

● 牟洪兰 林春实 靖继鹏

## 数据库建设是信息产业的核心

**ABSTRACT** The paper looks back to both the history and the present situation of Chinese data base construction, discussing chiefly the strategic nucleus of the country's present information industry—the Chinese character data base construction and pointing out that we should establish data bases of the highest of world levels through our own efforts. 5 refs.

**SUBJECT TERMS** Information industry—Present situation—China  
Chinese character data bases—Constructions

**CLASS NUMBER** G356.1

数据库建设应该放在什么位置,看来国内还没有达到共识。有人认为应突出发展信息技术,把信息技术作为信息产业的战略重点。有人认为数据库是信息产业的核心,联机检索系统机、网、库三大支柱,归结到一点,还是要有雄厚的数据资源。本文考察了我国数据库的历史、现状和发展道路,得出的结论是,无论在什么时候,都应把数据库建设放在核心位置,或者说战略重点上来考虑。

### 1 历史回顾

我国信息界有一种认识,最早的数据库出现于美国的 50 年代末 60 年代初。说是美国国防部利用电子计算机开发了相应的馆藏,意味着数据基地的建立,所以称其为数据库。又说日本的数据库服务始于 1970 年。这种说法是把计算机开始用于数据加工的时间认作是数据库的诞辰。由此认为,在此以前的,例如 30 年代的美国《工程索引》、更早些的《化学文摘》以及前苏联的《文摘杂志》,称之为“数据库”也未必是错误。我们认为,计算机技术只不过是加快了数据库建设的进程,提高了数据库的质量,而不应该认为是诞

生数据库的唯一标志。从这个意义上说,我国数据库的建设与苏、日差不多是在同一个时间起步的。50 年代初,我国限于“一边倒”的历史条件,曾把前苏联的《机械制造》文摘全译成中文。后不久改变做法,建立了自己的文摘杂志,这应该说是我国数据库建设的开始。文革期间,我国信息业处于冬眠状态。科学的春天到来之后,数据库建设又慢慢复苏。稍后不久,也就是 80 年代初,王安计算机引进我国,几种大部头汉语主题词表也相继问世,北京清华大学图书馆挑头开始了大学学报数据库的建设工作。形势应该说是不错的。遗憾的是,我国的信息部门,特别是起领导作用的中央信息部门,并没有把主要力量投入到大型商用的中文数据库建设工作中去,而是把主要注意力放在引进上,甚至以引进代替中文数据库的建设。1975 年开始引进《美国政府研究报告》。此后 10 年之间,引进西文库 2000 万个记录,而中文库只有 80 万个记录。西文库一统天下,大型中文商用数据库至今一个也没有。再从政府的投资看,1985 年~1991 年的 6 年间国家对信息业的投入呈负

增长趋势,增长了一6.9%,以后投资也未见有多大提高。所以说,从50年代初学苏联开始,到70年代引进美国的磁盘,到近期所谓现代意义的信息服务为止,40余年的时间过去了,本国中文商用数据库的建设仍处于幼年阶段,更谈不上自立基业,自成体系,这不能不引起我们深刻的反思。

## 2 数据库的现状

前面提到西文库一统天下。我国在除拉萨之外的50多个大中城市已设立了110多个国际终端,可以利用12个主要系统600多个数据库,能查到世界上2/3以上的信息。但事实上这些库使用起来并不理想。

西文联机检索系统多是赔本的买卖。例如,机械科技情报所联机检索系统,1992年前后每年订购磁带(盘)费15万元人民币,而收回的钱只有3万元;化工部情报联机情报检索系统CA磁带1991年订购费约10万美元,加上磁带使用费,年支出60万元人民币,但收回只有24~25万元人民币。

西文数据库的缺点难以“医治”。各个系统存在这样那样的问题,加在一起有:追溯年限短,更新周期长,查全率、查准率低,计算机老化,重复率高,倒排字段少,方法过于复杂等。自然,这些缺点会影响检索质量。这些缺点在其生产国可能不成为缺点,或者易于克服,而在我们这样的购买国就可能成为“不治之症”。特别是技术更新问题,我们购进的磁盘使用时间还不长,而生产国又代之以光盘了。这样,我们就只能跟在别人后面,被人拉着走路。

总之,引进的西文库既赔本又被动,而自己兴办的中文库,数量、容量少,实用性差。这就是我国数据库的现状。

## 3 中文数据库建设才是信息业的战略重点

我们认为,在信息业中,处于核心地位的永远是、而且只能是数据库的建设。而在数据库建设中,优先发展的应该是文献型商用数据库。为什么?信息工作归根结底是使知识

转化为物质的工作。信息之所以被认为是当代社会发展的三大支柱之一而与能源、物质并列,就在于它有转化为物质力量的效能。宋健同志说得好,信息工作“承接前人积累的知识,激发今天的智慧,创造新的学问,并把它扩散和活化到社会主义现代化建设的各项事业中去,形成强大的物质力量。”显然,没有强大的数据库,是不能完成这样的信息工作的。

战略重点应放在数据库,特别是中文商用数据库的建设上,这一点在目前尤为重要。有的著者从战略产业眼光提出,应突出发展信息技术,应把它作为信息产业发展的战略重点,还有同志认为计算机技术在信息业中是“瓶颈产业”。诚然,我国的计算机技术和数据库一样,都还比较弱。但应该把两者关系摆好。电子计算机尚未出世之前,如前所述,数据库早就以文摘杂志的形式出现了。美国《工程索引》在30年代就问世了。《化学文摘》则诞生得更早。可见,计算机技术对数据库并不绝对产生“瓶颈”作用,因为计算机一代又一代更新,对数据库的建设,无论在提高质量还是数量上,都产生了极大的推动作用。例如,计算机在美国与实力雄厚的文摘刊物相结合,使美国信息业在不到30年的时间里便登上了世界第一的宝座。但从世界各国的情况看,的确也出现过一些复杂的情况。例如,前苏联的文摘杂志,在世界上也名列前茅。但由于种种原因,其中包括计算机的功用没有得到政府领导人的重视,所以到1990年为止,还只有科学院系统的两个规模不大的专业数据库。可见在前苏联,计算机技术对数据明显地起到了“瓶颈”作用。不过那是僵化了的思想造成的结果,并不是什么规律性的东西。至于从我国目前的情况看,虽说信息技术(包括计算机技术和通讯技术)与数据库建设都还比较弱,但比较起来,前者似乎略强于后者。例如,电子版《邓小平文选》一、二、三卷的全部内容都储存在两张高密度软盘上,在普通微机上阅读和检索都十分方便,读者可以组

合任意关键词,以秒级速度准确无误地检索到所需的句子和段落。再如,我国光盘出版物 1992 年才一两种,1994 年便有几十种,其中包括《中国大百科全书》和《中国少年儿童百科全书》等。我国的卫星通讯、传真等通讯技术也有一定的成就。如此说来,目前对信息业而言,说数据库已经成了“瓶颈”岂不更符合实际一些。其实,战略产业这一概念本来就不是一成不变的。今天是“瓶颈”的东西,明天就可能不是“瓶颈”了。所以,无论从哪方面看。数据库的建设一点也不能放松,必须从战略的高度来看待我国数据库的建设问题。

#### 4 走自己的路

邓小平同志一向提倡走自己的路。他指出:“关键是本世纪内的最后十年,要为下个世纪前 50 年的发展打下基础,创造比较好的条件和环境。”我们想,在这个“比较好的条件和环境”中,为社会信息化打下基础应该是十分重要的一环。而在数据库建设方面,尤其要走出自己的一条路来,应该有世界上第一流的信息服务,特别是应建设第一流的数据库。

4.1 树立强烈的竞争意识,面向世界市场,以美国作为竞争对手。美国数据库最多。全世界商业数据库美国占一半,近 4000 个,服务公司 900 余家,联机终端 40 多万户,遍及世界各地。美国数据库的质量、品种、类型、规模、产值、使用效率以及工效比等,都占有优势,超过其他所有国家和地区的数据库。例如,美国《科学引文索引》系统的数据库,内容之丰富,功能之多,都堪称世界之最。尤以其产品包罗万象的科学图和图册堪称一绝。它可以指明一个国家或地区,以及某个科学家在世界科学发展中所做出的具体贡献,可以看出某门学科的前沿变化,简直是一部动态的、包罗万象的科学百科全书。因此,在我们的信息业冲向世界市场中,竞争对手就是美国。

4.2 统一安排,分散建库。我国是一个有 12 亿人口的大国,数据库不宜只集中在一两个

大城市,我国又是一个计划经济与市场经济相结合的国家,更应该做到统一安排建设问题。这里的统一,主要有两层意思。一层是中央统一安排国家的投资问题。宋健同志说过:“宁可停止一两个大项目,也要把信息产业搞上去。”另一层意思是在信息业的投资中数据库应该优先。因为国家对信息业的投资往往不能多,例如 1985 年对信息业的投资才 2.7 亿元人民币,如果数据库不占点优先地位,则资金之微是可想而知的。所谓分散建库,就是既不应该把力量集中在少数几个城市,但又不应该四处开花,搞“小而全”。而是根据国家的总体安排,在条件较好的地方分期分批建库,建一个成一个。既要量力而为,又要尽力而为。

4.3 建设世界一流水平的中文商业数据库。一流库的基本要求:无论是专业性或综合性的,都必需具有最充实的内容,尽可能囊括全人类的全部知识;有最好的查全率和查准率;有良好的网络性,多元检索性,远距离操作性;其信息产品具有强烈的适时性和切题性。要建成这样的数据库,当然需要一流的信息人才。其中特别是组织人才、文摘员兼翻译、标引员兼编辑。前面提到的美国《科学引文索引》之所以能够称雄世界,就是因为它有一位优秀的组织者加菲尔德博士和凝聚在他周围的一大群优秀学者和信息工作人员。

我国是一个发展中国家,具有中国特色的社会主义建设事业正在蓬勃发展,信息源日趋丰富。据报道,我国目前年出版发行图书 8 万种,1500 万册,研究报告 431 万件,报纸 1486 种,正式刊物 5880 种,学术论文 20 万篇,专利 220 万件。另外,还有档案文件 1.14 亿卷。从信息社会的高度来看,应把五千年的文化都算进来。作为信息专家,应该把五千年的中国文明放到世界知识宝库中去。我们之所以认为数据库建设是信息服务业的核心,是因为数据库作为一种强大的工具,能把整

(下转第 21 页)

生产第一线为农民服务,除介绍馆藏文献,推荐新图书、新知识、新信息外,最好还熟悉或掌握一些适用技术和本县(市)科技致富的典型事例,以增加农民学科学、学会利用图书馆的兴趣。也可请农技员和科技示范户现身说法,现场指导。(3)以点带面原则。由于受人力、财力和交通工具的限制,深入乡村不可能面面俱到,也不能经常“打一枪换一个地方”。要选准突破口,建立较长时期的服务点,以点带面,不断推进。(4)和农民读书活动相结合。如1986~1987年仙桃市图书馆开展了“百万农民读书读报有奖竞赛活动”,动员了31.3万农民参加,市图书馆和乡村200多个图书馆、室向农民提供了18万余册(份)科技书刊资料,活动开始后,该市出版的《仙桃农技》、《植物保护》两个刊物发行量增加了26万份。据统计,当年实现经济效益5.3亿元,占全市社会总产值的1/4<sup>[3]</sup>。

知识扶贫的效益是巨大的。它是一项投资很少、收效巨大、造福于民的知识工程。它需要得到国家在经费、交通工具(汽车)、知识信息检索工具(计算机)方面的支持。目前,开展知识扶贫的最大障碍是思想障碍。在市场

经济条件下,人们往往容易看到钱的力量而忽视精神的力量。知识扶贫工作是一项崇高的工作,没有崇高的精神境界是不可能去乐于奉献、作出牺牲的。公共图书馆界,尤其是县图书馆、山区图书馆之所以能够在缺经费、缺人员、没有交通工具的情况下,长年累月坚持下乡,为农民服务,就是因为没有忘记生我养我的土地,没有忘记中国的大多数人是农民,农民迫切需要知识和技术。长期坚持为农民致富奉献一切的图书馆工作人员是令人敬佩的。

#### 参考文献

- 1 宋健.把科技星火撒向农村.星火之光,1987(12)
- 2 江泽民.扶贫开发要坚持不懈、锲而不舍地长期抓下去.求是,1994(10)
- 3 李刚.开展“农民读书读报有奖活动”,是把知识星火撒向农村的有效措施.农业图书馆杂志,1987(3)

**李刚** 1968年华中农大农机专业毕业。现为湖北省图书馆副研究馆员,研究辅导部主任。通讯地址:武汉市,邮码430060。  
(来稿时间:1995-05-03.编发者:翟凤岐。)

(上接第36页)

个人类文明都组织和调动起来,为人类进入美好的信息社会打下一个良好的基础。

邓小平同志指出:“中国必须发展自己的高科技,在世界高科技领域占有一席之地。”并指出,应“开发信息资源,服务四化建设。”只要我们看准自己的路并坚定不移地走下去,中国信息业称雄世界的日子是不会太远的。

#### 参考文献

- 1 罗曼.信息产业发展的三大规律与启示.情报科学,1994,(15):26
- 2 陈道全.联机系统数据库大规模和正规化建设.情报学报,1993,12(3):195~199

- 3 刘东维,薛曙光.中国科技情报系统投资规模结构及其格局的演变.情报学报,1993,12(2):106~112
- 4 徐明忠等.信息服务业现状及其运行机制和发展模式.国外情报科学,1994,(1):42
- 5 赵少平.我国引进磁带自建西文数据库适应性经济性研究.情报学报,1993,12(4):241~249

**牟洪兰** 1982年北京大学图书馆学系毕业。现为吉林工业大学图书馆馆员。通讯地址:长春市斯大林大街114号,邮码130025。

**林春实** 毕业于东北师范大学图书馆学专业、图书情报专业,现为吉林工业大学图书馆馆员。通讯地址同上。

**靖继鹏** 吉林工业大学情报工程系系主任、教授。通讯地址同上。

(来稿时间:1995-03-22.编发者:丘峰。)