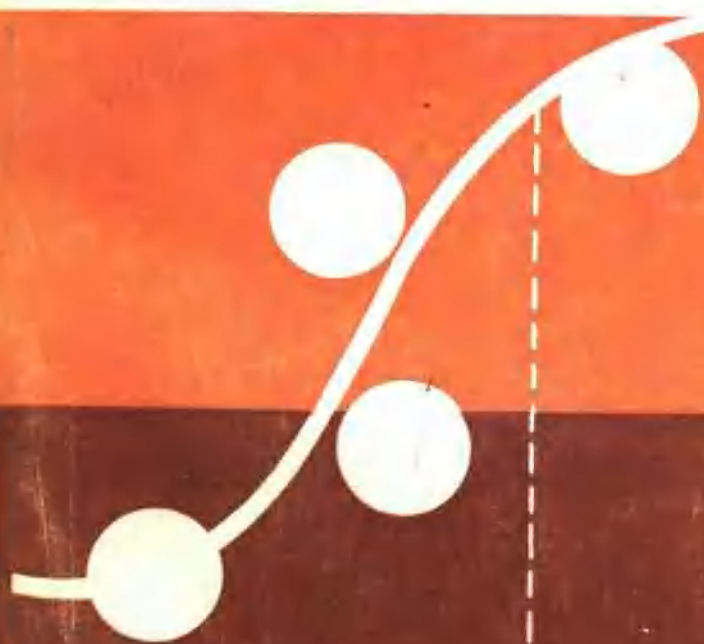


(华东地区高校计算机基础教学研究会推荐教材)

汉字dBASEⅢ、FOXBASE 管理信息系统开发技术

周莹年 郭彦发 丁 宏 李菊莲 编写



陕西电子编辑部

(华东地区高校计算机基础教学研究会推荐教材)

汉字 dBASE III、FOXBASE 管理信息系统开发技术

周莹年 郭盈发 丁宏 李莉莲

陕西电子杂志社

内 容 简 介

本书全面介绍了汉字 dBASEⅢ、汉字 FOXBASE 关系型数据库管理系统及其管理信息系统的开发技术。由于汉字 FOXBASE 与汉字 dBASEⅢ是兼容的，因而，本书首先介绍汉字 dBASEⅢ知识，最后介绍汉字 FOXBASE 的特点及其扩充功能。

全书共分十四章，其内容分为两篇：第一篇——基础篇，包括七章：前三章介绍计算机软硬件方面的基础知识，第四章至第七章系统地介绍汉字 dBASEⅢ系统配置，九类文件及其操作。第二篇——应用篇，包括七章：汉字 dBASEⅢ程序设计、汉字 dBASEⅢ管理信息系统的开发，实用性技术、多重数据库的应用技术、汉字 dBASEⅢ与高级语言的数据共享，汉字 dFORMAT 的应用、汉字 FOXBASE。

本书内容丰富，通俗易懂，各章都有大量例题，章末附有习题，书末还给出了必要的附录。

本书可作为大专院校及各类计算机培训班的教材，也适合中等以上文化程度的读者自学。本书更是企事业单位的领导干部、管理人员研究管理信息系统开发的良师益友。

前 言

汉字 dBASEⅢ是根据美国 ASHTON-TATE 公司于 1984 年 7 月推出的微型机关系数据库管理系统 dBASEⅢ汉化而成的。由于其功能齐全、直观易学、使用方便并适合我国国情，一经推出，就受到广大用户欢迎。

汉字 FOXBASE 是根据美国 FOX Software 公司于 1987 年 2 月推出的关系型数据库管理系统 FOXBASE 汉化而成的。由于它能与汉字 dBASEⅢ完全兼容而且速度快，功能有扩充，而引起各界用户的极大兴趣。

但是，以往出版的有关书籍都是介绍基本知识，基本命令，没有或极少涉及如何利用汉字 dBASEⅢ，汉字 FOXBASE 编制应用软件，开发管理信息系统这方面的内容。广大读者在编制软件，开发管理信息系统方面普遍感到困难，虽然学会了各条命令，但却无法付之实用。本书将着重介绍利用汉字 dBASEⅢ，汉字 FOXBASE 编制应用软件和开发管理系统的知识。为节省篇幅，本书首先介绍汉字 dBASEⅢ，最后介绍汉字 FOXBASE 的特点及其扩充的内容。

本书分为两篇共 14 章。第一篇是基础篇，用三章的篇幅介绍了计算机软硬件及操作系统的基础知识和操作系统的常用命令，使读者对计算机有一个总体的概念，以便于进一步了解计算机之所以能加工处理信息，并能如此神奇地完成各种各样的数值计算和数据处理的任任务。第四章至第七章是从管理信息系统数据库处理技术的角度出发，以数据库系统为横线，首先介绍数据库系统的组成、系统的运行环境及引导过程、数据库文件及其它八类文件的操作和辅助操作。这一篇的基本出发点，是使读者以数据库系统为纲来了解数据库系统的概念、功能、特点以及建立各种文件的方法和命令。这就使读者始终有一个数据库系统的概念。

第二篇是应用篇，这一篇反映了本书的重要特色。第八章介绍汉字 dBASEⅢ程序设计，包括程序的四种结构、自顶向下的结构化程序设计和程序设计风格的探讨。第九章汉字 dBASEⅢ管理信息系统的开发，用尽可能简单的方法向读者介绍管理信息系统开发的技术和要求。第十章、第十一章主要介绍数据库系统程序设计的实用性技术，包括：I/O 处理技术，查询检索技术、程序的调试和纠错技术、数据库的安全性和故障处理、程序的优化方法、数组运算的实现等。此外，还有多重数据库的应用技术。

汉字 dBASEⅢ尽管具有很多优点，得到了广泛应用，但对大量的复杂计算却力不从心，对于图形显示和访问外设更是无能为力。而高级语言既有较强的计算能力，也有较强的图形处理和访问外设的能力。于是，高级语言和汉字 dBASEⅢ配合使用，可能使编制的应用软件在数据管理、运算速度、计算能力、图形处理诸方面都取得较好的效果。本书第十二章以 BASIC、FORTRAN、PASCAL 和 C 语言等为例，介绍如何实现高级语言与汉字 dBASEⅢ的数据共享。

第十三章介绍汉字 dFORMAT 的应用，以帮助用户设计屏幕输入输出格式，并能自动地生成汉字 dBASEⅢ的格式文件(·FMT)。

第十四章介绍由最新多用户版本 MFOXBASE PLUS2.00 汉化而得的汉字

FOXBASE PLUS2.00, 简称汉字 FOXBASE。它与汉字 dBASEⅢ完全兼容, 而且速度倍增。本章着重介绍汉字 FOXBASE 与汉字 dBASEⅢ的区别, 相同部分请参阅前面各章对汉字 dBASEⅢ的介绍。

本书从计算机基础知识, 汉字 dBASEⅢ到汉字 FOXBASE 对关系数据库管理系统进行了全面而系统的介绍, 使读者得到较高层次的应用知识。我们希望读者学习本书后, 能够在研制开发管理信息系统时得到益处。

本书根据作者 85 年以来的教学经验, 在内部教材的基础上, 结合科研的经验整理而成。在编写过程中, 广泛收集了各类读者和用户的愿望、要求和经验教训, 并广泛参考了我们所能收集到的国内外参考资料。

本书的第一、二、三章由周萱年编写, 第四、五、六、九、十四章由郭盈发编写, 第十一、十二、十三及第十章第二、三、四、五、六节由丁宏编写, 第七、九章第十章第一、七、八、九节以及附录由李菊莲编写。周萱年为主编。

在编写过程中, 我们得到了杭州电子工业学院和计算机系领导的支持, 学院计算机中心部分同志提出了一些宝贵意见。在此, 一并表示衷心的感谢。

由于编者水平有限, 加之时间仓促, 错误和不妥之处在所难免, 敬请读者批评指正。

周 萱

一九八九年四月

目 录

第一篇 基础篇

第一章 计算机基础知识	1
第一节 数据和信息	1
第二节 计算机硬件基本概念	12
第三节 运算速度的有关因素	20
习题	21
第二章 计算机语言	22
第一节 机器语言	22
第二节 汇编语言	23
第三节 高级语言	24
第四节 计算机的硬件和软件	27
习题	28
第三章 操作系统	29
第一节 操作系统的发展过程	29
第二节 研究设计操作系统的几种观点	39
第三节 操作系统常用命令	
习题	54
第四章 汉字 dBASE III 基本知识	55
第一节 数据库系统的组成	55
第二节 汉字 dBASE III 及其系统文件	56
第三节 汉字 dBASE III 运行环境及其引导过程	57
第四节 用户文件	58
习题	61
第五章 汉字 dBASE III 数据库文件及其操作	62
第一节 数据库文件	62
第二节 字段的五种类型	64
第三节 数据库文件的数据存贮结构	66
第四节 命令的一般形式及常用符号约定	67
第五节 全屏幕编辑状态	68
第六节 数据库文件的建立	69

第七节	数据库文件的打开与关闭	70
第八节	记录指针及其移动	71
第九节	多个工作区	73
第十节	增加记录	74
第十一节	库文件的编辑	76
第十二节	库文件的查询	80
第十三节	统计与汇总	82
第十四节	数据库文件内容的输出	83
习题	86
第六章	汉字 dBASE III 其它八类文件及其操作	88
第一节	记忆文件	88
第二节	索引文件	90
第三节	命令文件	94
第四节	格式文件	95
第五节	内存变量及内存变量文件	106
第六节	标签格式文件	111
第七节	报表格式文件	114
第八节	文本文件	118
习题	119
第七章	汉字 dBASE III 辅助操作	121
第一节	常量表示法	121
第二节	表达式	121
第三节	函数	123
第四节	文件操作命令	130
第五节	几条通用命令	137
第六节	状态设置命令	138
第七节	用 CONFIG.DB 文件设置参数	149
第八节	显示系统状态命令	151
习题	152

第二篇 应用篇

第八章	汉字 dBASE III 程序设计	156
第一节	自顶向下设计程序模块	156
第二节	数据库文件的设计	158
第三节	汉字 dBASE III 程序的建立和执行方法	161

第四节	程序的顺序结构	163
第五节	程序的分支结构	164
第六节	程序的循环结构	177
第七节	子程序调用与菜单设计	188
第八节	主程序与子程序之间的直接数据交换与内存变量的属性	194
第九节	程序设计风格探讨	206
习题	209
第九章	汉字 dBASE III 管理信息系统的开发	213
第一节	管理信息系统的规模及建立条件	213
第二节	管理信息系统的开发过程	215
第三节	系统分析	215
第四节	系统设计	220
第五节	系统实施	222
第六节	系统运行	224
习题	225
第十章	实用性技术	226
第一节	数据库文件的应用	226
第二节	输入的设计技巧	234
第三节	输出报表的处理技术	236
第四节	查询检索的设计技巧	243
第五节	数组运算的实现	247
第六节	自动生成报表的通用程序	248
第七节	程序的调试与纠错技术	254
第八节	程序的优化技术	260
第九节	数据的安全性及故障处理	263
习题	270
第十一章	多重数据库操作	272
第一节	工作区的概念和规则	272
第二节	SET RELATION 命令的应用	274
第三节	UPDATE 命令的应用	277
第四节	JOIN 命令的应用	278
习题	280
第十二章	汉字 dBASE III 与高级语言的数据共享	282
第一节	数据共享的实现方法	282
第二节	数据库文件与文本文件的数据交换法	284

第三节	BASIC 语言程序与文本文件交换数据法	288
第四节	FORTRAN 语言与文本文件的数据交换法	293
第五节	PASCAL 语言与文本文件的数据交换法	294
第六节	C 语言与文本文件的数据交换法	296
第七节	外部程序的调用——RUN 命令	297
习题	298
第十三章	汉字 dFORMAT 的应用	299
第一节	CdFORMAT 的运行环境及技术指标	299
第二节	CdFORMAT 的组成和工作流程	299
第三节	CdFORMAT 的操作说明	300
第四节	应用举例	306
第十四章	汉字 FOXBASE	309
第一节	基本性能	309
第二节	功能扩大的命令	315
第三节	新增加的命令	319
第四节	功能扩展了的函数	325
第五节	新增加的函数	326
第六节	编译汉字 FOXBASE 的命令文件	335
第七节	多用户汉字 FOXBASE	337
第八节	配置文件 CONFIG.FX / .DB	342
附录一	dBASE III 技术指标	346
附录二	ASCII 码字符表	347
附录三	汉字 dBASE III 命令集(字母序)	348
附录四	常用 CCDOS 命令	353
附录五	汉字输入方法简介	354
附录六	汉字 dBASE III 全屏编辑控制键及其功能	360
附录七	EDLIN 使用简介	362
附录八	汉字 WORDSTAR 使用简介	364

第一篇 基础篇

第一章 计算机基础知识

第一节 数据和信息

一、数据和信息

所有计算机，包括大、中型和小型的，其基本工作原理是相同的。任何计算机系统提供的工作环境都取决于它的硬件和操作系统。

一个计算机系统可以说是一个信息和数据处理系统。所谓信息，其含义是非常广泛的，它与术语“数据”不同。信息是反映事物的状态和特征的，是客观世界各种事物状态的事实的集合。客观世界在不断地运动和变化，事物的状态和特征也在相应地变化，因而就不断地产生信息。人们通过获取的信息来区别不同的事物，区别事物的发展和变化，从而认识事物改造事物，所以说信息是人类认识世界改造世界的知识源泉。

人类文明寓于“信息革命”之中，信息革命对人类社会各方面的发展都有重要影响。人类社会已经经历了四次信息革命：

第一次信息革命是语言的发展。当我们的祖先学会说话的时候，他们就获得了一种比只是做手势及发简单声音更为高明得多的表达思想及概念的手段。他们所学到的一切传给他们的孩子们，从而使人类知识的积累一代代地增长起来；语言是人类智力成就的根源，而且是人类不同于其它动物的标志。

第二次信息革命是书写的发展。书写文件远比人的记忆更精确和持久，而且容易复制和传递。书写使信息可靠地保存很长的时间，而且能使两个从来没见过面的人互相交换信息。

第三次信息革命是中世纪末期印刷术的发明带来的。在此以前，信息是难以广泛传播的，虽然已经有了书，但每册付本都是手抄的，因此书就成了宝贵而稀有之物。印刷术完全改变了这种状况。书籍、杂志、报纸成批地生产，很多国家有文化的人的比例已经很大。印刷品到处被用来传递信息，给人知识、给人以娱乐以及用来影响和支配人。

第四次信息革命是由于计算机的产生和发展而发展起来的。计算机出现才 40 多年，然而它已经使得整个工业显得面貌陈旧，而且在其它方面也带来根本变革，几乎每个人都波及到了。

信息是人类社会一种宝贵的软资源。能否大量地有效地利用信息是社会发展的标志之一。从信息论的观点看，古代社会和现代社会的最大差别就在于信息利用的多少；信息传递的速度和信息处理的质量。古代生产力水平低，仅利用到很少量的信息，而现代社会由于生产力水平高，科学技术发展迅速，能充分地大量地利用信息。特别是电子计算机、高级技术及大规模和超大规模集成电路的发展，提供了处理各种形式信息的有力工具。所以信息不仅是国家、社会和全体人民的宝贵资源，也是科学技术发展进步的途径。

信息和数据这两个概念常常容易混淆不清，在一个数据处理应用系统中，信息和数据

都具有它们自己的特征，在系统设计中要很好地考虑这些特征。计算机是一种信息处理机。对于一个人机系统，信息属性可以归纳如下：

1. 目的性

信息必须有目的地传给人或机器。人们在事务中的活动是各种各样的，因此需要的信息也是多种多样的，信息必须有目的地有选择地传给人或传给机器才有其意义和价值。

2. 时间性

信息是有时间性的，在一定时间之内它是有价值的，过了这段时间它就失去了意义。从时间来分信息可以分成历史的信息、现行的信息及未来的信息三类。有了历史的信息，根据现行的信息可以预测将来的信息。信息的时间性就涉及到信息的保存时间，重要的信息就作长期保存；暂时有用的信息，就作短期保存。

3. 可靠性

当统计评价参数时，要求参数的实际值必须在选定概率所确定的范围之内，才能保证信息的正确性。

4. 精确性

不同的决策对信息要求的精确度不同，有的只要求知道信息的范围，有的要求数量精度。如财务管理要求信息的精度就很高，而市场预测就不可能做到精确的要求。

5. 发生率

不同的事务信息的发生率不同，有的信息的发生率高且有规则，有的信息发生率低而无规则，而且发生率与时间有关，有高峰期，有低峰期。根据信息的发生率，安排信息的处理形态及传输速度。

6. 适用性

收集信息的范围、深度、广度是否合适，就涉及到信息处理用的结果是否满足使用要求，这就是适用性。

其它的属性还有密集性、相关性、组织性、期待性。对于机器来说信息的属性还有静态和动态，线性和非线性，连续和分离等。这些信息的属性在系统设计时，在信息利用时都必须很好地加以考虑。

数据是描述事物状态特性的特定符号。符号的形式有各种各样，归纳起来可分两类：一类是数值化的数据如实验数据、统计数据、凭证数据、十进数据、二进制数等。另一类是非数值化的数据如文字、图象、报表、声音等等。

数据在数据处理应用系统中有它的生命期(Life cycle)，如图 1-1 所示。在企业的环境中，对企业内部和外部事物的观察记录，经验描述，收集产生数据，然后把它存贮在文件中或数据库中。如果收集、记录的数据是无价值的或经过校验是错误的，则将它消灭，这

个数据的生命周期也就结束。如果收集或记录的数据是正确的，有用的，则数据就在它的生命周期中从一个过程传送到另一个过程。还可以根据需要对数据进行计算、分类、综合、加工、排序，最后提供有用的信息，有的还需要再反馈到输入，直至结束它的生命周期，数据“消灭”为止。

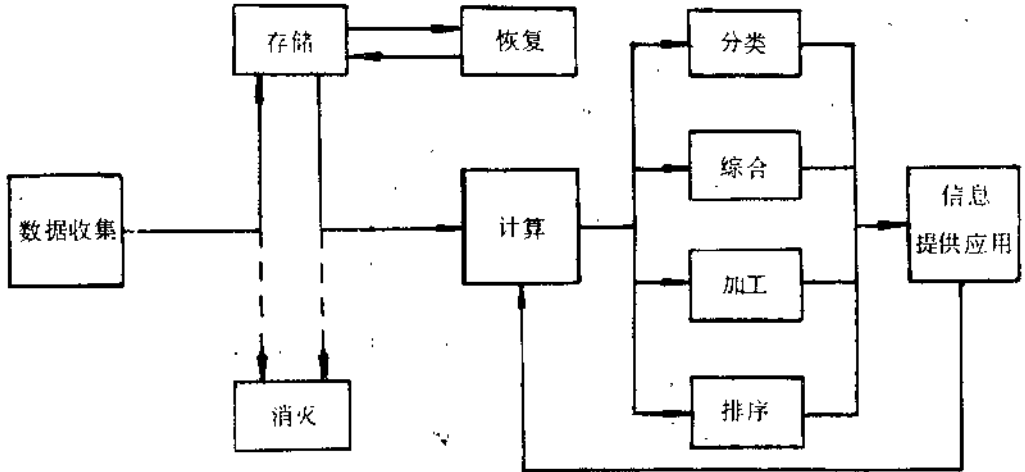


图 1-1 数据的生命周期

信息和数据的关系，对于一个数据处理应用系统来说，客观存在的事实即信息，可以用符号序列来表示，这种符号序列可以是字符串或数字串，这种字符串或数字串就称为数据。数据是用以载荷信息的数字、字母和符号。这就是信息和数据的关系。

用计算机来处理各种形式的信息，首先要把信息代码化。通过记录信息代码的纸带、卡片和源程序一起输入计算机，然后由“翻译”系统将它们转换为二进制的数字即机内代码——这就是一种特定的数据，计算机再对它进行运算处理。所以一切形式的信息在计算机内都是以特定的数据来表示的。而经过加工处理后的数据再转换或被解释成为更有用的信息。

例如一个人事登记表如表 1-1 所示

表中每个人的编号、姓名、性别、年龄、学历、职务都是用文字和数字记录下来的数据。姓名、性别、学历、职务是非数值数据，而职工号、年龄则是数据型的数据。

表 1-1 人事登记表

职工号	姓名	性别	年龄	学历	职务
0001	李名	男	30	大专	助工
0002	江卫东	男	48	大学	讲师
0003	张丽	女	26	大学	助教
:	:	:	:	:	:

计算机进行信息处理时，按“输入——处理——输出”的模式进行。例如用计算机对上述人事登记表进行处理，要求分别统计出一个单位男女职工人数，必须把人事登记表代码化后输入到计算机，经过分类统计，最后输出统计结果，这就是用户所需要的信息。计算机进行分类统计的过程就是信息处理的过程。而且“处理”的过程又分为若干步骤等等。其中每一步骤中都有明确的要求，“处理”就是按这些要求进行。所以，计算机好比一个黑匣子，这个黑匣子有一个入口和一个出口。从入口“输入”必要的信息，从出口“输出”所要得到的信息，而“处理”过程是在黑匣子(即计算机)中实现，见图 1-2。

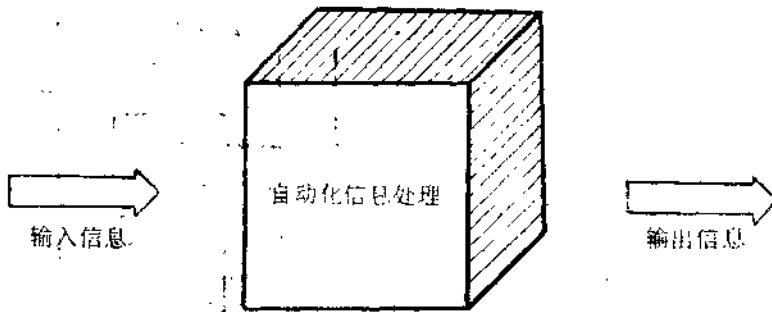


图 1-2 计算机的功能

如何实现计算机的信息处理呢？必须编制软件(程序)。在编制软件(程序)时，重要的是要确定什么是输入，需要处理的是哪些内容，需要输出的又是什么。适当理解和选择 I. O 将避免数据处理系统中很多麻烦。

二、数据的编码(数据表示)

数字电子计算机所处理的对象就是数字，或者说是那些已经用数字进行编码表示的信息。既然这样，我们就有必要谈谈数的表示系统(简称数制)。

在所有的计算机中，数是用一种特殊的方式来进行编码表示的，这就是二进制制。其简称二进制。

也许大家会问：人们不是已经用惯了通常的十进制数了吗？为什么还要用别的方法来代表计算机中的数？这是由计算机的内部结构决定的。计算机所能存贮的信息的最小单位是一个二进制数位，通常称为一个比特(bit)。比特是“二进制数位”英文 Binary digit 的简称。一个比特的数据可以是“0”或“1”，也可以表示“开”或“关”，“真”或“假”。所以一个 bit 可以用来叙述同一概念的两种不同状态。一个比特可以用称作触发器或存贮细胞的电子元件存贮起来。即一个比特就是一个触发器或存贮细胞所能代表的信息量。但是，计算机的程序和数据往往包含很大的信息量，而这些都得由计算机的存贮器存放起来。特别需要指出的是，每个数字或一个字符，往往需要一组 bit 来进行存贮和处理。如表 1-2 中的例子。

在微型计算机中，上述一组(即 8 个)bit 作为一个单位叫做一个拜特(Byte)，也叫字节。由表 1-2 已经看到，除数字外，一个 Byte 可以存放任意一个英文字母的代码，这种编码的方法叫做 ASCII 码。由于一个字节是由 8 个 bit 组成的，因此，可以表示 256 种不同的组合。但是 ASCII 码(American Standard Code For Information Interchange 美国

表 1-2 现实的数和字母以及它们在计算机内所存储的代码

现实世界的数字或字母	计算机里所存储的代表数字和字母的代码
+127	01111111
-128	11101000
D	01000100
I	01001001
M	01001101

标准信息交换码的简称)规定每个字节的最高位为奇偶校验位, 所以 ASCII 字符集能够表示 128 个字符, 每个字符对应于一个字符码。字符码表示该字符集中字符的排列顺序(即次序号)。ASCII 码字符集见附录二。

计算机内所采用的电子元件或运算决定了计算一般都要采用二进制或八进制、十六进制, 但是, 使用计算机的人类本身却不是采用二进制, 而是采用十进制, 在人和机器之间最好有一个联系的“桥梁”。所以, 一般计算机所输入的数据是用十进制, 由计算机把它们转换成二进制, 再进行运算, 运算完以后又变回到十进制输出。作为用户来说, 可以不了解计算机内部结构以及进制的转换。但是, 作为一个当代科学的忠实读者来说, 为了使自己对计算机工作的基本原理有大概的了解, 还是必须学一点数与数制较换的知识。下面仅列出几个表加以说明, 有兴趣的读者可以翻阅“计算机原理”方面的书。

表 1-3 八进制中所用的数字集合与二进制、十进制的比较

二进制	八进制	十进制
0	0	0
1	1	1
	2	2
	3	3
	4	4
	5	5
	6	6
	7	7
		8
		9

表 1-4 八进制中的不同数字及它们代表的三位二进制的数字组

八进制数的数字集合	所代表的三位二进制数字组
0	000
1	001
2	010
3	011
4	100
5	101
6	110
7	111

表 1-5 十进制、十六进制数字集合及其所代表的四位二进制数字组

十进制的数字 集 合	十六进制的数字 集 合	所代表的四位二进制 数 组
0	0	0000
1	1	0001
2	2	0010
3	3	0011
4	4	0100
5	5	0101
6	6	0110
7	7	0111
8	8	1000
9	9	1001
	A	1010
	B	1011
	C	1100
	D	1101
	E	1110
	F	1111

三、汉字在计算机中的存贮及 I/O 时字型转换

(一) 简述点阵表示法的设计原理

1. 西文字形点阵表示法的设计原理

点阵表示法是随着屏幕显示技术的发展而发展的。

目前屏幕显示技术均采用电视扫描方式，按固定的机内时钟，由显示控制器周期地扫描屏幕缓冲区，根据屏幕缓冲区单元每一位的 0 或 1 信息，产生加到 CRT 阴极射线管栅极的脉冲，控制电子枪在屏幕上作相应点迹扫描，以显示图象。

屏幕缓冲区的内容的变化，会引起屏幕上图象显示的变化。如果将屏幕看成一个大的点阵，则该点阵即可分割成若干个显示字母(字符)形状的小点阵。例如：一般的显示屏幕，都是高分辨率的绿色磷 P31 不反光荧光屏，其图象分辨率为 640×240 个象素。

假如将屏幕点阵分割成一组组小点阵，每个小点阵为 7×8 或 8×10 个象素。每个点阵中显示一个西文字母(或字符)的形状，则每个字母(或字符)可设计为 5×7 或 7×9 的点阵。其余的部分正好留给光标显示的位置和作为字母与字母之间的间隔。由于屏幕上每行 640 个象素，共 240 行象素，即前面所说的屏幕上图形分辨率为 640×240 个象素，而每个西文字母的小点阵占 8×10 个象素。因此，整个屏幕上可以显示 24 行，每行 80 列西文字符。见图 1-3 和图 1-4 所示。

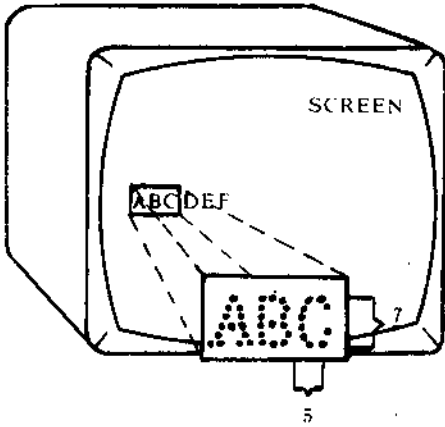


图 1-3 屏幕显示 5×7 点阵字符

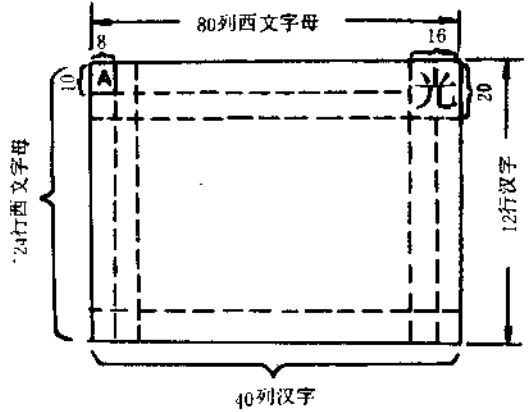


图 1-4 中西文信息屏幕字形显示举例

2. 汉字字形点阵表示法的设计原理

基本原理与西文字形点阵表示法相同，只是 7×8 的小点阵无法满足汉字字形的要求。显示汉字用的字形，一般以 14×16 点阵，即四个 7×8 点阵为最小极限。

仍以每行 640 个象素的屏幕为例，若将其分割成一组 16×20 个象素的点阵，则每个点阵可显示一个 16×18 点阵的汉字(其余部分作为光标显示位置和字母间的间隔)。整个屏幕可显示 12 行，每行 40 个汉字(即 $240 / 20 = 12$ 行， $640 / 16 = 40$ 列)。因此，对于一般的显示屏幕，满帧可显示 400~500 个汉字。

(二) 点阵式汉字库

尽管每个汉字都是由横、竖、点、撇、捺、折、挑、勾等笔画构成，没有统一的规律；尽管汉字可以从结构上分类，还有偏旁部首之分，但拼法不唯一，部件大小、左右、上下很难统一。以字形的角度看，规律性的东西甚少，所以说汉字的字形具有复杂性。但是汉字字形也有许多优点，最突出的是：“汉字是方块字”，即每个汉字无论笔画多少，每个汉字的大小都相同，均可放在一个固定大小的方格之中。因此，汉字字形的特点可归纳为“字形复杂，大小相同。”每个汉字的字形在计算机中可以看作一个同样大小的方块“图形”。

3. 汉字字形的数字化

(1) 汉字点阵字形：就是把汉字图形置于网状方格上，每格对应存贮器的一个二进制位，有笔画的格对应为“1”，无笔画的格对应为“0”。

由于汉字都可以写在同样大小的方块中，故可以把这个方块划分成许多小格，组成一个“点阵”，每个小方格是点阵中的一个点。例如 16×16 点阵，是把一个方块横向、纵向均分成 16 格，一共有 256 个小方格，称为 256 个点，组成了 16×16 的点阵。点阵中的每

个点都可以有“黑”(代表“1”)“白”(代表“0”)两种颜色，用这种点阵就可以描出汉字的字形，称为“汉字字形点阵。”目前我国汉字字形点阵，已形成了 16×16、24×24、32×32、64×64 等规格化字形点阵，但也有采用 16×18、20×20 等非规格化字形点阵的。

(2) 点阵的数字化：很容易用数字表示点阵，称为点阵的数字化如图 1-5，点阵中每一个点用一位二进制表示。若采用 16 进制，则四个点为一位十六进制数。而一个字节是指八位二进制数，故可用两位十六进制数表示，所以一个字节能代表点阵中的 8 个点。于是 16×16 的汉字点阵要用 32 个字节的信息表示。

汉字字形数字化以后，就可以把字形转化为一串数字，这串数字被称为“汉字字形的数字化信息”，简称为“字形信息”。

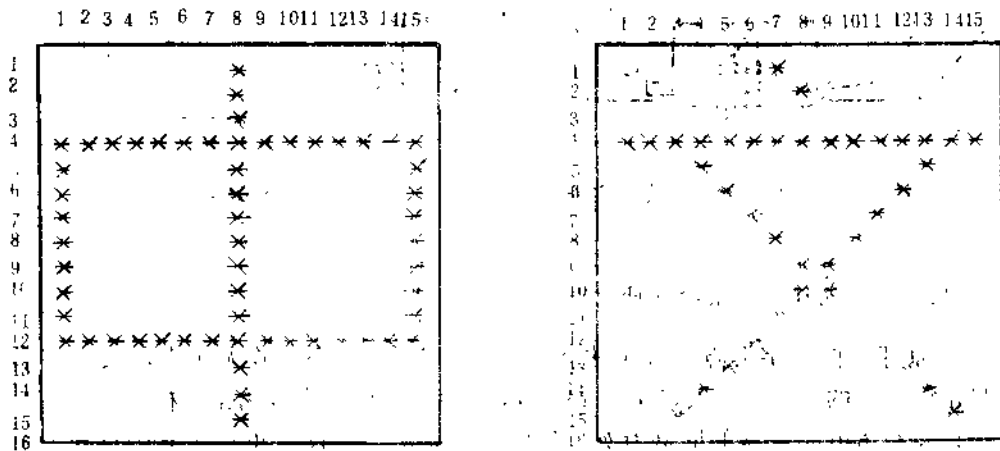


图 1-5 点阵的数字化

字库中存储的是这种点阵的代码，每个点对应的一个存储位为“1”或“0”。实际上，字库中所存的点阵信息是逻辑上相对应的，也就是说，把文字点阵图形表示为计算机字时，是将点阵划分为若干字节。

图 1-5 示出的“中”、“文”两字用 15×16 点阵表示，占用 32 个字节的存储单元。若用二进制数和十六进制数表示“中”、“文”两个字，则为：

中		文	
二进制数	十六进制数	二进制数	十六进制数
0000000010000000	0080	0000000100000000	0100
0000000010000000	0080	0000000010000000	0080
0000000010000000	0080	0000000000000000	0000
0111111111111111	7FFF	0111111111111111	7FFF
0100000010000001	4081	0001000000000100	1004
0100000010000001	4081	0000100000001000	0808
0100000010000001	4081	0000010000010000	0410
0100000010000001	4081	0000001000100000	0220