参考文献

- 1 杜桂玲. 图书馆信息资源建设中的数据库建设. 四川图书馆学报, 2006
- 2 王兰成, 蒋丹, 刘庆辉. 全文数据库建库原理与应用技术. 情报学报, 2005
- 3 王军. 我国数据库产业发展战略研究. 河南图书馆学刊, 2000(1)
- 4 周塞峰, 刘志学. 试谈我国数据库产业的现状和前景——《人民日报》信息中心发展之启示. 北京图书馆馆刊, 1997(2)

谢琳惠 洛阳大学图书馆馆长, 副研究馆员。通讯地址: 河南洛阳。邮编 471023。

(来稿时间: 2007-04-16)