

经全国中小学教材审定委员会2004年初审通过
普通高中课程标准实验教科书

信息技术 · 选修 4

数据管理技术

SHUJU GUANLI JISHU

祝智庭 主编



中国地图出版社

数据挖掘技术

李海英 编著

清华大学出版社

经全国中小学教材审定委员会2004年初审通过
普通高中课程标准实验教科书

信息技术 · 选修 4

数据管理技术

SHUJU GUANLI JISHU

祝智庭 主编



中国地图出版社

本套教材主编：祝智庭
本套教材副主编：刘观武 任友群
本册教材主编：祝智庭
本册教材副主编：任友群

普通高中课程标准实验教科书

信息技术·选修4

数据管理技术

祝智庭 主编

中国地图出版社 出版

北京市宣武区白纸坊西街3号 邮编：100054

网址：<http://www.sinomaps.com>

北京市北关闸印刷厂印刷 新华书店发行

开本：890毫米×1240毫米 16开 印张：6.75 字数：156千字

2005年1月第1版 2006年6月第4次印刷

ISBN 7-5031-3813-9/G·1512

定价：12.66元（含1张CD-ROM）

版权所有 侵权必究

图书与光盘如出现质量问题，请及时与本社联系。

编写说明

本套教材根据教育部《普通高中技术领域课程标准》(信息技术部分)编写,供高中阶段学习使用。本套教材共分6册:必修模块为《信息技术基础》,选修模块依次为《算法与程序设计》、《多媒体技术应用》、《网络技术应用》、《数据管理技术》和《人工智能初步》。

整套教材以“知识引领、活动穿插;任务引领、知识渗透;工具支持、资源配套;评估跟进、形式多样”为编写思路。从解决学生日常生活、学习中的实际问题入手,运用信息获取、加工、管理、表达与交流的基本方法,在以主题活动、探究性学习等多种形式的学习过程中逐步提升学生的信息素养,从而实现知识与技能、过程与方法、情感态度与价值观三个方面的培养目标。为了支持学习和创作表达的过程,有利于发展性评价,教材中引入了基于网络环境的“电子学习档案袋”。每册教材配学习光盘(CD-ROM),以便于使用光盘中的资源,且在教材中作了相应的提示。通过信息技术学习网站(<http://itedu.tjjy.com.cn>)提供互相交流的平台,并及时更新和拓展教学资源。学习的测评由电子学习档案袋、电子作品和在线考试三种方式组成,网站提供了上传电子作品的应用软件、电子学习档案袋安装软件和支持在线考试的相关资源。

本册教材为选修4模块,供36学时使用。

本套教材由华东师范大学教授、博士生导师祝智庭任主编,特级教师刘观武和华东师范大学副教授任友群博士任副主编。

本册教材主编祝智庭,副主编任友群,编者马艳、刘姝弘、王旭卿、毛黎莉、高淑印、齐国英。

欢迎广大师生通过电子邮件(infotech@sinomaps.com或tjjy@tjjy.net)与我们交流,提出意见和要求,指出差错或不足,共同推动信息技术课程教材建设。

天津市教育教学研究室

中国地图出版社

2004年6月

前言

为什么要学习数据管理技术？答案其实很简单。

自然的演化与社会的发展都在不断地产生着无穷无尽的信息，人类的进步要求我们处理越来越多的数据。快速处理大量数据的需求甚至已经成为我们日常生活的一部分，而以数据库技术为代表的数据管理技术就是这样一种简单高效地处理数据的技术。

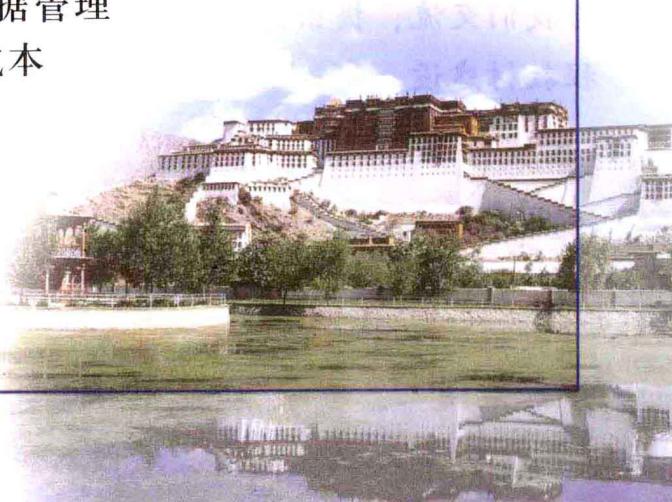
学习数据管理技术是一个无止境的过程。以前，设计数据库都是专业人员的工作，他们通常是具有数学、计算机科学或系统设计背景的、经验丰富的程序员。而在这里，我们作为初学者，将学习怎样适当地设计、实现和应用数据库。这既是数据管理技术进步的结果，也是社会发展对我们提出的新要求。我们不必担忧，但我们必须努力，就像图灵曾说过——我们的目光所及，只能在不远的前方，但是可以看到，那里有大量需要去做的工作。

在本书的学习过程中，我们将一起开展访谈，设置情境，扮演角色，合作攻关；一起营造学习环境，使用信息技术工具完成具有挑战性的任务。我们还会借助光盘和网站等丰富资源开展学习。

我们的学习不是单纯的记忆和练习，而是要通过完成一系列具体的任务，提升信息素养。在这里，我们对数据管理技术的学习是任务驱动式的学习。我们一起在做中学，接受一个又一个任务的挑战。我们要用数据管理技术来处理诸如濒危动物和世界遗产等大量信息，我们还要对完成任务的情况进行自我评价和互相评价。

应该看到，所有使用的数据库工具都是用于特定目的的。一旦我们理解了某个工具如何使用，那么，不管该工具用在什么情况下，它都将以类似的方式工作。书中的智慧拓展、经验坛、即时充电等都是我们的脚手架，而我们最终的目的是脱离脚手架而独立攀登。

记得一位先哲说过：“如果造物主在造物之前咨询我的话，我会建议用更简单的方法。”数据管理技术就是这样的方法。因此，当我们完成本书所有的任务时，我们就可以用这句话来回答开头的问题了。



目录

第一单元 认识数据库——世界遗产在中国

1

第一节 初探数据管理	2
第二节 数据模型和数据规范化	10
第三节 进入数据库	18
单元小结	28

第二单元 建立数据库——“濒危动物”数据库

29

第一节 设计数据库	30
第二节 建立数据库	41
第三节 应用结构化查询语言	52
单元小结	60

第三单元 应用数据库——“濒危动物”数据库应用系统

61

第一节 设计数据库应用系统	62
第二节 数据库程序设计基础	69
第三节 制作数据库应用系统	82
第四节 测试与评价	98
单元小结	101

中英文对照表

102

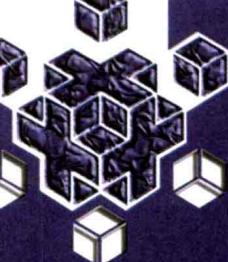
第1

单元·认识数据库 ——世界遗产在中国

我们与数据有着千丝万缕的联系。你是否有过这样的经历？在通讯录众多的人名地址中费力地搜寻朋友的电话号码，在数量众多的照片中翻找自己需要的那张照片，自己最喜欢听的歌曲文件却不知放到了计算机的哪个角落……在这个信息社会里，如何有效地管理数据，为我们的学习和生活提供便利，是我们需要去关心和面对的。

本单元的学习目标：





第一节 初探数据管理

虽然数据无处不在，可是我们了解数据吗？知道数据是如何被管理的吗？数据管理会为我们带来什么样的好处呢？通过本节的学习，我们将对数据和数据管理有所认识，了解数据管理技术的发展历史，掌握数据库和数据库管理系统的概念及其相互关系，并通过查询“在中国的世界遗产”这一主题活动，进一步体验数据管理。

在《信息技术基础》的学习中，我们已掌握了管理数据的一些基本方法，比如，使用资源管理器、专门的管理软件（如迷你通讯录）和数据库等管理数据。下面，我们将系统地学习有关数据和数据管理的知识，进一步体验数据管理的作用及方法。

一 认识数据和数据管理

1. 数据

随着社会的发展，日常生活和工作中需要处理的数据急剧增加，我们每天都会接触到大量的数据，比如图书馆中有关图书的大量数据、超市中成千上万种商品的数据、各单位关于人员和财务方面的数据、学校中教师和学生的有关数据、企业中产品生产和销售的数据等。其实，我们上网收发邮件、看新闻、听音乐，也是在和数据打交道。

数据是对客观事物属性的描述。但是，数据在未被处理之前其具体含义是不易被看出的。我们来看图 1-1-1 中的数据。

从表面上看，我们说不出它有什么实际意义。“16”表示什么呢？它是地区编号还是物品代码？我们无从知晓。但是，如果我们以某种方式对它进行处理，它便具有了实际的意义，从而成为对我们有用的信息。表 1-1-1 中的数据被处理成了天气预报的表格，这时，数据便转换成了信息，这些数据也发挥了它们的作用。因此，数据必须经过某种方式的处理才有意义，才能转化为信息。

16 7 <3

图 1-1-1 数据示例

表 1-1-1 数据转换成信息

中央气象台城市天气预报：10月 23 日 20 时至 10 月 24 日 20 时

城市	天气现象		温度(℃)		风向风力
	夜间	白天	最高	最低	
北京	多云	小到中雨	16	7	<3 北

数据的表现形式多种多样，除了上例中的文字和数字以外，还有图形、图像、声音和视频等形式。例如，上面的天气预报既可以用文字来表示，也可以通过语言（声音）来播报，还可以使用视频（天气预报节目）形象直观地加以表达。

2. 数据管理

数据被加工的过程一般称为数据处理，它主要指数据的收集、存储、加工和传播等活动。要从大量的数据中方便地获取信息，仅仅进行加工还不够。例如，要从许多气象预报中查找某一个城市的天气预报，通过上例中的表格去查找会很困难。因此，我们还应该对数据进行有效的管理，比如，建立一个气象信息数据库，将这些气象信息存储在数据库中，这样查询起来就非常方便了。

天气预报的例子从一个侧面说明了进行数据管理的目的之一——方便查找和检索信息。下面的两个案例，可以让我们进一步体会到数据管理的优势所在。



案例 奶牛场的数据管理

数据管理可以为奶牛场提高奶牛的产奶量提供支持。首先，通过输入数据将生产中的大量数据保存起来，然后，使用计算机进行计算和分析结果。例如，可以对当日产奶情况的数据进行分析，然后通过程序来优化和设计奶牛饲料配方；也可以根据奶牛饲养标准和每头奶牛的体重、产奶量和乳脂率等数据计算奶牛每日的饲喂量，并可对饲喂的营养是否符合营养标准进行验证评定。此外，还可以进行奶牛选配指导和奶牛育种值的评估等。通过对奶牛场数据的有效管理，以往很多难以了解的问题，得到了迅速反馈，从而为管理者快速、准确地作出生产决策提供了依据。



案例 中国地质调查局的数据管理

中国地质调查局为了查找矿产资源，需要对大量地质信息进行综合处理、分析和评价。传统方法是由许多专家根据自身的经验进行人工评估。人工评估不仅周期长，不利于及时发现和开采矿产资源，也无可避免地带有主观色彩，甚至出现判断失误，造成大量经济损失。为此，早在20世纪80年代，地质行业就开始使用计算机进行数据管理，通过对大量的地质信息进行自动处理和评价，从而帮助人们预测哪些地方最有可能蕴藏着矿产资源。

从上面的案例中不难看出，数据管理的主要目的是获取信息和分析信息，以指导我们的行动或帮助我们作决策。数据管理涉及对各种数据进行分类、组织、编码、存储、检索和维护等一系列活动，它是数据处理的中心问题。至于采用什么样的方法和技术进行数据管理，则与数据管理的目标、数据量的多少和数据的形式等多种因素有关。

二 数据的计算机管理

1. 数据管理技术的发展

人类进行数据管理已经有很长的历史，最早是使用手工管理的方法，如手工编制图书目录。后来，开始使用一些机械设备和简单的电子设备，例如，使用穿孔卡片和卡片处理机进行人口统计。计算机出现以后，人们才找到快速、高效和全面管理数据的现代化工具。利用计算机进行数据管理的技术经历了三个阶段：人工管理阶段、文件管理阶段和数据库（DataBase，简称DB）管理阶段。



单元 ◆ 认识数据库 —— 世界遗产在中国

第一阶段：20世纪50年代中期以前是人工管理阶段，也称自由管理阶段。这一阶段，计算机主要用于科学计算，没有操作系统和管理数据的软件。因此，处理的数据无法保存，应用程序自己管理数据，数据不具有独立性，无法实现数据共享。程序与数据之间的对应关系如图1-1-2所示。

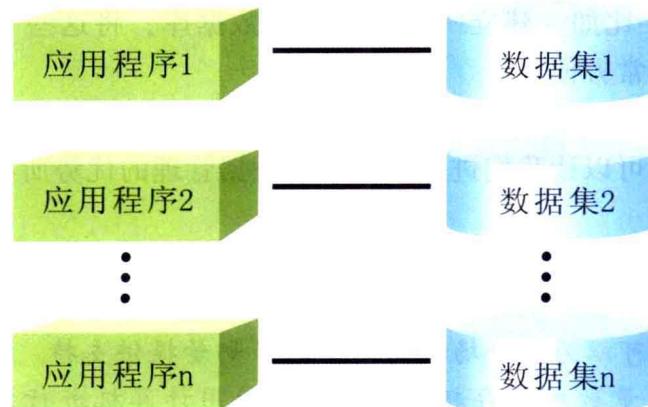


图1-1-2 人工管理阶段应用程序与数据之间的对应关系

第二阶段：20世纪50年代后期到60年代中期进入文件管理阶段。随着计算机硬件性能的改进和软件技术的发展，计算机开始大量地应用于数据处理工作。操作系统中已经有了专门的数据管理软件，一般称为文件系统。文件系统将数据组织成相互独立的数据文件。程序与数据之间由文件系统提供存取方法进行转换，使程序和数据之间有了一定的独立性，如图1-1-3所示。但是，数据文件的结构仍依赖于程序，一个文件基本上对应于一个应用程序。当不同的应用程序使用部分相同的数据时，也必须建立各自的文件，而不能共享相同的数据，因此，造成数据的重复存储，浪费存储空间。

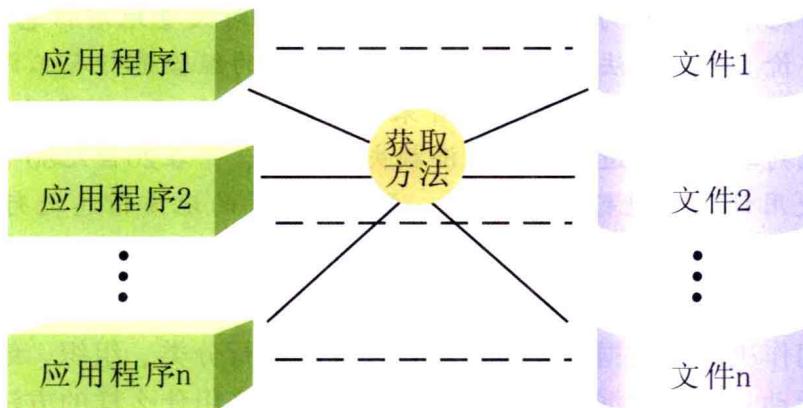


图1-1-3 文件管理阶段应用程序与数据之间的对应关系

第三阶段：20世纪60年代后期以来，计算机用于管理的规模越来越大，数据量急剧增长，为了解决多用户、多应用共享数据的需求，数据管理技术应运而生。从文件管理到数据库管理，标志着数据管理技术的飞跃。用数据库管理数据具有明显的优点，主要表现在：

- ◆ 数据的共享性高，数据冗余度低。此阶段，数据不再面向某个应用而是面向整个系统，因此，数据可以被多个用户和多个应用共享使用，共享性大大提高。由于数据被共享使用，故减少了数据冗余，节约了存储空间。

◆ 提高数据的一致性和完整性。采用人工管理和文件管理时，由于数据被重复存储，很容易造成数据的不一致。在数据库中，由于数据可以共享，因而提高了数据的一致性和完整性。

◆ 数据独立性高。数据与程序相对独立，即使数据的物理存储和逻辑结构改变了，应用程序也不用改变。这样简化了应用程序的编制、维护和修改等工作。

数据库管理阶段应用程序和数据之间的关系如图 1-1-4 所示。

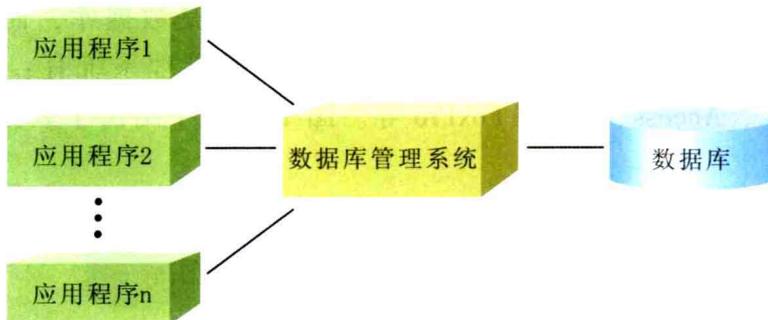
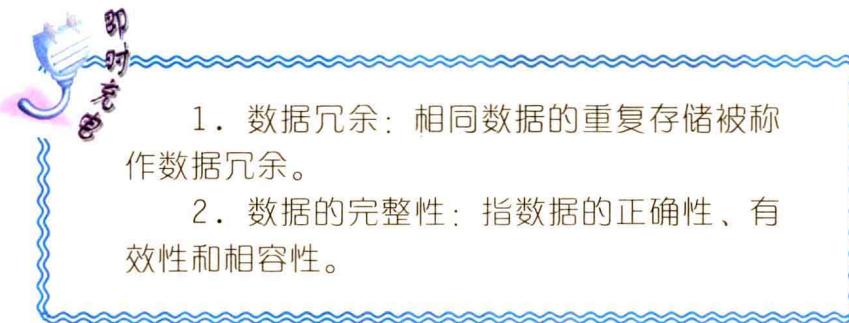


图 1-1-4 数据库管理阶段应用程序与数据之间的对应关系

由于数据库具有上述优点，目前，数据库已经普遍应用于各行各业，数据库技术得到了更为广泛的应用。



数据管理技术是信息技术发展的基础。以数据库系统为代表的数据管理技术，既满足了社会发展对信息技术的需求，又推动了信息技术的迅猛发展。但是，随着信息技术的普遍应用，非结构化内容数据（如文档、音视频流等）的大量增加，已有的数据管理技术已远远不能满足应用需求，新的数据管理技术和方法正在不断涌现。相关的知识和内容可以参看配套光盘“知识扩展”栏目。

2. 数据库

通常，我们将数据库理解为大量的、可共享的、相互关联的数据的集合。数据库中的数据按照特定的形式存储、组织和使用，它们具有较低的冗余度、较高的数据独立性和扩展性，并且可以为用户共享。比如，一个学校所有教师的信息（姓名、性别、任教年级、所教科目）和学生的信息（学号、姓名、性别、年级、班级）等，构成了需要处理的数据；这些数据按照一定的规则组织在一起并为用户共享，即构成了数据库。

3. 数据库管理系统

数据库管理系统（ DataBase Management System，简称 DBMS）是对数据库中的数据进行管理和控制的软件。DBMS 使用户可以方便地建立和使用数据库，它还负责数据的安全性保护和完整性检查、数据的访问控制和查询优化、数据库自动恢复等。DBMS 的主要功能包括以下几个方面：

- ◆ 数据定义：包括定义数据的长度、数据的类型和数据的运算等。
- ◆ 数据操纵：实现对数据库的基本操作，如插入、查询、删除和更新等操作。
- ◆ 数据库的运行管理：控制数据库的运行和用户的并发访问。
- ◆ 数据库的建立和维护：包括数据的输入和转换，数据库的存储、恢复和性能监视等功能。

数据库和数据库管理系统是两个不同的概念，常常有人将它们混淆起来。例如：经常有人说 SQL Server 是数据库，事实上，SQL Server 并不是数据库，而是一个数据库管理系统，它是用来管理数据库的。目前较流行的数据库管理系统有：SQL Server、Oracle、Sybase、MySQL、Access、Visual FoxPro 等。图 1-1-5 中给出了数据库和数据库管理系统之间的关系。

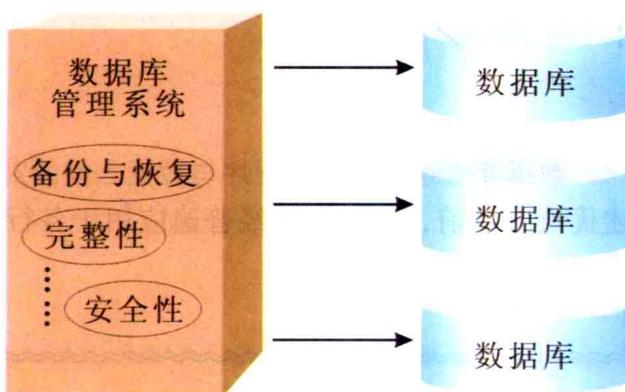


图 1-1-5 数据库和数据库管理系统的关系



知识拓展

多媒体数据库

20世纪80年代以来，许多新的数据类型（图像、音频和视频等）的处理、管理和应用越来越受到人们的重视。为了对这些新的数据类型进行管理和应用，需要建立相应的数据库，即包含多种媒体的多媒体数据库。与传统数据库相比，多媒体数据库不仅用于存放文本数据，还用于存放音频数据、静止图像数据、视频与动画数据等多种不同媒体及其整合的数据。例如，多媒体教材库就是一种多媒体数据库，它所存放的数据为用于教育的各种教育信息，其中，教学大纲和学生的评价数据多为文本信息，学习的记录除了包括文本信息外，还包括图像、音频和视频等各种媒体信息。

分布式数据库

分布式数据库（Distributed DataBase，简称DDB）就是计算机网络环境中各结点中的数据库的逻辑集合。所谓“分布式”，有两方面的含义：一方面是指数据库中的数据不是存储在同一台计算机中；另一方面是指数据库中的数据在逻辑上是相互联系的，是一个整体。分布式数据库的一个典型应用是银行转账，如图1-1-6所示。

这个应用需要将一个分行账户中的若干金额划转到另一个分行的账户中，即：从一个结点的数据库划转到另一个结点的数据库中，故需要同时更

新两个结点上的数据库。这种涉及两个或两个以上结点的应用称为全局应用或分布应用。支持分布应用是分布式应用系统一个最主要的特征。分布式数据库不仅要求数据的物理分布，而且要求这种分布是面向处理和面向应用的。此外，分布式数据库还应该具有独立处理的能力，即：可以对本地数据库执行某些应用。

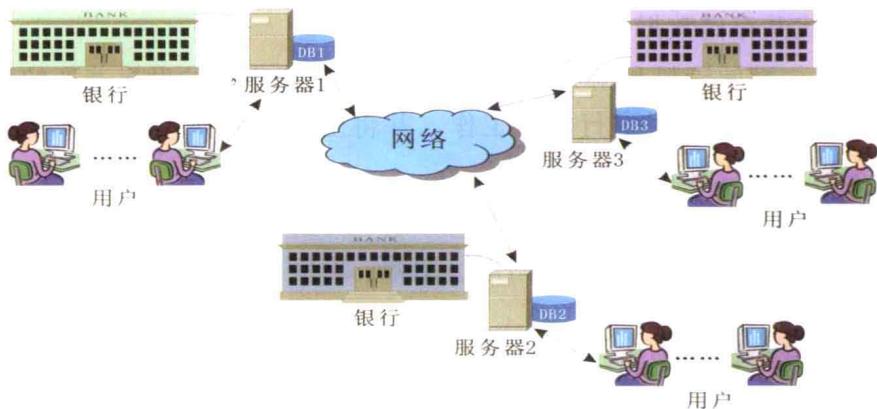


图 1-1-6 银行分布式数据库系统的示意图

数据挖掘

随着数据库技术的不断发展及数据库管理系统的广泛应用，数据库中存储的数据量急剧增大，在大量的数据背后隐藏着许多重要的信息。数据挖掘就是从大量的实际应用数据中，提取隐含在其中的、人们事先不知道但又可能有用的信息和知识的过程。数据挖掘的主要目的是提高市场决策能力，检测异常模式，在过去的经验基础上预言未来趋势等。例如，通过对大量气象资料和销售资料的处理及分析，德国的啤酒商发现，夏天气温每升高1℃，就会增加230万瓶的啤酒销量，而日本人则发现，夏季30℃以上的天气每增加一天，空调的销量便增加4万台。另外，前面提到的中国地质调查局的数据管理的案例也是数据挖掘的一个典型应用。



查找和管理数据——在中国的世界遗产

我国有着五千年的悠久历史，自然的馈赠、祖先的余泽使我们成为名副其实的世界遗产大国。图1-1-7和图1-1-8中展现了我国两处著名的遗产——布达拉宫和黄山的迷人风姿。下面，我们就通过完成以下活动来了解我国世界遗产的主要情况，进一步体验数据管理。

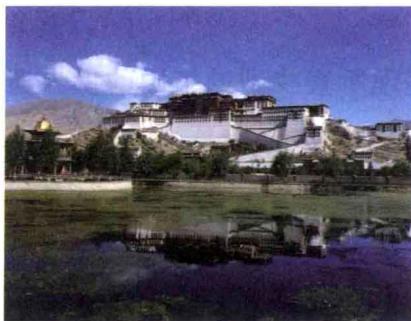


图 1-1-7 布达拉宫



图 1-1-8 黄山

1

单元 ◆ 认识数据库 — 世界遗产在中国



查询在中国的世界遗产的有关情况。

四至五人为一个小组，确定各自的任务，分工协作，按照以下步骤开展活动。

步骤 1：查找并存储数据

通过上网搜索、翻阅资料和浏览光盘等多种途径查找中国的世界遗产的相关资料，并存储在计算机中。查找数据的同时进行如下工作，并将工作结果存入电子学习档案袋的“我的资料”中，以“定稿”方式发布。

- ◆ 分类记录数据的来源（如网络、图书、报刊和电视等）。
 - ◆ 记录获得的数据类型（文本、图形、图像、音频和视频等）。



中国世界遗产网	http://www.cnwh.org
中国全景网	http://www.chinavr.net
文通网	http://www.culturalink.gov.cn
世界遗产专题集邮网	http://www.guwh.com
联合国世界遗产中心	http://whc.unesco.org
配套光盘素材	可以运行配套光盘中的中国景点及世界遗产数据库应用系 (kbsc\1\d1\vichan\browser.exe) 查询有关资料, 应用系统界面如图 1-1-9 所示。

配套光盘素材 可以运行配套光盘中的中国景点及世界遗产数据库应用系统(kbsc\d1\vichan\browser.exe) 查询有关资料，应用系统界面如图 1-1-9 所示。



图 1-1-9 中国景点及世界遗产数据库应用系统

步骤 2：组织和管理数据

选择一种合适的数据管理方法，分类整理所查找到的中国的世界遗产资料，并将整理结果存入电子学习档案袋的“我的资料”中。

供参考的分类方式如下：

- ◆ 按照世界遗产类别分类。
 - ◆ 按照所在省份分类。
 - ◆ 按照批准日期分类。



世界遗产可以分为自然遗产、文化遗产、自然和文化双遗产、文化景观四类。另外，还有一类是口头和非物质世界遗产。一般以前面四种分类为主。

步骤③：汇报与讨论

各组制作汇报报告并存入电子学习档案袋的“我的感受”中，以“定稿”方式发布。各组汇报工作成果，介绍本组采用的数据管理方法，分析其优缺点，然后进行如下讨论：

- ◆ 整理出来的有关世界遗产的数据可以为谁提供服务？
- ◆ 哪种数据管理方法可以使我们通过网络方便地进行数据检索和查询？
- ◆ 如果管理所有的世界遗产资料，采用哪种数据管理方法更合适？

步骤④：小组之间的评价

参考表 1-1-2 给出的评价标准，各组进行互相评价，并将评价结果存入电子学习档案袋的“我的感受”中，以“定稿”方式发布。

表 1-1-2

任务完成情况评价表

评价项目	优秀	中等	较差
确定分类整理项目情况	围绕任务主题，自主确定恰当的整理项目	参考教材给出的项目，自己设计部分整理项目	没有自己设计整理项目，完全按照教材确定整理项目
收集和保存数据的情况	从多种电子和非电子的渠道收集数据，并正确地标明了出处；资料保存完好且有备份	从少量的电子和非电子渠道收集信息；资料保存完好	只从单一的渠道收集数据，资料保存不完全
组织和管理数据的情况	学生为有效管理数据，自己开发了数据库，数据排序和检索方便	学生使用电子表格、文件夹等方式管理数据，数据检索基本方便	数据排放无序，查找数据不方便
汇报与表达	提交的资料完整，有效地使用综合媒体以多种方式展示了自己的成果，并发布到网上；语言表达清楚、准确	提交的资料较为完整，使用综合媒体以多种方式展示了自己的成果；语言表达比较清楚、准确	提交的资料不完整，仅用单一媒体展示了自己成果（如书面报告）；语言表达不够清楚、准确

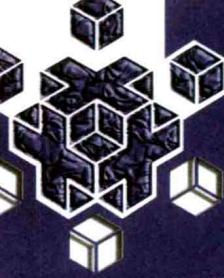
1. 数据和信息有什么不同？文件管理和数据库管理的主要区别是什么？

2. 本周你做了哪些直接或间接涉及数据库的事情？如果没有数据库，本周你的生活会有什么不同？

3. 访问中国国家图书馆 (<http://www.nlc.gov.cn>)，你认为这个站点是由数据库支持的吗？如果是，你认为这个站点的哪些特性和功能借助了数据库技术？你还能发现类似的站点吗？



实践与思考



第二节 数据模型和数据规范化

什么是数据模型？数据规范化的目的是什么？怎样实现数据规范化？通过本节的学习，我们将了解数据模型的含义及四种常用的数据模型，掌握实体、属性和联系等基本概念及实体间的三种联系，并通过实例分析，明确数据规范化的思想。

在实现任何类型的数据库之前，都需要为即将存储的数据设计一个正确的模型。数据模型为构建物理数据库描绘了蓝图，它以一种非常明确和简洁的方式清晰地勾勒出数据资源的需求和关系，从而帮助我们了解与数据需求有关的所有事实。根据一个设计良好的数据模型构建的数据库，可以实现最低限度的数据冗余、最大限度的数据完整性和更好的数据共享等。

一 数据模型初探

1. 数据模型

数据模型（data model）用来描述数据库中数据存储的方式。按照不同的应用目的，数据模型可分成三种类型，即概念数据模型（conceptual data model）、逻辑数据模型（logic data model）和物理数据模型（physical data model）。

概念数据模型又称概念模型，它侧重于描述客观世界的事物及事物之间的联系，而与具体的数据库管理系统和计算机平台无关。下面，我们讨论概念数据模型，它是整个数据模型的基础。

概念模型中涉及以下几个基本概念：

- ◆ **实体**（Entity）：客观存在的可以相互区别的事物，如一家音像店、一张CD唱片等。实体也可以是比较抽象的事件，如购买一张CD唱片。同型实体的集合称为实体集。例如，全部CD唱片就是一个实体集。
- ◆ **属性**（Attribute）：实体所具有的某一特性。例如，音像店实体可以由音像店的名称、地址和电话等属性组成。
- ◆ **联系**（Relationship）：实体之间的联系或实体内部各属性之间的联系。例如，顾客与音像店之间存在着买卖联系，音像店与CD唱片之间存在着销售联系。

通常，我们使用实体-关系图（E-R图）来表示概念模型。实体-关系图使用基本的图形符号来表示数据及数据之间的关系。图1-2-1画出了音像店销售CD唱片的E-R图。

图1-2-1中各种图形符号的含义如下：

- ◆ 矩形框：用来表示实体，框内为实体名称。
- ◆ 椭圆：用来表示属性，并用连线与实体或联系连接起来。