

关飞霞

中国数据库产业的现状与发展

ABSTRACT Based on an analysis of the present status, the author discusses the future prospects of the industrialisation and internationalisation of the development of databases in China under coordinated impacts of the macro-regulation and market orientation. 17 refs

KEY WORDS Database Market economy Future prospects

CLASS NUMBER G258.94

当前,信息高速公路建设热潮席卷全球。信息高速公路即“国家信息基础设施”(NII),包括信息设备、信息资源、通信网络和人才资源四个方面。通俗地讲,就是人、机、网、库四个要素。其中,信息资源(数据库是最主要的组成部分)是最重要的因素,而目前在中国,它又是最薄弱的环节。因此,在中国,加强数据库的建设是当务之急。而当前中国市场经济深化的形势,又为数据库建设提供了良好的契机。既是挑战,又是机遇。

本文试图在对中国数据库产业现状的分析基础上,展望其未来发展。

1 现状

中国数据库的建设已有20多年的历史,经历了三个发展阶段。

第一阶段(1974~1979年):主要是引进、学习与借鉴国外数据库理论和成果,着重引进和解决汉字处理技术。

第二阶段(1980~1993年):研究和建设中文数据库。主要集中在科技文教部门。

第三阶段(1993年至今):以中国第一家数据库专业公司——“万方数据公司”的正式成立为标志,开始以独立经济人的身份走向市场。

国家科委科技信息司、国家信息中心等权

威部门对中国数据库进行过两次大规模调查。

第一次是在1992年,共上报806个数据库,经筛选收入《数据库指南》的有137个。可用的不到1/10,即能在一定范围内使用的数据库只有几十个,且容量大部分在10MB以下。

第二次即最近的一次是在1995年,截止到1995年10月31日,共收到数据库调查表1562份,去掉重复的、国外引进的、仅用于内部使用的及规模太小的数据库,收录到已经出版的《中国数据库大全》中的数据库有1038个。一半左右已在一定范围内提供不同程度的服务,且容量大部分在10~100MB之间。

从这两次调查结果可看出,中国的数据库建设已有很大的发展:

数量增加了,特别是可用数据库大大增加;

规模扩大了,总容量有很大提高;

结构变化了,从过去的科技数据库为主,已开始向以经济和社会的数据库为主过渡;

分布改进了,各地开发的数据库已超过半数;

服务方式增多了,脱机的光盘数据库已成为数据库销售的主要方式,并还在继续增加。然而同国外数据库的发展相比,还很落后。

国际上数据库的发展经历了三个阶段:从60年代初到70年代初,为第一阶段,数据库的内容主要是科技方面,且由政府资助;从70年

代中期到 80 年代中期, 为第二阶段, 数据库的内容以商业、金融为主; 80 年代后期以来, 为第三阶段, 数据库开始为家庭和社会服务, 且出现了多媒体数据库等。

显而易见, 我国数据库目前的发展水平只相当于发达国家 70 年代的水平, 仍存在着许多问题:

(1) 自用的数据库多, 商用的数据库少; 科技数据库多, 商用数据库少。

(2) 单机、脱机使用的数据库多, 联网使用的少, 共享程度低。

(3) 质量不高。表现在: 规模小, 覆盖面窄, 标准化程度低, 用户界面不够友好, 更新频率低。

(4) 经济效益差。具体表现在: 数量多, 产值少; 投入多, 产出少。

(5) 二次(书目、文摘等)信息数据库多, 全文、事实、多媒体等一次信息数据库少。

(6) 引进的数据库多, 自建的数据库少, 出口的更少。

(7) 处于无序状态。缺乏必要的宏观调控, 使数据库的结构、分布不合理, 各自为政, 低水平重复。

笔者以为, 造成上述问题的根本原因在于: 没有摆脱旧的计划经济体制的影响。不适应新的市场经济的要求。而市场经济的要求有两方面: 一是市场化、产业化; 一是必要的宏观调控。二者的相辅相成, 在中国目前市场经济的初期阶段, 尤为必要。

市场化就要求数据库的建设充分面向市场, 向产业化方向发展, 它要求数据库更多地为商业服务, 要求信息的大量共享, 要求标准化、规范化, 它的发展趋势必然是国际化的要求。

作为市场化的必要补充, 必要的宏观调控可以使数据库的结构分布趋于合理, 从无序走向有序。

2 展望

目前, 中国对信息业投资中重硬件设备、

网络建设, 轻资源开发、数据库建设, 造成数据库与信息系统建设投资比例严重失调, 数据库建设严重滞后, 影响了信息业的协调发展。

显然, 数据库建设已成为制约中国信息业发展的瓶颈。但与此同时, 又为数据库建设带来充满生机的发展空间。

这就需要在市场化与宏观调控之间找到一个平衡点: “稳住一头、放开一片”。

2.1 宏观调控

早在 1984 年, 邓小平同志就指出: “开发信息资源, 服务四化建设”。

江泽民主席强调: “四个现代化, 哪一化也离不开信息化”。

1996 年来, 国务院信息化工作领导小组组织制定了《国家信息化“九五”规划和 2010 年远景目标(纲要)》, 提出了中国信息化工作的指导方针, 即“统筹规划、国家主导、统一标准、联合建设、互联互通、资源共享”的二十四字方针。

1997 年 4 月, 国务院召开首次全国信息化工作会议, 提出了当前国家信息化建设的主要任务。会议把信息资源的开发和利用放在首位, 将其作为国家信息化的核心任务。指出要大力组织具有民族优秀文化传统的中文信息资源上网。

国家科委也于 1997 年 4 月初提出了“关于加强信息资源建设的若干意见”, 指出加强数据库建设等意见。

1998 年初, 国务院大刀阔斧地进行机构改革, 精简、合并机构, 却新设立了信息产业部。这预示着困扰中国数据库建设的一系列问题将逐步得到解决。

(1) 国家资助建库。对于国家必不可少的资源, 虽然用的人很少, 但关系到国家未来的发展, 如具有基础性、全局性、先导性、战略性的数据库, 以及社会公益型数据库等, 国家要支持, 或主办, 或资助。

(2) 制定法规政策。一些发达国家已形成各自的数据库产业政策及有关法规, 如欧共体制定了《数据库法律保护指令》。中国尚无数据

库法。数据库的产业化及良性发展,迫切需要有关的法规政策加以引导、规范。目前,国务院信息办正在起草国家信息化总体规划的信息资源开发利用子规划、政策法规子规划等,并发布了《中国互联网络域名注册暂行管理办法》。

(3) 制定标准。数据库的联网、共享,必然要求标准化。按照与国际标准接轨的原则,加强加快中国国家标准的制定与颁布,使数据库产品标准化、规范化,质量得以提高。

还应设立国家级数据评估组织,以评估数据库产品的质量,特别是其标准化的程度。国际科学数据委员会(CODATA)就设有数据评估的专门组织。开发数据库较早的国家均有相应的数据评估组织。

(4) 逐步公开政府部门非保密的信息。在中国,政府部门掌握着巨大的信息资源,占信息资源总量的 80%,但大部分“养在深闺人未识”。而在发达国家,政府部门采集的信息有一些是免费为公众服务的,或以很低廉的价格出售。我们应在确保安全保密的前提下,逐步公开政府部门可以公开的公共信息,以便为数据库公司提供稳定的、高质量的、权威的数据源,从而有利于数据库质量的提高,有利于数据库业的发展。这也是政府义不容辞的责任与义务。政府信息机构不能以自己掌握的信息资源参与竞争,因为这是垄断性的、不公平、不正当的竞争,不利于数据库业的发展。

2.2 市场化

除了国家支持、资助的基础性、战略性、公益性数据库外;对于一般性的为社会、家庭服务的数据库,特别是“短、平、快”(经济、金融等)的数据库,就要面向市场,走社会化、产业化、发展道路。

(1) 企业群。中国的数据库建设,还远未形成一个产业,因为没有足够的企业投身到这一行业中。数据库的产业化,呼唤着企业群的出现。

可喜的是,已经有一些公司在数据库建设方面做出了有益的探索。例如,万方数据(集

团)公司。其前身是中国科技信息研究所的标本馆,它在没要国家和上级部门的经费的情况下,在比绝大多数数据库差得多的条件下,以以库养库、合作建库的方式,成功地建成了《中国企业、公司及产品数据库》(CECDB)。靠的便是市场无穷的力量。1993年,在没有国家投资的 CECDB 的基础上,作为中国第一家数据库专业公司的万方数据公司正式成立。继续用以库养库、合作建库的方式,不断开拓新品种,目前已发展到 10 余个数据库。它的 CECDB 以光盘形式在欧美地区上市,其英文版已进入 DIA-LOG 和 DATASTAR 两大国际联机系统。CECDB 的用户已遍布世界 40 多个国家和地区。它还在全国信息机构联合承建了 China Info(中国信息)资源系统,已于 1997 年 8 月正式开通。

“万方”在计划经济中率先以“以库养库”的方式开发项目,在科技改革的初期阶段率先以市场模式进入市场,如今的目标则是在激烈的市场竞争中占领市场。

原国家科委已正式宣布筹建中国最大的、能与国际接轨的中国信息市场(CM)。这必然为越来越多的数据库企业的涌现、为数据库的产业化,创造良好的市场条件。

随着企业群的形成与发展,必然会出现下面的一些发展趋势:

数据库企业的功能将分离,即向专业化方向发展。中国现在的数据库建设远未规模化、产业化,存在的只是一个个“大而全”、“小而全”的建库单位。而产业化必然要求专业化。随着企业群的出现,“大而全”、“小而全”的模式必然会被打破,企业的功能将出现分离,从而分化出专门的数据库生产商、销售商、服务商等。企业功能的分化即专门化,必然会提高数据库生产及服务的质量,有利于整个产业的发展。

新的信息技术将被普遍采用。如分布式数据库,它使应用能透明地对分布的数据进行操作,它适应数据在逻辑上和物理上分布的实际情况,提供处理的高效性与高度的可访问性,

而对用户来讲,却完全像一个非分布式的系统。面向对象的数据库,它把面向对象技术与数据库技术结合起来,利用类的设施来描述复杂对象,利用类中封装的方法来模拟对象的复杂行为,能很好地适应CAD、CASE、GIS等新的应用、Web应用、新的商业应用等。多媒体数据库,它将解决信息媒体的多种化、多媒体数据集成、多媒体数据与人之间的实时交互等难题。基于数据库的知识获取技术(KDD),是机器学习、专家系统、数据库、统计学、模糊学等多种技术交叉的一个领域,它绕过了专家系统中知识获取的瓶颈,形成了用数据库作为知识源的一整套新的策略和方法。新的检索技术,如多媒体数据库检索技术、全文本和超文本检索等,也将成为发展趋势。

数据库的服务方式将采用联机为主,光盘为辅,较少采用文字版、软盘等其他方式。这是因为联网已成为全世界不可阻挡的潮流,联机服务方式将成为用户首选的方式,而光盘则是作为联机的补充方式而存在。

可以预见,数据库的进一步发展及产业化,其直接起因必然是企业群的形成。有竞争又有合作的企业群的出现,必将带来数据库产业的繁荣。

(2) 用户。毋庸讳言,用户的存在及需求,必将刺激数据库的发展。

目前,计算机在中国已得到极大的普及,正越来越多地进入办公室和家庭。但大部分个人计算机只是当作打字机与游戏机用。这虽然是种巨大的浪费,但也培养了大量潜在的数据库用户。因为这些微机的操作者与使用者,将越来越不满足于微机目前的功能,必将寻求联网,也就意味着查询网上数据库的信息。

同时,“金”字系列工程在不断建设与发展,中国的通讯网络已初具规模。目前已有四大互连网络,即:中国教科网、中国科技网、中国公用网、中国金桥网。我们用三年的时间,走完了国外30年的发展路程。这些网络与专门应用系统为数据库的联机服务奠定了雄厚的物质基础,也必然吸引大量的网络用户。

但是,一方面,我们的信息高速公路或称中速甚至低速公路上近存在着“有路无车”、“有车无货”或“有货低质”等尴尬问题;另一方面,通讯费用的高昂,又使用户望而却步。因为联机费用由两部分组成,通讯费与数据库使用费。通讯费的居高不下,阻碍了联机用户的发展。

解决通讯费用过高的问题,将有赖于通过邮电通信部门的改革,即引入竞争机制,打破行业垄断来实现。1997年2月,在日内瓦召开的世界贸易组织会议上,68个国家的代表承诺不同程度地开放本国电信业务。可见,各国电信业务走向更加开放的大趋势已经明显。可喜的是,原邮电部部长也指出:“开放市场是我国邮电通信的原动力”。而中国联通公司的成立在电信领域初步形成了既竞争又联合、相互促进、共同发展的局面。相信随着中国市场经济的深入发展,邮电部门将进一步改革,从而制约网络发展的通讯费用高昂的问题将得到解决,大量潜在的联机用户可望转化为真正的现实用户。

显然,对于中国的数据库产品,不仅存在着大量潜在的国内用户,也同样存在着广大的国外用户。

然而,中国现在出口的数据库还很少,能够走向世界的数据库公司也寥寥无几。

诚然,中国应是中文数据库最大生产国与出口国。但是,如果我们不加紧数据库的建设,本应属于我们的市场就可能被别的国家捷足先登,取而代之。这并非危言耸听。因为国际舆论认为,21世纪的中国将成为全球最大的电子信息市场。电子信息市场已成为各大跨国公司争夺的焦点。REED、D & B、IHS、DOW JONES和KNIGHT-RIDDER等国外信息服务商已纷纷在中国建立了办事处。他们拥有资金、信息处理技术、在全世界的销售渠道和市场声誉等优势,一旦全面进入中国市场,就会给我们带来极大冲击。

但在建设中文数据库方面,我们毕竟是集天时、地利、人和之最大优势,只要善于抓住时

机, 必然会赢得广泛的国外用户。对于有些数据库, 我们还可以采用中、英文双语的形式发行。我们不仅要保住国内市场, 还应尽量开拓与占领国外市场。

数据库的建设, 不仅仅是技术问题, 也是经济问题——是争夺市场的焦点, 还是政治文化问题。能不能建设好数据库(即信息资源), 关系到国家的命运前途——能不能自立于世界之林。

现在, 保护和发扬民族优秀的文化传统, 已引起世界各国特别是发展中国家的普遍重视。1996 年 5 月在南非召开的“信息社会与发展大会”上, 对这一问题的呼声很高。很多国家的代表强调, 要保持语言和文化的多种性, 保持全世界多民族的优良文化传统。

新加坡政府为了抗衡西方新闻媒介对其负面报道, 将大型数据库 Singaport Infom ap 输入 Internet, 向全球的网络用户介绍新加坡的发展成就。目前, 在 Internet 上已形成有关新加坡社会政治的论坛。

因此, 面临外来的文化侵袭与威胁, 靠“堵”、靠“躲”都行不通, 只能上网进行正面对抗。况且, 在多元化的世界格局中, 我们也有责任组织具有民族文化传统的数据库上网, 将中国的文化介绍到全世界。

相信, 随着中国市场经济的进一步发展及改革的全面推进, 中国的数据库建设定会克服重重困难, 走过黎明前的黑暗, 成为名副其实的朝阳产业。

从“国家经济信息化联席会议”(1993 年)到“国务院信息化工作领导小组”(1996 年)以及“中华人民共和国信息产业部”(1998 年)的成立, 这一国家机构的意味深长的变迁本身就是最好的预言。

参考文献

1 王俊兰等 国内外数据库建设与政策研究 课题

研究报告, 1990 10

- 2 国家计委, 国家科委, 国家信息中心. 中国数据库大全 北京: 中国计划出版社, 1996
- 3 陈通宝 我国数据库产业和电子化信息资源 情报学报, 1996(10)
- 4 吴家柱 中国数据库产业现状及其存在的主要问题 现代图书情报技术, 1998(2)
- 5 张帆 雾天里的朝阳——中国数据库业透视 中国信息导报, 1995(4)
- 6 “中国信息市场”和“全国科技信息网络”建设在全力推进 中国信息导报, 1995(3)
- 7 在奋进中的万方数据公司 中国信息导报, 1995(3)
- 8 包锦章 数据库产品商品化的探求 现代图书情报技术, 1997(5)
- 9 国家科委关于加强信息资源建设的若干意见 中国信息导报, 1997(5)
- 10 朱伟 中国国家信息基础设施的现状 情报学报, 1997(12)
- 11 胡启立 我国信息化建设的几个问题 中国信息导报, 1997(8)
- 12 邹家华 加快推进国家信息化 信息世界, 1997(8)
- 13 党小卉 中国邮电通信驶入快车道——国家邮电部部长吴基传谈我国邮电通信业发展及市场前景 经济与信息, 1995(8)
- 14 国家信息网络篇 信息经济与技术, 1998(1)
- 15 李德毅 数据库技术展望 计算机世界, 1995-03-22
- 16 Date, C. J. An Introduction to Database Systems 6th ed., Addison-Wesley, 1995
- 17 Michael Stonebraker, Dorothy Moore Object-Relational DBMSs: the Next Great Wave Morgan Kaufmann, 1996

关飞霞 北京大学信息管理系。通讯地址: 北京海淀区。邮编 100871。

(来稿时间: 1998 4 28。编发者: 刘喜申)