

上海市计算机学会组织编写

# 计算机与信息科学 十万个为什么

## 4

## 数据库与信息检索

《计算机与信息科学十万个为什么》

丛书编辑委员会 编著



清华大学出版社

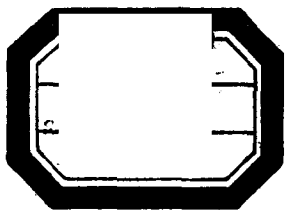
上海市计算机学会组织编写

计算机与信息科学十万个为什么

4

# 数据库与信息检索

《计算机与信息科学十万个为什么》 编著  
丛书编辑委员会



清华大学出版社

(京)新登字 158 号

J5236.08

## 内 容 简 介

本书以“十万个为什么”的体例形式简明、通俗地论述了数据库与信息检索的社会应用,数据库与信息检索系统的基本概念、基本原理、实现和操作方法,介绍了该领域的最新研究动态。具有简明、通俗、实用,并有一定的趣味性等特点。主要内容包括数据库部分 74 条,信息检索部分 50 条。

本书可供希望了解计算机应用领域中的数据库和信息检索基本知识的各行各业人员阅读,也可作为从事数据库应用的工程技术人员的参考书。

版权所有,翻印必究。

本书封面贴有清华大学出版社防伪标签,无标签者不得销售。

## 图书在版编目(CIP)数据

计算机与信息科学十万个为什么 (4): 数据库与信息检索/《计算机与信息科学十万个为什么》丛书编辑委员会编. —北京: 清华大学出版社, 1997

ISBN 7-302-02735-8

I. 计… II. 计… III. ①电子计算机-基本知识②信息技术-基本知识③数据库-基本知识④情报检索-基本知识 IV. TP3

中国版本图书馆 CIP 数据核字 (97) 第 24851 号

出版者: 清华大学出版社(北京清华大学校内, 邮编 100084)

因特网地址: [www.tup.tsinghua.edu.cn](http://www.tup.tsinghua.edu.cn)

印刷者: 北京市丰台区丰华印刷厂

发行者: 新华书店总店北京科技发行所

开 本: 787×1092 1/32 印张: 12.625 字数: 291 千字

版 次: 1998 年 1 月 第 1 版 1998 年 1 月 第 1 次印刷

书 号: ISBN 7-302-02735-8/TP·1420

印 数: 00001~10000

定 价: 14.00 元

《计算机与信息科学十万个为什么》

丛书编辑委员会

名誉主任：陈至立

名誉主编：何成武 张兆琪 李超云

主 编：张吉锋(兼主任)

副 主 任：吕传兴

常务编委：王心园 孙一霖 吕传兴 陆 皓

吴洪来 郁宝忠 张吉锋

编 委(按姓氏笔划排序)：

尹芳平 方起兴 朱益梅 刘淦澄

何守才 何礼义 张鹏飞 陈涵生

林慎机 赵怀成 徐桂珍 宣国荣

程 亮 童 颖 鲍振东

# 《计算机与信息科学十万个为什么》

## 各篇、各专题编委

篇、专题名	主 编	副主编	主 审
引路篇	郁宝忠	王心园	吴洪来
综合应用技术篇	陈一民	劳诚信 董长洪	王心园 王家增
专业技术篇			
软件与接口技术	吴洪来	徐国定	夏宽礼
计算机网络与数据通信	方起兴	谢承德 荆金华	张根度
多媒体	何守才	顾  宁	郑衍衡
计算机辅助设计	宣国荣	郁文忠	仲  毅
管理信息系统与 办公室自动化	何礼义	夏雨人	张吉锋
数据库与信息检索	陆  皓	周  宁	周广声
人工智能	童  颖	朱关铭	缪淮扣
计算机安全与保密	鲍振东	赵一鸣	覃  光

# 序

人类进入了信息时代。

随着以计算机和现代通信技术为核心的信息技术迅速发展和广泛应用,信息资源得到进一步开发和利用,大大推动了人类社会各个方面的发展,并对人们的工作、学习和生活产生了深刻的影响。

在当今世界,发达国家和新型工业化国家都在大力发展信息技术,竞相规划和建设本国的信息基础设施,加速信息化进程,力争在这场世纪之交的大竞争、大发展中立于不败之地。当前,我国也在大力发展信息技术,加强信息产业和信息基础设施的建设,以迎接信息时代对我们的挑战。

实现信息化,关键是人才。我们不但需要有一批熟悉信息系统与信息资源开发,致力于信息化建设的专家,更需要有一大批能掌握计算机与信息技术,会用并用好信息系统与信息资源的应用人才,还要在全社会普及计算机及信息知识,增强信息化意识,使人们学会并适应在信息社会环境中工作、学习和生活。

在科学技术普及的事业中,书籍有着极其重要的作用。书

籍是人类进步的阶梯。一部好的科普丛书能启迪人们的心智，激发人们进一步学习科学技术的兴趣和奋发向上的精神。早在 20 世纪 60 年代，上海的科学家们就编写了一套《数、理、化、生十万个为什么》，对推动科学文化普及做出了有益的贡献。今天，上海市计算机学会继承和发扬这一优良传统，组织了 100 多位专家学者，通过调查研究，编撰成这套《计算机与信息科学十万个为什么》丛书。丛书内容丰富，形式新颖，图文并茂，深入浅出，面向青少年和广大群众，是计算机与信息科学普及教育难得的好教材。丛书不仅能让人们获得从事计算机与信息技术所需的知识和技能，更重要的是，还能使人们从中受到科学思想、科学精神、科学态度和科学方法的教育。因此，这也是一部宣传社会主义精神文明的普及读物。

一部好书可以影响一代人。《计算机与信息科学十万个为什么》的编写出版，得到清华大学出版社以及中国计算机报社的大力支持，这是一件有意义的事情。采用“十万个为什么”的方式普及计算机知识，是一种十分有益的尝试。希望编写丛书的专家学者不断跟踪科技发展趋势，不断修改、扩充和更新丛书内容和媒体形式，使丛书越出越好，以满足广大读者的需要。

我们希望有更多的专家学者和学术团体参加科技普及工作，编写出版普及读物，传播科学知识，为推动两个文明建设，为把我国的经济建设转移到依靠科技进步和提高劳动者素质的轨道上来做出自己的贡献。

祝贺《计算机与信息科学十万个为什么》丛书出版成功。

陈至立

# 丛书前言

计算机与信息科学的普及应用是一个国家现代化程度的重要标志,计算机与信息科学的发展水平和开发能力更是现代国家国力的重要表现。党中央和国务院在制订发展我国高新技术产业政策时,把计算机与信息科学列为优先发展的领域之一。举国上下,越来越多的人认识到学习和掌握这一领域的知识和技能的重要性,为了能在激烈竞争的现代社会生活中不落人后,为了能为实现社会主义四个现代化建设做出更大贡献,利用一切可以争取到的机会,甚至动用家庭有限的收入作投资,创造条件,学习计算机科学知识和技能,在全国各地正在形成越来越高的热潮。在这种形势下,作为在计算机和信息科学领域里耕耘多年的识途老马们,自然有义不容辞的责任,为青少年和广大迫切需要学习计算机与信息科学的已入门和未入门的群众做出自己应有的贡献,帮助大家克服困难,少走弯路,尽快占领这一领域的各个高地与顶峰。

帮助青少年和广大群众掌握计算机与信息科学的基础知识和基本技能是实现科技兴国战略任务的重要组成部分。当今,计算机与信息科学早已从为少数科学家所专有演变为解



决科学、技术、工程和日常生活各方面问题的强有力的技术和工具。掌握计算机科学的原理和应用技能既可以为儿童、青少年进一步学习科学技术打下良好基础,也是广大群众掌握现代工具、提高生活技能的有效途径。为了有效地实现这一目标,最重要的是激发大家学习和掌握计算机知识与技能的兴趣和睿智,而不是简单地让大家学会几条指令的操作或某几个软件的使用,因为这些东西会随着科技的进步、版本的更新很快“过时”,唯有掌握学习的能力和探讨计算机科学技术的兴趣是长久不衰的。

基于上述原因,上海市计算机学会在清华大学出版社、电子工业部计算机与微电子发展研究中心、中国计算机报社的大力支持下,聘请了 100 多位长期从事计算机与信息科学各个领域的教学和科学研究、有丰富经验、学有所成的专家、教授,从计算机与信息科学的宝库中,选取了人们在学习、工作以及生活中经常会碰到的问题,力图用生动有趣的例子、浅显明白的道理、通俗准确的语言来回答这些问题,来描述计算机与信息科学的过去、现在和未来,真可以说是历经寒暑,十易其稿,最终编写而成这套《计算机与信息科学十万个为什么》丛书。

丛书包括“引路篇”、“综合应用技术篇”和“专业技术基础篇”三篇,分为九个分册出版。作为丛书的第一本,“引路篇”(154 个问题)提供的是学习计算机和信息科学的综合基础知识与技能,同时,也说明本篇的编写宗旨是“引路”,即起到“引进门”的作用。“综合应用技术篇”单独成册(83 个问题),主要是通过典型例子说明计算机技术是如何应用于日常生活的各个方面的。“专业技术基础篇”分为七个分册,包括以下 12 个

专题:

1. 软件基础(80 个问题)
2. 接口技术(18 个问题)
3. 数据库(74 个问题)
4. 信息检索(50 个问题)
5. 数据通信(97 个问题)
6. 计算机网络(66 个问题)
7. 管理信息系统(39 个问题)
8. 办公室自动化(52 个问题)
9. 多媒体(41 个问题)
10. 计算机辅助设计(82 个问题)
11. 人工智能(72 个问题)
12. 计算机安全与保密(61 个问题)

上述各篇各专题中问题的选取原则是,既针对青少年和广大群众当前学习的实际需要,又照顾到计算机科学日新月异的发展。编写力求做到科学性、通俗性、趣味性并重,既有原理的说明,也有应用技术的指导。考虑到本丛书的基本读者群体是青少年和具有中等及中等以上文化程度的有志学习计算机知识的广大群众,每个题目都完整地说明一个知识点,绝大多数条目在知识结构上是相对独立的,在学习时不需要以其他题目的知识作为基础,这样的安排也是为了适应业余学习间隙性的需要。

每一分册条目的编号按以下原则编排: 条目编号的形式为  $x.y$ ,  $x$  表示本书在丛书中的分册序号,  $y$  表示该条目在本书中的序号。

本丛书的每一部分都包含了少量的计算机与信息科学的

前沿知识,这种局部超前的安排,不仅是为了适应广大青少年进一步学习计算机和信息科学的需要,也有助于具有一定基础的各类专业人员扩展自己的知识面。而且,由于计算机科学及相关技术发展迅速,今天的所谓前沿知识,随着时代进步,技术发展,将成为日后大众的普及知识。

本书在编写时致力于提高读者的知识水平与学习能力,尽可能用简洁、准确的文字讲清原理,使读者在理解的基础上激发进一步学习的主观能动性,进行创造性的学习。我们认为,经常研究“为什么”,可以养成自己提出问题、解决问题的习惯,学会举一反三、融会贯通、启迪思路,一旦读者自己弄清了“为什么”的道理,一些具体操作通过相应的操作手册是容易掌握的。

希望本书能够成为青少年和广大群众前往计算机与信息科学殿堂时的登堂台阶,这就是我们的用心所在。

## 《计算机与信息科学十万个为什么》

丛书编辑委员会

1997年5月

## 编 者 的 话

数据库是计算机与信息科学中发展最快的重要分支之一,它已成为计算机信息系统与应用系统的重要技术基础和支柱。最早的数据库系统出现在 20 世纪 60 年代末,但发展到现在,数据库技术已走向成熟阶段,其重要标志是已有了大量商品化的数据库软件产品面市,广大用户成功地应用各种数据库软件来开发适合本部门应用所需要的信息系统,从而在信息科学领域中发挥了巨大作用。然而,尽管广大用户会操作使用某一种数据库软件,但对它们的一些基础知识和应用技术还缺少系统、全面和完整的理解。为了普及推广计算机知识,根据丛书的编纂宗旨,以“十万个为什么?”的形式编写数据库与信息检索的条目。在数据库部分,既着重介绍数据库的基本知识,又兼顾它们的应用和操作。尤其是人们常常容易混淆的一些基本概念,如“什么是数据库?”“什么是数据库系统?”“什么是数据库应用系统?”等结合生活实际去讲清基本原理。书中较多的条目是针对传统领域的数据库知识,但对近年来新兴领域的数据库,如知识库、多媒体数据库、面向对象数据库、演绎数据库、分布式数据库、主动数据库、工程数据库

等等,都逐一给予介绍。我们还对数据库系统的功能和特性,如数据库的共享性、独立性、安全性、完整性、故障恢复和并发控制等知识,虽然其部分内容较多偏于专业范畴,在书中还是尽可能作了通俗的说明。数据库应用领域十分广泛,我们仅对最常见的数据库应用,如管理信息系统、办公自动化系统、计算机集成制造系统等作了介绍。相信读者定能举一反三。在怎样使用数据库方面,书中举例介绍了建库、检索、更新、删除和插入等基本操作原理及方法。

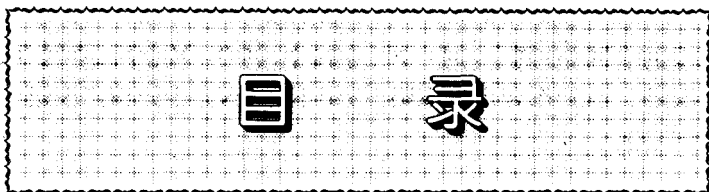
信息检索是以信息检索系统为依托来进行信息的收集、加工、存储、查找,并向各类用户提供具有权威性、时效性、准确性的信息服务,是人们在信息社会工作和生活中的重要工具,利用计算机进行信息检索已成为人们获取信息的重要方法。数据库又是信息检索系统最重要的信息源,也是系统中最基本的组成部分。因为只有采用数据库技术,才能使信息检索系统具有管理大量、持久和共享数据的能力,具有高效、准确的事务处理的能力。本书从信息检索的社会应用中选择大家经常接触到的文献检索、电子出版物、大学信息、客票、金融、旅游、城市管网等方面,用科学原理和通俗语言去叙述信息检索的功能和它们对人类生活所起的作用。同时介绍联机检索系统、光盘检索、机读目录、交互电视、电子数据交换等实现信息检索的手段。在应用中涉及到的一些信息检索的基本知识,如系统的建立、电子出版物的制作、检索方法和系统评价等,也作了相应的介绍。本书总共编写了124条。在编写深度的掌握上,大部分条目(占总数的65%)是面向中等以上文化水平的非计算机专业人员;也有小部分条目牵涉较多或较深的专业知识,我们力图把本领域中既是热点又是难点的问题,尽量

阐述清楚。希望即使不是从事数据库和信息检索领域工作的专业人员,看了也有所启示和受益。

本书由陆皓任主编,周宁任副主编,周广声任主审。参与编写的有陆皓、周宁、周广声、陈福生、林兰、郑志新、褚纪红、张驰等。由于作者水平有限,书中不妥、不当之处请广大读者批评指正。

编 者

1997年5月



## 数 据 库

### 数据库的基本知识

- 4.1 什么是数据库? ..... 1
- 4.2 为什么说数据库是现代的信息源? ..... 4
- 4.3 为什么计算机中的 0 和 1  
可以用来表达任何信息? ..... 7
- 4.4 采用数据库技术的信息管理系  
统能给应用部门带来哪些好处? ..... 10
- 4.5 数据库与数据库系统是一回事吗? ..... 11
- 4.6 人们为什么把 dBASE 称为大众数据库? ..... 13
- 4.7 FoxBASE, FoxPro 为什么  
比 dBASE 更受用户青睐? ..... 17
- 4.8 怎样正确选择你需要的数  
据库管理系统(DBMS)? ..... 18
- 4.9 为什么各种档次的计算机都要  
配置数据库管理系统(DBMS)? ..... 22

- 4.10 数据库管理系统(DBMS)与操作系统(OS)有什么关系? ..... 24
- 4.11 配置数据库管理系统(DBMS)时对计算机硬、软件环境的要求是什么? ..... 27

### 数据库的基本操作

- 4.12 不懂数据库语言的用户能否与数据库打交道? ..... 29
- 4.13 在数据库管理系统(DBMS)中怎样对数据库中的数据进行操作? ..... 32
- 4.14 怎样在数据库管理系统(DBMS)中建库和装数据? ..... 34
- 4.15 怎样修改数据库中的数据? ..... 37
- 4.16 为什么把数据库的检索、更新、删除和插入统称为数据操纵? ..... 39
- 4.17 怎样用结构化查询语言(SQL)实施数据操纵? ..... 41
- 4.18 怎样用 QBE 实施数据操纵? ..... 45
- 4.19 怎样用 xBASE 实施检索操作? ..... 47
- 4.20 怎样用 xBASE 实施更新操作? ..... 51
- 4.21 为什么在数据库管理系统(DBMS)中要有排序和索引功能? ..... 54
- 4.22 数据库管理系统(DBMS)能提供哪些统计功能? ..... 58
- 4.23 数据库管理系统(DBMS)能提供哪些报表功能? ..... 62



## 数据库应用知识

- 4.24 为什么关系数据库系统受到用户欢迎? ..... 64
- 4.25 什么是关系代数? ..... 66
- 4.26 什么是关系演算? ..... 72
- 4.27 为什么关系数据库管理系统(RDBMS)  
    必须具有选择、投影和连接运算功能? ..... 77
- 4.28 Oracle 是个什么样的数据库管理系统? ..... 81
- 4.29 Sybase 是个什么样的数据库管理系统? ..... 86
- 4.30 Informix 是个什么样的数据库管理系统? ..... 89
- 4.31 DB2 是个什么样的数据库管理系统? ..... 95
- 4.32 INGRES 是个什么样的数据库管理系统? ..... 97
- 4.33 什么是层次数据库系统和网状数据库系统? ..... 101
- 4.34 微机 DBMS 与大型机上  
    的 DBMS 有哪些不同? ..... 103
- 4.35 在数据库中为什么能检索声音、图像信息? ..... 105
- 4.36 为什么光盘数据库能得到迅速发展? ..... 107
- 4.37 什么是知识库? ..... 109
- 4.38 为什么要进行知识检索? ..... 110
- 4.39 为什么说多媒体数据库是“立体化”的数据库? ... 111
- 4.40 多媒体数据库实用化的关键是什么? ..... 114
- 4.41 多媒体数据库为什么是  
    最有发展前途的数据库? ..... 116
- 4.42 为什么说面向对象数据库  
    (OODB)是第三代数据库? ..... 119
- 4.43 什么是演绎数据库? ..... 122
- 4.44 什么是分布式数据库? ..... 125