

上海市计算机学会组织编写

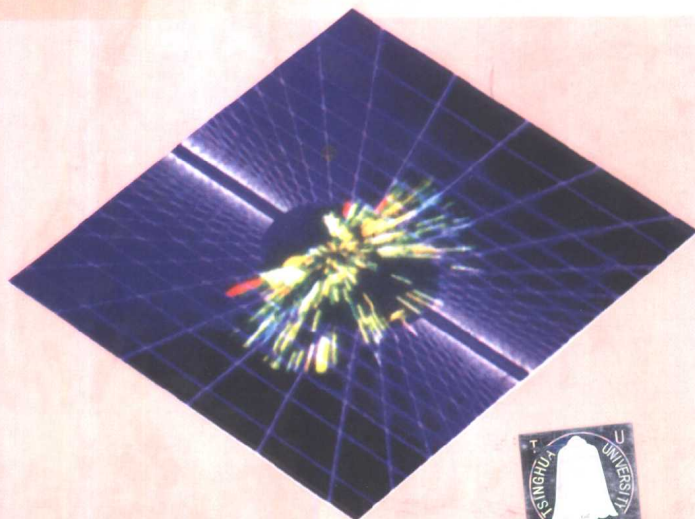
《新编计算机与信息科学十万个为什么》

丛书编委会 编著

十万个为什么

新编
计算机
与信息科学

5 数据库
与信息检索



清华大学出版社
<http://www.tup.tsinghua.edu.cn>

新编计算机与信息科学十万个为什么

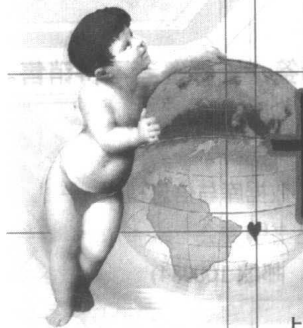
丛书编委会 编著

计算机

与

信息科学 (新编)

十万个为什么



上海市计算机学会组织编写

⑤ 数据库与信息检索

清华大学出版社

地址: www.tsinghua.edu.cn

(京)新登字 158 号

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基本概念、基本原理、系统实现和基本操作方法;介绍了该系统的主要应用和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。主要包括:数据库的基本知识、数据库的基本操作、数据库应用知识、数据库设计知识、数据仓库与数据集市、数字化图书馆、信息检索基本知识和社会应用、网络检索和信息资源管理等。

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,也可以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激光防伪标签,无标签者不得销售。

书 名:新编计算机与信息科学十万个为什么·5·数据库与信息检索

作 者:《新编计算机与信息科学十万个为什么》丛书编委会

出版者:清华大学出版社(北京清华大学学研大厦,邮编 100084)

<http://www.tup.tsinghua.edu.cn>

印刷者:世界知识印刷厂

发行者:新华书店总店北京发行所

开 本:787×1092 1/24 印张:20 字数:449 千字

版 次:2000年9月第1版 2000年9月第1次印刷

书 号:ISBN 7-302-03931-3/TP·2299

印 数:0001~5000

定 价:200.00元(全套)

《新编计算机与信息科学十万个为什么》

丛书编委会

名誉主任：陈至立

名誉主编：施伯乐 张兆琪

主 编：张吉锋(兼主任)

副 主 任：吕传兴

常务编委：王心园 吕传兴 陆 皓 吴洪来

郁宝忠 张吉锋 程耀华

编 委 (按姓氏笔画排序):

尹芳平 方起兴 孙德文 何礼义

陈一民 陈涵生 宣国荣 施鹏飞

徐桂珍 高黎新(兼秘书) 童 颖

鲍振东 张鹏飞

《新编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	高黎新	王心园
专业技术基础篇			
系统结构	孙德文	徐伟民	张吉锋
软件基础	吴洪来	徐国定	夏宽理
数据库与信息检索	陆 皓	周 宁	周广声
办公自动化与管理信息系统	何礼义	黄天敏	张吉锋
计算机网络与数据通信	方起兴	谢承德	张根度
		荆金华	俞嘉惠

篇、专题名	主 编	副主编	主 审
因特网、内联网和家庭网络	方起兴	谢承德	张根度
		荆金华	俞嘉惠
多媒体技术	施鹏飞	程耀华	郑衍衡
计算机辅助设计	宣国荣	李启炎	仲 毅
		彭澄廉	
人工智能	童 颖	缪淮扣	朱关铭
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技术专家，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。

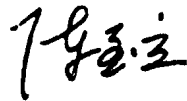
在科学技术普及的事业中，书籍有着极其重要的作用。书籍是人类进步的阶梯。一部好的科普丛书能启迪人们的心智，激发人们进一步学习科学技术的兴趣和奋发向上的精神。早在 20 世纪 60 年代，上海的科学家们就编写了一套《数、理、化、生十万个为什么》，对推动科学文化普及做出了有益的贡献。今天，上海市计算机学会继承和发扬这一优良传统，组织了 100 多位专家学者，通过调查研究，编撰成这套《新编计算机与信息科学十万个为什么》丛书。丛书内容丰富，形式新颖，图文并茂，深入浅出，面向青少年和广大群众，是计算机与信息科学普及教育难得的好教材。丛书不仅能让人们获

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。《新编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作出自己的贡献。

祝贺《新编计算机与信息科学十万个为什么》丛书出版成功。

A handwritten signature in black ink, appearing to read '陈至立' (Chen Zhi), written in a cursive style.

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。

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，最终编写成这套《新编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 12 个分册出版。作为丛书的第一本，“引路篇”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，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 10 个分册，包括以下 12 个专题：

1. 软件基础
2. 接口技术
3. 数据库
4. 信息检索
5. 数据通信
6. 计算机网络
7. 管理信息系统
8. 办公自动化
9. 多媒体
10. 计算机辅助设计
11. 人工智能
12. 计算机安全与保密

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，随着时代的进步和技术的发展，将成为日后大众的普及知识。

本丛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。

希望本丛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。

《新编计算机与信息科学十万个为什么》
丛书编委会

2000年4月

• IX •

编者的话

数据库系统是实用性很强的系统。数据库技术的进步推动着数据库应用领域的扩展和应用层次的提高；而广泛应用新技术又促进了数据库技术的进一步发展。从第一版出版到现在仅仅一年多时间内，国际数据库技术领域涌现出许多技术热点，其中之一是面向对象技术与数据库技术的结合，有了对象关系数据库系统和纯对象数据库系统。另一热点是数据仓库、联机分析处理和数据采掘，这是随着数据库应用的深化，为解决如何有效地将数据库中的大量数据变成有用的信息为决策支持服务而提出的问题。它们不仅已进入理论上深入剖析，更是在实践中建立系统，并在市场上已有众多解决方案的产品。随着 Internet 技术的发展，WWW 上可用数据源的数量迅速增加，与 Internet 有关的数据库技术应运而生，特别是在信息检索领域，如何方便快速查找因特网上的资源，成为亿万网民最为关心的问题，这样，网络信息检索工具也就层出不穷，这些新知识和新技术就成为本书第二版要扩充的内容。与此同时，我们也大刀阔斧地把第一版中一些陈旧或过时的条目加以删除或修改；对仍给予保留的条目也仔细加以审订，务求内容正确，文字通顺。本书第二版总共编写了 168 条，其中新增的 56 条，修改的 20 条，占总数 45%；而保留条目为 92 条，占总数 55%，但原书的基本宗旨和风格不变，是以“十万个为什么”的体例形式，简明、通俗地论述数据库与信息检索的基本概念、基本原理、实现和操作方法，也介绍了该领域的最新研究动态。我们是紧紧跟踪科技发展趋势，力求反映当代技术水平，充实和更新本书内容，做到既简明、通俗、实用，又有一定的趣味性等特点。

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、聂玉美、林蓉、陈传艺、王玉波等。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

2000年9月

目 录

数 据 库

数据库的基本知识

5.1 什么是数据库?.....	3
5.2 为什么说数据库是现代的信息源?.....	6
5.3 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?.....	8
5.4 数据库与数据库系统是一回事吗?.....	10
5.5 什么是关系代数?.....	12
5.6 什么是关系演算?.....	18
5.7 怎样正确选择你需要的数据库管理系统 (DBMS)?.....	22
5.8 为什么各种档次的计算机都要配置数据库管理系统?.....	26
5.9 数据库管理系统 (DBMS) 与操作系统 (OS) 有什么关系?.....	28
5.10 配置数据库管理系统时对计算机硬、软件环境的要求是什么?.....	31
5.11 为什么在建立数据库时要采用数据压缩技术?.....	33

数据库的基本操作

5.12 你是否知道与数据库打交道的工具是数据库语言?.....	37
5.13 怎样在数据库管理系统中建库和输入数据?.....	39

5.14	怎样修改数据库中的数据?	42
5.15	为什么把数据库的检索、修改、删除和插入统称为数据操纵?	45
5.16	怎样用结构化查询语言 (SQL) 实施数据操纵?	46
5.17	怎样用示例查询语言 (QBE) 实施数据操纵?	50
5.18	为什么在数据库管理系统中要有排序和索引功能?	52
5.19	数据库管理系统能提供哪些统计功能?	56
5.20	数据库管理系统能提供哪些报表功能?	60

数据库应用知识

5.21	为什么关系数据库管理系统受到用户欢迎?	62
5.22	为什么要把关系数据库管理系统 (RDBMS) 分为几类?	64
5.23	为什么关系数据库管理系统必须具有选择、投影和连接运算功能?	67
5.24	FOX 系列数据库软件为什么又添新“狐狸”?	71
5.25	Oracle 是个什么样的数据库管理系统?	73
5.26	Sybase 是个什么样的数据库管理系统?	79
5.27	Informix 是个什么样的数据库管理系统?	83
5.28	DB2 是个什么样的数据库管理系统?	87
5.29	INGRES 是个什么样的数据库管理系统?	89
5.30	MS SQL Server 是个什么样的数据库管理系统?	93
5.31	什么是层次数据库系统和网状数据库系统?	95
5.32	为什么说面向对象数据库 (OODB) 是第三代数据库?	98
5.33	什么是后关系数据库技术?	101
5.34	为什么要研制开发对象关系数据库?	104
5.35	什么是 DB2 通用数据库?	107

5.36	为什么要研制开发纯对象数据库?	111
5.37	微机 DBMS 与大型机上的 DBMS 有哪些不同?	114
5.38	在数据库中为什么能检索声音、图像信息?	116
5.39	为什么光盘数据库能得到迅速发展?	118
5.40	什么是知识库?	120
5.41	为什么要进行知识检索?	121
5.42	为什么要研究开发智能化数据库?	123
5.43	为什么说多媒体数据库是“立体化”的数据库?	125
5.44	多媒体数据库实用化的关键是什么?	128
5.45	多媒体数据库为什么是最有发展前途的数据库?	131
5.46	什么是演绎数据库?	133
5.47	什么是分布式数据库?	136
5.48	什么是主动数据库?	141
5.49	什么是工程数据库?	143
5.50	为什么说数据库管理系统是数据库的“大管家”?	146
5.51	为什么说数据库管理员不是一个人, 而是一个团体?	148
5.52	数据库系统的动态工作过程是怎样的?	150
5.53	数据库和数据库应用系统是一回事吗?	152
5.54	管理信息系统 (MIS) 是数据库应用系统吗?	155
5.55	办公自动化系统 (OA) 是数据库应用系统吗?	158
5.56	计算机集成制造系统 (CIMS) 是数据库应用系统吗?	162
5.57	数据库机是数据库机器吗?	165
5.58	为什么说 PowerBuilder 是新一代数据库应用开发工具?	168
5.59	自建的数据库如何上网服务?	171

数据库设计知识

5.60	数据库为什么一定要与数据模型联系在一起?.....	174
5.61	什么是数据管理参考模型?.....	178
5.62	为什么 SQL 语言被定为国际标准的数据库语言?.....	182
5.63	为什么网状数据库语言被定为国际标准的数据库语言?.....	184
5.64	怎样用陈氏的实体联系方法 (E-R 方法) 来设计数据库?.....	187
5.65	什么是现实世界、信息世界和数据世界?.....	189
5.66	为什么要建立数据字典?.....	192
5.67	为什么要研究查询优化?.....	194
5.68	为什么说数据转换是数据库建立与应用的基础?.....	198
5.69	为什么在不同数据存储模式的数据库之间要进行数据转换?.....	201
5.70	为什么说数据共享是推动数据库产业发展的强大动力?.....	205
5.71	什么是数据库的数据独立性?.....	207
5.72	什么是数据库的可靠性?.....	209
5.73	在数据库管理系统(DBMS)中保证数据库(DB)安全可靠的措施是什么?.....	211
5.74	什么是数据库的完整性?.....	214
5.75	怎样使数据库在发生意外故障后能迅速有效地恢复?.....	217
5.76	为什么在数据库管理系统中要有并发控制?.....	219
5.77	数据库中为什么会存在“幽灵”?.....	221

数据仓库与数据集市

5.78	什么是数据仓库和数据仓库系统?.....	224
5.79	已经建立了数据库,为什么还要建立数据仓库?.....	229