第 1 章　　绪 论

1.1　数据库的概念

1.1.1　数据库的定义

现实世界充满着信息。所谓信息泛指反映特定事物的消息、情报或知识，这种消息、情报和知识以可被感受的声音、文字、图像、符号等为表征，并可通过各种方式传播。信息是对客观事物的反映。

以特定的符号表达的信息称为数据。数据是信息的一种量化表示，数据反映信息，而信息依靠数据来表达。表达信息的符号可以是数字、文字或图形。计算机所能接受、处理和存放的信息须是数字化的信息，因此，必须人为地把信息转换成可以被计算机接受的数据，也就是以二进制形式存储在计算机内并被计算机加工处理的数据。

一定数量的数据存储于特定的存储介质之上，以特定的结构相互联结于一体，形成了数据库。数据库是以一定的组织方式存储在一起，相互关联而独立于应用程序之外，并能为多个用户所共享的数据集合。

1.1.2　数据库的特点

1.有序性

需要全本请访问 www.ertongbook.com

数据库不仅存储数据，而且连同数据之间的逻辑关系一起存储，也就是说数据库中的所有信息不是无序的堆积，而是相互之间存在着一定的逻辑关系，构成一个有序的信息系统。正是这种有序性，使得用户可以方便地查检和利用数据库中的数据。

2.共享性

数据库的数据一经输入，即可多次多重输出，可以为多个用户所共同利用，甚至是同时利用。有效地克服了信息传递的时空障碍。

3.独立性

数据的逻辑组织和数据的物理存储方式与用户的应用程序无关,一旦数据结构改变,与这些数据有关的程序不需要重新编写和调试，可节省大量的开支。这样，也使得同一个数据库可以被不同的应用程序所使用。

4.完整性

数据库所聚集的数据应全面地反映特定主题范围内的相关信息，形成完整的覆盖面，而且通过不间断的数据维护活动，始终保持数据的这种完整和正确。

5.最小冗余性

数据库数据之间多维的逻辑结构使得每条数据可以通过不同的途径得以反映，从而避免了数据的重复存储，最大限度地减少了数据的冗余度，节省空间和数据更新的开支，同时保证了数据的一致性，提高了信息存储利用的效益。

1.2 数据库的类型

1.2.1 按数据库性质划分

按其提供数据的性质，数据库可以分为文献数据库、数值数据

库、事实数据库和多媒体数据库。

1. 文献数据库

文献数据库是指报道各类文献及文献信息的数据库，包括书目数据库和全文数据库。

(1) 书目数据库

书目数据库是只存储有关主题领域各类文献资料的书目信息，以二次文献的形式报道文献的数据库，如题录数据库、文摘数据库、引文数据库、期刊目次数据库以及图书馆馆藏目录数据库等。书目数据库以简略的形式向用户报道文献的信息，提供查找、获取文献的线索。这类数据库信息量大，信息密度高，文献报道范围广，数据连续性、累积性强，是用户快速查找文献的有效工具。

书目数据库中的数据来源于期刊论文、会议论文、研究报告、专利文献、学位论文、图书、政府出版物、报纸等各种不同的一次文献，是经过加工、压缩的派生性数据。

书目数据库的数据结构一般比较简单，记录格式也较为固定。因此，其建库速度往往比较快，建设费用相对比较低。

文摘索引数据库和图书馆馆藏目录数据库是最常用的书目数据库。

文摘索引数据库的内容性质与书本式文摘索引相同，主要是简要地通报有关领域特定时期内发表的文章，供人们查阅与检索。它提供确定的文献来源信息，即能准确鉴别相对应的原始文献。但是，它一般不提供原始文献的收藏地点。

图书馆馆藏目录数据库又称为“机读目录”，即 MARC(Machine - Readable Catalogue)。它主要报道特定图书馆实际收藏的各种文献资料的书目信息和存储地址。它既是一般用户查找图书馆资料的工具，又是图书馆业务部门的业务管理工具。它所报道的数据内容详细，除描述文献本身以外，还有许多附加信息，如业务加工信息、管理信息、馆藏信息等，记录格式也比较统一。我国

各类图书馆的馆藏目录数据库都按照《中国机读目录通讯格式》(CNMARC)著录。

(2) 全文数据库

全文数据库是存储文献全文或其中主要部分，以一次文献的形式直接提供文献的源数据库。通常将经典著作、法律条文及案例记录、重要的科技期刊、新闻报道和百科全书、手册、年鉴,以及图书馆所藏的其他重要文献的全部文字或主要文字转换成计算机可读形式，建成数据库。用户可以从中直接检出所需的原始文献。

与书目数据库相比，全文数据库有许多特点和优点，主要表现在以下方面：

①检索直接。能直接检索出原始文献或解决问题所需的文献资料，不必进行二次检索（即根据检出的书目信息再去查找原文）。

②报道详尽。文献的正文部分或附属部分都可以检索获得和显示，用户可以直接查看到文献正文中的每一段、每一句和每一个词，甚至还可以看到某些边缘性信息。

③传递快速。用户可以通过检索系统迅速地浏览、检索和获得文献原文，不受地理位置的限制。

④标引全面，使用简单。绝大多数全文数据库多利用计算机进行全文自动抽词标引，生成倒排档。能为用户提供标题、著者、关键词、摘要等多重检索途径,使用方便。

⑤用户接口多为菜单驱动型，或采用较简单的检索命令，易学易用。

⑥检索语言多用自然语言，少数用受控语言。检索方法除使用布尔检索以外，位置检索占有相当突出的地位。

而且，随着信息处理技术和储存技术的发展和突破，原先困扰全文数据库发展的存储空间、传递速度、文本转换和信息内容稳定性、可靠性等技术问题已得到基本解决，全文数据库获得了迅速的发展和普及，成为数据库技术和产业的重要发展方向。

2.数值数据库

除了以文献单元为存储和处理单元的文献数据库外，还有一些以数值、事实、术语或图像为存储和处理单元的数据库，即非文献数据库。非文献数据库包括数值数据库、事实数据库、术语数据库和多媒体数据库等。

数值数据库是一种以自然数值形式表示、计算机可读的数据集合，主要记录和提供特定事物的性能、数量特征等信息。其信息报道范围常覆盖某一大类的专业领域。如在军事上，部队的实力、装备、武器的性能等信息；在商业和经济领域中，特定产品的性能特征、价格趋势、国家经济增长率等数值信息 在科技领域,物质的物理化学性质、结构、频谱及实验数据、计算公式等。

数值数据库所记录的数值数据是人们从文献资料中分析提炼出来的，或是从实验、观测和统计工作中直接得到的。数据库生产者将这些数据收集起来，经过核实、检验和加工整理，按一定的方式组织起来，利用计算机进行存储和检索，就成了数值数据库。如果数据库中还含有定义数值或说明这些数据项所必需的文字（文本数据），则又称为文本一数值数据库。与文献数据库相比，数值数据库是人们对信息进行深度加工的产物，它可以直接提供解决问题时所需的数据，是进行各种统计分析、定量研究、管理决策和预测的重要工具。

数值数据库对所收录的数据的可靠性要求比较高，有时还需要列出数据的误差估计、数据来源和实验条件等，以减少用户使用中的误差。数值数据库存储的数据通常成组排列，他们本身并不被检索，但它们与字母或数字形式表达的可检索的关键词或叙词相联结，成为可检索的数据。数值数据库检索的结果是一个或一组特定的数值。

为了满足用户的不同需要，许多数值数据库还常附有特定的检索软件包,以提供数值的运算、统计分析、分类、排序和重组等功

需要全本请访问 www.ertongbook.com

能。这就使数值数据库系统除了具备一般检索功能外，还常能提供如下一些功能：

(1)数据运算能力。可以根据现有数据建立函数关系，进行数值的计算，例如可以要求系统以某一年为基准年，重新确定时序的指数；要求系统进行不同计量单位的换算；要求计算时达到某种精确度,等等。

(2) 图形处理能力。例如，有的系统能提供分子结构式或晶体结构图、光谱图等，有的可以处理非线性座标等。

(3)报表生成能力。能根据用户的需要自动生成各种报表，或对输出数据进行排序和重新组织，或允许转存数据，建立新的文档。

(4) 数据分析能力。例如，为复杂的现象构造仿真模型；或利用各种数值分析技术进行预测、可靠性计算和假设检验；甚至可利用对数据的观察和统计分析产生新的思想等。

(5)试验模拟能力。能够模拟具有数学模式的试验；模拟化学反应过程等。

3.事实数据库

事实数据库是存储有关事物(如人物、机构、事件等)的一般指示性描述的参考数据库。因此又称之为“指示性数据库”或“指南数据库”。其主要用途是供用户查询特定事物的发生时间、地点、过程等简要情况。

事实数据库的类型很多。按信息类型划分，常有收录各种人物传记信息的人物传记数据库，收录各种公司的生产与经营活动信息的公司名录数据库，存储各种基金信息的基金指南数据库，存储各种技术标准或规程信息的技术标准指南数据库，存储各种计算机软件目录信息的软件数据库，存储各种产品和商品信息的产品指南数据库以及面向各行业管理的管理信息数据库，存储各种武器装备的技术性能、制造技术、使用方法的武器装备数据库等

等，多种类型的事实数据库。

事实数据库中存储的数据一般是用来描述人物、机构、研究项目、产品或商品等非文献信息源的。与其它数据库相比，它有以下特点：

(1)提供的不是信息文献，而是各种类型的实体信息，每个款目都是对事实的描述。

(2)数据库记录长，字段多，许多字段都可以按数字或字母顺序排序。在大多事实数据库中，排序输出是打印前的典型操作。

(3) 虽然大多数事实数据库可能提供有主题检索手段，但与文献数据库相比，主题检索的作用较为次要，而名称检索更为重要和流行 另外,由于事实数据库本身是由‘ 纯情报 ’组成,无需像书目数据库那样提供文献出处。

(4) 事实数据库能提供非常多的检索点。由于用户对原始情报需求的多样，而且对原始情报的查找是直接输入原始数据特征，而寻求在记录中的匹配。因此，凡是记录中看来对用户有用的事实情报元（指能单独构成一个事实或检索点、有完整意义的数据）都有可能成为检索点，提供检索，这就使事实数据库的检索途径非常多。同时，根据用户的需求，事实数据库能提供不同事实情报元的不同组合的输出，这样输出形式和功能也会有很多。

(5)查准率较高，而查全率较低。事实数据库的检索中，检索输入的是直接的原始数据，输入后，只有和原始数据匹配的数据才能被检索出来，这样就使得其查准率较高，而查全率就较低了。

4. 多媒体数据库

多媒体数据库是相对于传统的仅支持单一媒体的数据库而言,将图像、图形、文字、动画、声音等多种媒体数据合为一体,并统一地进行存取、管理和应用的数据库。多媒体数据库可以对客观世界存在的对象所表现出来的信息进行更生动、更全面的保存和查询。用户除了能浏览相应对象的文字描述，还可以观看其对应

的图形或动态的图像，听到对应的声响。提供多媒体数据存储和检索功能的系统将是一个更直观、更真实的科学数据库系统。

1.2.2 按数据库的结构模型划分

1.层次模型数据库

层次模型是用树型结构来表示数据之间联系的模型。它把客观问题抽象为一个严格的自上而下的层次关系。在这类数据库结构中，有且仅有一个无父结点的根结点，其它结点有且仅有一个父结点，它反映了数据间一对多的联系。

层次结构模型的数据库具有层次分明、结构清晰的优点。它适合于描述客观存在的事物中有主、细目之分的结构关系。其缺点是只能反映数据间单一的“一对多”联系。

2. 网状模型数据库

网状模型是用网状结构来表示数据之间联系的模型。在这类数据库结构中有一个以上结点无父结点；同时至少有一个结点有多于一个的父结点。它能反映现实世界中较为复杂的多对多的关系。

网状结构模型的数据库表达能力强，它既能反映数据间的纵向联系，又能表达数据间的横向联系。缺点是数据库在概念、结构和使用上都比较复杂，而且对计算机的硬件环境要求较高。

3. 关系模型数据库

关系模型是用二维表格数据来表示数据之间联系的模型。这种模型的数据库既能反映属性间一对一的联系，又能反映属性间一对多的联系，还能反映属性间多对多的联系。

关系模型的数据库具有数据结构简单，概念清楚，符合自然习惯，而且其格式单一，一律为表格框架，通过公共属性可建立关系之间多种的联系。

4. 面向对象模型数据库

面向对象数据库是近年来人们把面向对象技术与数据库技术结合起来研究和开发的一种新型的数据库理论。其基本原理是把传统数据库中的数据和对数据的操作过程一体化，组合成易于赋予语义的对象，并把它作为最基本的处理单位。

面向对象的数据库具有表示和构造复杂对象的能力，能够有效地应用于计算机辅助设计、计算机辅助制造、音频信息及视频信息的存储、检索与显现等管理过程。因此，面向对象的数据库将成为未来数据库技术发展的一个重要方向。

1.2.3 数据库的其它类型

除了上述各类数据库外，还可以按其它不同的标准，把数据库分为多种类型。

按数据库内容的专业划分有：综合性数据库、专业性数据库和专类性数据库等。

按数据库的载体类型划分有：磁带数据库、磁盘数据库、光盘数据库和联机数据库等。

按数据库所含语种划分有：单语种数据库和多语种数据库。

1.3 数据库建设的意义

1.3.1 数据库建设是信息产业的核心

信息、计算机和通信网络是现代信息社会的三大支柱产业，计算机和通信网络是硬件，而信息是软件，是核心。硬件设施可自己研制开发，也可以引进购买，但是一个国家的信息资源，是很难买来的，必须靠自己开发建设。数据库是信息资源采集、处理、存储和传输的枢纽，在信息资源网中，占有举足轻重的地位。没有数据库,建立起来的通信网络和计算机网络等‘信息高速公路”,无‘车”跑,没‘货’拉,只能成为摆设品,无法发挥应有的效益。

需要全本请访问 www.ertongbook.com

1.3.2 数据库建设是图书馆自动化管理的标志

实现计算机管理是图书馆自动化的重要标志之一。而图书馆要实现计算机化管理，作为管理对象的图书文献必须数字化，建立文献数据库。在此基础上图书采访、编目、流通、查询等项目的自动化才成为可能。一定规模的文献数据库和完备的书目数据，能有效深化和完善图书馆的服务功能，如提供多途径检索，提高检索效率,从而提高服务效率 提供定题服务、跟踪服务等专项服务 提供联机检索，甚至可以为读者提供建立团体或私人藏书书目数据库等项服务，使图书馆的服务突破传统的手工管理的局限。

1.3.3 数据库建设是文献信息网络化的基础

图书馆网络化就是通过馆际协作与业务辅导关系，有计划地将分散在各地区、各系统的各种类型图书馆组织起来，在集中领导下统一行动,从而形成一个既有分工、又有协作,纵横交错、脉络贯通的图书馆体系。图书馆网络化是图书馆发展的必然趋势。构成图书馆网络的主要特征之一，就是拥有相应的文献数据库，这是网络资源中的主要组成部分。数据库的建立，使图书馆在网络上既能接受外来的信息，又能向网络传递本数据库的信息数据，形成双向信息流动，完全符合信息高速公路的要求，为图书馆进入信息高速公路做好积极的准备工作。

1.3.4 数据库建设是信息资源共享的前提

建设图书馆网络的目的是实现文献资源共享和图书馆之间其他业务的合作和协作。文献资源共享必须以书目资源共享为前提，各种书目数据库的产生，使文献信息资源共享和馆际合作成为可能。高质量的书目数据库，不仅有利于国内的文献资源共享，还可以进入世界，推动世界范围的文献资源共享和国际书目控制。

1.3.5 数据库建设是军队军事训练信息网建设的主体工程

当前全军军事训练信息网建设工程已经全面启动，这是我军军事训练、教育现代化的一个重大步骤。全军军事训练信息网建设的主体工程是军事训练信息资源的数字化建设，而信息资源数字化建设的核心内容就是信息数据库的建设。这是全军军事训练信息网发挥效益的关键因素，是实现信息资源共享的必要条件，是网络用户的迫切要求。按照总参首长指示和全军军事训练信息网领导小组的统一部署，全军正在建设一批特色鲜明的军事训练信息数据库，将全军部队训练信息、院校教育信息、文献信息进行分析梳理、整序加工和数字化处理，形成一个结构优化、功能强大的信息资源体系和信息服务体系，提高部队训练和院校教育信息资源服务保障能力。

1.4 数据库建库模式

根据参与数据库建设的部门的数量，数据库建库可分为集中型、分散型、集中与分散并行等几种模式。

1.4.1 集中模式

集中型建库模式就是由信息服务部门本身或是在国家赞助扶持下集中一定技术力量建设数据库，并为用户提供服务的模式。这种建库模式最大的特点是整体性较强，数据库的建设从全局出发，较易形成大型、综合性的数据库，对于国家的信息产业规模化发展具有很大的推动作用。另外，由于是对数据进行集中处理，在很大程度上能保证数据的质量。尤其是在一些文献基础工作薄弱的国家，缺乏专业的信息服务人才，选择集中建库的方式可以在一

定程度上弥补这一缺陷。

然而，集中型建库也存在着不少问题，主要表现在：

(1)学科覆盖面窄。设备与技术力量的集中，使得数据的收集受到了限制，不可能做到包罗万象，使得数据库的覆盖领域往往只能涉及几个有限的学科，难以满足教学科研和社会的需要。

(2）难以形成有专业特色的数据库。在科技高速发展的今天，人们特别需要各种专、精、深的数据信息。由于集中建库模式受建库人员知识结构的局限，在这方面有很大的不足，难以做到每个数据库在专业上达到精深的程度，数据库很难形成特色。

(3)建库速度缓慢。集中建库将大量的建库任务集中在少数建库人员身上，建库力量有限，人手不足，往往影响到建库的速度。

(4)无法形成庞大的用户群。每一种事物从出现到被社会承认都要经过认识、认可、熟悉使用的阶段，在这一过程中，由于集中建库模式涉及的领域窄，很难做到在社会上大面积推广，其用户面有限，影响数据库的商品化。

1.4.2 分散模式

分散建库模式就是各信息服务部门各自建立自己的数据库，对信息进行加工处理，向用户提供信息服务。目前我国许多专业单位自建的许多数据库都是采用这种分散的建库模式建立起来的。

(1)分散建库模式往往是有关专业单位建设自己专业所需的数据库，而且都由本专业人员参与建库，因而其建设的数据库一般具有较强的专业特色，大部分数据库都提供了全文检索。

(2)分散建库所建的专业数据库，在专业领域内能够具备信息覆盖面广、信息收集全的特点，检索的可信度较强。

(3 充分发挥了社会各方面建设数据库的积极性，有利于数据库建设的快速起步，迅速形成国家和社会整体的数据库群。同时

可以在一定程度上为国家节省开发信息产业的费用，建立特色齐全、内容全面的数据库。

(4) 分散建立的数据库大多是本着满足本单位需求的原则进行建设的，针对自身的需求收集信息数据，利用专业技术优势进行数据库建设，因此数据库的针对性、实用性强，能够较好地满足本单位用户的利用需求。而且自建的数据库在使用手续上也更方便,更便宜。

分散的建库模式也存在许多不足：

(1)从全局来看，建成的数据库太分散，大部分数据库都处在自产自用的状态，无法形成大型的数据库。

(2) 标准化程度不高。各建库单位各自为政，遵循自己的建库标准行事。各数据库之间缺乏兼容，甚至无法进行联机检索，不能达到互补的效应，缺乏整体的信息提供能力，最终无法实现信息资源的网络化，不利于数据库的长远发展。

(3) 不利于建成大型综合的数据库，使众多的数据库停留在小、散状态，降低了数据库的价值，最终难以形成数据库建设的规模化、产业化。

(4) 数据库的建设缺乏整体的规划和分工，容易造成封闭的“小而全”模式,各信息服务部门从本身的需求出发建设数据库,难免出现大量的重复浪费。

1.4.3 集中和分散并行模式

集中和分散并行的模式是指统一规划、分担建库、集中处理、分散服务的模式。利用这种模式，在国家和行业的统一规划下，各信息服务部门分工负责，系统地收集本单位或本单位专业范围内的信息资料，利用他们在技术和资源上的优势，大力开发自产数据库。同时以一定范围（如地区或专业领域）为限设立信息协调中心，对各数据库的建设加强协调管理，按统一的标准对各单位建立

的数据集中加工处理，融合成较大规模的数据库群。然后再分散提供给用户使用。如我军军事训练信息网的文献信息数字化工程就是采用了集中管理分布式建库的建设模式。

集中和分散并行的建库模式是一种比较合理的建库模式，它集中了集中与分散建库的优点，摒弃了不合理的部分，其主要优势表现在以下几点：

(1)有利于形成合理的数据库分布。

(2）有利于调动各方面的建库积极性。

(3)克服了信息数据涉及领域窄的缺点

(4)标准化程度高，适合国际信息产业的发展。标准化程度的高低是走向信息资源网络化的关键，影响着数据库建设的方向和步伐。集中和分散并行建库模式在调动各方建库积极性的同时，能够较好地规范各建库单位的技术标准，做到统一标准，使数据库建设从一开始就步入标准化、网络化的轨道。

(5）节省人力、物力和财力，避免多家数据库重复收录造成的浪费。在系统内各建库单位明确分工，明确各自所建数据库收录信息数据的范围。这样使得各数据库所收集数据既无交叉重复的现象，又能有效避免遗漏，使科技力量以及信息资源得到最佳的搭配。不仅信息数据收集齐全，而且所耗用人力、物力和财力也将是最小的。

1.5 数据库技术和数据库产业的发展

1.5.1 国外数据库技术和数据库产业的发展

事实已经充分说明，数据库技术和产业的发展是与计算机技术、通信技术的发展应用密切相关的。从本世纪 50年代开始到现在，国际上数据库产业的发展大体经历了三个阶段。

1.50－60年代的形成时期

50 年代初期，随着计算机的商品化及其推广应用，在传统的文献工作中开始出现了数据库的概念和技术，例如 1951 年美国联邦调查局运用成套穿孔卡片所建立的数值数据库。然而，给予数据库产业以重要推动的，是按字母顺序排列的、以磁存储器为载体的书目数据库。60 年代初，这种原用于文摘和索引生产的数据库开始被改造为计算机控制的照相排版系统。 1960 年，美国国家医学图书馆着手设计其 MEDLARS 系统，1964 年使用该系统成功地进行了医学文献的批处理检索。其后，美国化学文摘服务社、美国国家科学基金会、生物科学情报服务社、工程索引公司等机构相继都开发了相应的数据库，并以磁带的形式发行。 1969 年美国国会图书馆发行了 MARC 书目磁带。特别是洛克希德的 DIALOG 系统和 SDC 的 ORBIT 系统，成为两个著名的大型联机数据库系统。

这一时期数据库的检索是以批处理方式脱机进行的。这种方式是将用户的提问累积到一定的数量，成批地运行多个检索课题和检索策略。这对于大多数用户的追溯性查找来说，花销很大，速度缓慢，而且在检索进行中无法实验检索策略和检验检索结果，查找结果常常难以令人满意。

2.70 年代成熟时期

进入 70 年代后，美国有更多的政府部门开始生产数据库，数据库的数量迅速增长，到 1975 年总数已超过 300 个，80 年达到 600 多个。数据库的内容也逐渐趋于多样，涉及自然科学、社会科学、人文科学以及人们普遍关心的事项或大众化的课题。所开发的商业数据库，包括销售、金融、经济、工业厂商名录，以及诸如房地产、保险业务等各种专业的课题。在数据的形式上，非书目数据库日益受到关注，各种指南和参考性数值数据库相继问世，全文数据库开始出现并日渐增多。

由于计算机技术和磁盘技术的发展，1970 年前后批处理脱机

检索方式逐渐向联机检索方式转化。 1971 年美国空军的 NASA 实验室使用修改过的软件，在纽约州的医学图书馆，对 MEDLARS 数据库最早进行了联机实时检索。洛克希德公司也开始为多家数据库生产者提供 DIALOG 服务，到 1974 年，该系统提供公共联机服务的数据库已达到 18 个。联机检索可对磁盘上的资料随机访问和采用交互方式修改检索策略，消除了批处理脱机检索的盲目性，得到了广大数据库用户的极大欢迎。

在此同期，欧洲情报产业界也开始认识到“联机革命”，数据库生产和联机检索，在英国、法国、联邦德国等国家迅速发展起来。

3.80 年代以来的发展时期

80 年代以来，全世界的数据库产业迅猛发展。由于更多的商业性机构进入数据库市场，呈现出日益增加的拥挤和竞争，1980－1984 年的 4 年间仅美国数据库总量就由 600 个一跃而超过 2400 个，翻了两番。同时，美国对数据库的垄断局面逐渐开始被打破，日本和欧洲的一些国家，独立自主地建立自己的数据库产业和联机产业，不依赖于他国，取得了巨大的成就，其中法国生产数据库的机构超过 100 个；日本至 1985 年，已拥有 1290 个数据库，其中 1/6 为本国所生产。另外，前苏联、澳大利亚、加拿大等国家也都拥有了很多自主发展的数据库。

这一时期，计算机技术、通讯技术和大容量信息存储技术的发展，使数据库联机检索跨入了网络化、国际化的崭新阶段。世界上最大的联机系统—DIALOG 系统的数万个终端遍布世界各大中城市。而且随着个人计算机销售量的激增，针对生活、家庭的数据库和联机服务应运兴起，用户一伸手便可以检索到远在地球另一端的收录有数千万条信息数据的数据库，在某种意义上，实现了世界范围的信息资源共享。数据库已成为现代信息社会的重要支柱。

1.5.2 我国数据库建设的发展

1.我国数据库建设的三个阶段

我国数据库建设起始于 70 年代中后期。发展至今，大致经过了 3个阶段。

第一阶段是 1974 年 – 1979 年。这一阶段的特点是引进、学习和借鉴国外数据库理论和成果，特别是引进和解决汉字处理技术。1974 年，中国科技信息研究所引进日本高千穗 TK70 汉字处理系统,并在 TK70 小型计算机上，开始试建我国的文献数据库和检索系统，成为我国数据库产业建设的一个起点。此后， 1977 年 1 月，在国家计委支持下，中国科技大学和国家计委计算中心联合在安徽省黄山召开了我国首届“数据库技术研讨会”。对于推动我国数据库建设的起步，进行了理论的准备，对数据库产业的发展探索了道路。

第二阶段是 1980 年 – 1993 年。1980 年,化工信息研究所等率先正式开始研究和建设第一批中文数据库。 1983 年由中国人民大学萨师煊教授主编的我国第一本《数据库系统概论》出版。同年清华大学倡议并牵头，组织了全国 360 多所高等学校参加研建“中国高等学校学报文献库”。特别应该提出的是 1986 年全国计算机应用工作会议召开，会议确定了“七五”期间国家要重点建设包括国家经济信息系统、国家科技信息系统在内的 12 个国家级信息系统的建设目标。此举为我国数据库建设营造了政府支持的气候和环境,此后,全国各地、各部门、各行业、各种类型的数据库,诸如文献型、事实型、数据型数据库,纷纷涌现。自 1986 年至 1988 年，每年新建数据库数量达到 50 – 100 个，其中以 1987 年为最。统计数字表明,我国‘七五’期间,在数据库建设方面总投资达 10 亿人民币,至 1991 年全国共建成 800 多个数据库。但是从统计中也可以看到，这一时期各行业数据库建设的发展很不平衡，建成的

需要全本请访问 www.ertongbook.com