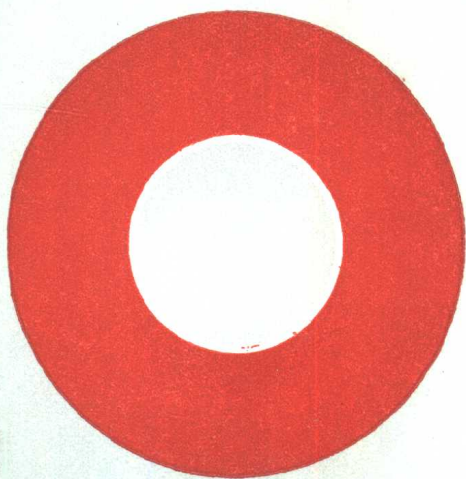
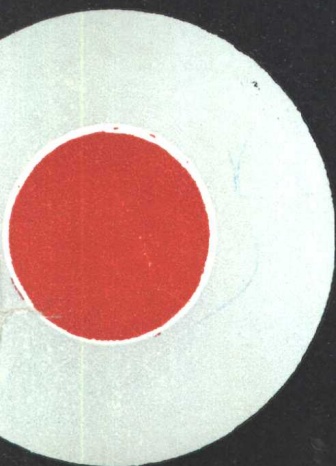




用磁帶 交換 數據的 技術

許健民

• • • • • 高 等 出 版 社 •

用磁带交换数据的技术

许健民 编著

高等教育出版社

内 容 简 介

本书介绍用磁带记录、存储和交换资料的有关技术问题。全书共分七章。第一至第三章讨论数据在磁带上的存放格式，第四、五章介绍选择物理记录和逻辑记录类型和长度时需要考虑的因素，第六章叙述用FORTRAN语言实现磁带读写的有关问题，第七章给出建立磁带资料库需要考虑的基本问题。

本书可供利用计算机从事数据处理工作的广大科技人员参考。

用磁带交换数据的技术

许健民 编著

责任编辑 庞金波

• • •

系 统 出 版 社 出 版

(北京西郊白石桥路46号)

北京昌平环球印刷厂印刷

新华书店北京发行所发行 全国各地新华书店经售

• • •

开本：787×1092 1/32 印张：2.5 字数：52千字

1986年10月第一版 1986年10月第一次印刷

印数：1—4000

统一书号：13194·0348 定价：0.62元

序 言

气象部门不是一个直接创造物质财富的部门，其主要任务是向社会提供天气、气候等信息和咨询。这种信息通过各部门的活动而产生经济效益和社会效益。因此，从某种意义上来说，气象部门是个信息部门，它的大量工作是围绕着获取、传输、交换、加工、储存、分发信息数据而进行的。气象部门的信息数据有两个重要特点：一是信息量大，随着观测手段的改进和增多，信息量越来越大；二是需要进行大量地交换。各地气象台站所获取的气象信息不仅是自己用，也不仅与邻近地区和省份交换，而且要进行全国性和国际性的交换。

当前记载气象信息最主要、最普遍的载体是磁带。这是由于它非常适合于计算机的使用，也便于携带、储存、交换。为此，世界气象组织已经制定了国际气象部门通用的标准磁带格式。这犹如有了国际气象电码一样，使大家有共同的“语言”。使用了这一标准格式，各计算机都能识别和读出磁带上的数据内容。近几年来，我们国家也把大量储存的资料转到了磁带这一载体上。这对今后气象资料的加工、使用都是方便的。虽然，数据和信息的载体将来会进一步的发展、改变。例如复制成高密度的光盘。但从目前情况看，这种改变不会很快。据我所知，至少世界气象组织现在还未考虑制定光盘的国际标准。将来即使有变化，从磁带转入新的

载体也是方便的。因此，利用和掌握磁带交换数据的技术，是当前气象业务、科研单位必不可少的一项基础知识。

许健民同志写了这本“用磁带交换数据”的技术小册子，对如何利用磁带记录数据的知识、技术、标准等作了较为全面和通俗易懂的介绍。我相信，这本书的出版对气象工作者正确地使用磁带载体来进行气象信息数据的交换是有益的。它还可以作为气象部门广大台站和科研单位的一本很好的学习材料。

骆继宾

1986年2月于北京

前 言

数字电子计算机的出现，有力地推动了其它科学技术的发展。它的应用大致可以概括为三类：科学计算、数据处理和自动控制。其中数据处理是指对各种数据和信息的加工、合并、分类、检索等工作。由于每一个业务单位都需要保存和处理大量数据，因而数据处理的工作量在计算机应用的三个方面中占比重最大。据统计，用作数据处理的计算机在美国约占百分之七十，在罗马尼亚占百分之九十以上。北京气象中心有四台大型计算机从事资料处理工作，每天要接收、分析、储存和转发大约八十兆比特的气象资料。如何保存、分发，并有效地利用这些资料已成为日常业务中的基本课题。随着我国四个现代化的逐步发展，像北京气象中心这样需要做大量资料处理工作的业务单位将会越来越多，数字电子计算机的应用必将愈加深入广泛。

我国目前气象资料服务的主要方式是资料的整编、出版。用这种方式分发资料，用户得到的信息不能直接输入计算机进行再处理。随着电子计算机的逐渐普及，越来越多的资料将需要并可能在信息化媒体上进行交换。

用磁带交换资料有存储密度大，输入输出快、携带方便以及对环境条件要求不高等优点。因此磁带已成为国际上交换气象资料的标准媒体。

计算机型号非常多，资料在机器内存中和输出到外部媒体上时所用的格式目前尚未统一。所以要读出一盘从别的型

号的计算机写入的，资料格式细节不明的磁带往往不是一件很容易的事。为了推广在信息化媒体上交换资料的技术，我们编写了这个小册子。介绍用磁带媒体记录、存储和交换资料的基本知识。

在气象行业中，计算机专业的科技人员数量不多，应用程序一般由气象科技人员自己编写，这种情况在我国目前是很普遍的。作者在实际工作中注意到，有些用计算机算题的科技人员对机器了解甚少，碰到错误不能分析，也不善于自觉地避免发生错误。针对这种情况，这本小册子不仅写应该怎样用磁带进行输入输出，而且通过对机器输入输出处理的分析，指出为什么要这样做。这本小册子的读者对象就是在工作需要利用计算机进行数据处理的广大科技人员。希望这里所提供的关于使用计算机的基本知识有助于初学者了解计算机，并把它用好。

本书第一至第三章讨论数据在磁带上存放的格式，第四、第五章介绍选择物理记录和逻辑记录类型和长度时需考虑的因素，第六章叙述用FORTRAN语言实现磁带读写的有关问题，第七章给出建立磁带资料库需要考虑的基本问题。

在这项工作进行的过程中，张德祥、赵振纪、陈建军等同志曾与作者进行过多次有益的讨论，应显勋同志审阅了全部手稿，王世昌、黎连业同志阅读了部分章节。对于这些同志的支持和帮助，作者一并表示感谢。

由于作者水平有限，错误之处在所难免，请读者批评指正。

目 录

序言

前言

第一章	数据格式	1
§1	信息和数据	1
§2	比特、字节和字	1
§3	计算机内部的字符代码	3
§4	计算机内部的十进制数代码	7
§5	定点二进制数格式	8
§6	浮点二进制数格式	9
§7	计算机外部的字符代码	11
§8	通信传输编码	11
§9	小结	13
第二章	磁带和磁带机	14
§1	用磁带做信息媒体的优点	14
§2	磁带的物理特性	15
§3	差错检验	16
§4	磁带机	17
§5	记录方式	18
§6	磁带记录格式	19
§7	小结	21
第三章	磁带文件	22
§1	文件、记录和记录项	22
§2	文卷和记录块	22

§3	磁带上的物理记录	24
§4	磁带上的逻辑记录 (1)	24
§5	磁带上逻辑记录 (2) ——跨块记录	27
§6	记录型式对输入输出效率的影响	29
§7	读写操作和读写处理的配合对输入输出效率的影响	31
§8	磁带标号	31
§9	小结	36
第四章	物理记录长度选择	37
§1	记录组块的必要性	37
§2	限制记录块长度的因素	42
§3	多个输入输出缓冲区的应用	45
§4	小结	48
第五章	逻辑记录格式选择	48
§1	字符代码格式	48
§2	二进制比特流记带 (1) ——全字长数原码	50
§3	二进制比特流记带 (2) ——浮点数压缩码	50
§4	二进制比特流记带 (3) ——定点数压缩码	52
§5	二进制比特流记带 (4) ——半字长定点数原码	53
§6	检查和检验	54
§7	记录索引	55
§8	非定长逻辑记录的组块	55
§9	字符格式和二进制比特流格式读写所占CPU时间和磁带空间的比较	56
§10	小结	57
第六章	用FORTRAN IV实现磁带读写	58

§1	读写语句的执行过程	58
§2	作业控制语言及其与程序语言的关系	59
§3	FORTRAN 读写语句	60
§4	作业控制语句中定义记录类型和长度的参数	63
§5	小结	67
第七章	建立磁带资料库需要考虑的几个基本问题	67
§1	资料格式的选择	67
§2	资料内容	68
§3	质量控制	69
§4	带库管理	69
	参考文献	70

第一章 数据格式

这一章讨论数据在计算机内存中以及在机器之间编码传递时的格式。

§1 信息和数据

计算机中的信息基本上可以分成指令和数据两大类。指令是指经过适当编码并作为一个单元引入计算机的信息。每种指令规定计算机执行一种操作或运算。数据则是指用字母、符号和数字表示的情报。数据又可以分成数值的和非数值的两类。数值数据（数）在计算机中的存放格式有区域十进制、装配十进制、定点二进制、浮点二进制四种。图 1.1 表示计算机中基本信息类型之间的关系。

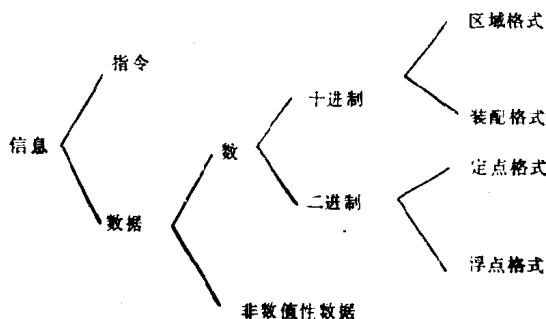


图1.1 计算机中的基本信息类型

§2 比特、字节和字

数字电子计算机都采用二进位计数制计数。因为用一个

物理元件来表示一个二进制数时，只要求它具有两种不同的稳定状态分别表示 0 和 1 就可以了。显然，制造一个具有两种不同稳定状态的物理元件比制造一个具有多种不同稳定状态的元件容易得多。比特 (Bit) 就是二进位计数制的信息单元。它是英文二进制数 (BI-nary digiT) 一词的缩写。

对一般用户来说由计算机加工处理的最小信息单位为字节 (Byte)。字节也是计算机中主存储器的最小地址单元。

字节的长度基本上是根据使用要求和硬件的代价考虑决定的。早期的计算机字节长一般是 6 个二进制位。因为 6 位是将 26 个英文字母和 10 个二进制数字进行编码所需的最小位数。它还可以允许扩充 28 个专用字符来表示标点符号、数字符号等。另外，在计算机的指令中，一般用一个字节表示操作码，指出应该执行的操作。早期的计算机操作码数量较少，用 6 个二进制位来区分就足够了。

随着计算机技术的发展，6 位字节所能代表的字符数和操作码数不够用了。于是 IBM 公司在 60 年代中期设计 360 系列计算机时，决定把字节长度改成 8 位。8 位字节比 6 位字节有以下优点。首先，8 位字节可以有效地存放两个十进制数而无需浪费空间，而 6 位字节只能存放一个十进制数，浪费了两位；其次，现在计算机中操作码的种类已增加到 200 个左右，6 位字节容纳不下了；再则，8 位字节不仅便于用字符表示信息，也便于用二进制数直接进行输入输出。由于有这些优点，8 位字节目前已被大部分计算机厂商采用。

字 (Word) 是一组连续的字节，用来存放一个数。字的长度 (字长, Word length) 是字节长度的整数倍。6 位字节计算机的字长可以是 12、18、36、48 或 60 比特。8 位字节计算机的字长可以是 16、32 或 64 比特。对于字长为 32

位的 IBM-360/370 或 M 系列机而言，16 位字长的字称作半字，64 位称双字。在科学计算中，字是表示数的基本单位。

§3 计算机内部的字符代码

计算机使用的字符 (Character) 共有三类：

数字 0 1 2 …… 9

字母 A a B b …… Z z

专用字符 · , = + 空格 ; , ……

早期的 6 位字节计算机内存中的字符代码不统一。8 位字节计算机就比较一致了。它们一般采用 EBCDIC 码或 ASCII 码作为字符代码。

EBCDIC 码是 IBM 公司提出来的，目前使用最广泛。它是英文扩充的二进制编码的十进制交换码 (Expanded Binary Coded Decimal Interchange Code) 一词的缩写。EBCDIC 码所有 8 位都用作信息，一个字节可以表示 256 种不同的字符。每个字符代码可以用两个十六进制数表示。左边一个 (高位) 十六进制数称为字符代码的信息区，简称区；右边 (低位) 的十六进制数称为数字部分，简称数字。当用 EBCDIC 码表示数字时，信息区内固定放 1111，即十六进制数 F (15)。从 0 到 9 的各个数编码为 F 0 到 F 9，其中 0 到 9 为数字部分。

有的计算机用 ASCII 码加工信息。ASCII 一词是英文美国信息交换标准代码 (American Standard Code for Information Interchange) 一词的缩写。它原来是一种用于通信线路传送信息的七位代码。因为机器字节比原先的 ASCII 码还多一位，故让信息区内权重为 2 的位 (左边第 3 位) 等于权重为 8 的位 (最左边一位)。这样就把 ASCII 码

扩展到8位,记作ASCII-8。当用ASCII-8码表示数字时,信息区内放0101,即十六进制数5。从0到9各个数编码为50到59。

表1.1和表1.2中分别列举了十六进制代码表和常用的EBCDIC、ASCII-8和ASCII代码表。完整的字符代码表可从文献〔6〕的附录中找到。在表1.2中,当我们说字符A的EBCDIC码为C1时,这意味着字母A的EBCDIC码为11000001。依此类推,字母A的ASCII-8和ASCII码分别为10100001和1000001。读者可以从表1.2中注意到,ASCII码是7位码,它与ASCII-8的关系就是ASCII-8在ASCII码的第二列与第三列之间加了一列,这一列恰好等于ASCII码的第一列。

表 1.1 十六进制代码

数 字	二 进 制 表 示 法
0	0 0 0 0
1	0 0 0 1
2	0 0 1 0
3	0 0 1 1
4	0 1 0 0
5	0 1 0 1
6	0 1 1 0
7	0 1 1 1
8	1 0 0 0
9	1 0 0 1
A	1 0 1 0
B	1 0 1 1
C	1 1 0 0
D	1 1 0 1
E	1 1 1 0
F	1 1 1 1

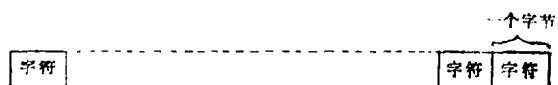
表 1.2 常用的 BCD、EBCDIC、ASCII-8
和 ASCII 代码

代 码 种 类 字 符	BCD		EBCDIC		ASCII-8		ASCII	
	6		8		8		7	
	2	4	4	4	4	4	3	4
A	3	1	C	1	A	1	4	1
B	3	2	C	2	A	2	4	2
C	3	3	C	3	A	3	4	3
D	3	4	C	4	A	4	4	4
E	3	5	C	5	A	5	4	5
F	3	6	C	6	A	6	4	6
G	3	7	C	7	A	7	4	7
H	3	8	C	8	A	8	4	8
I	3	9	C	9	A	9	4	9
J	2	1	D	1	A	A	4	A
K	2	2	D	2	A	B	4	B
L	2	3	D	3	A	C	4	C
M	2	4	D	4	A	D	4	D
N	2	5	D	5	A	E	4	E
O	2	6	D	6	A	F	4	F
P	2	7	D	7	B	0	5	0
Q	2	8	D	8	B	1	5	1
R	2	9	D	9	B	2	5	2
S	1	2	E	2	B	3	5	3
T	1	3	E	3	B	4	5	4
U	1	4	E	4	B	5	5	5
V	1	5	E	5	B	6	5	6
W	1	6	E	6	B	7	5	7
X	1	7	E	7	B	8	5	8
Y	1	8	E	8	B	9	5	9
Z	1	9	E	9	B	A	5	A
0	0	A	F	0	5	0	3	0

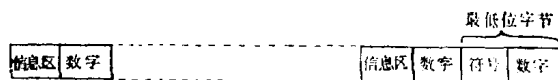
续表

代 码 种 类	BCD		EBCDIC		ASCII-8		ASCII	
位	6		8		8		7	
字 符	2	4	4	4	4	4	3	4
1	0 1		F 1		5 1		3 1	
2	0 2		F 2		5 2		3 2	
3	0 3		F 3		5 3		3 3	
4	0 4		F 4		5 4		3 4	
5	0 5		F 5		5 5		3 5	
6	0 6		F 6		5 6		3 6	
7	0 7		F 7		5 7		3 7	
8	0 8		F 8		5 8		3 8	
9	0 9		F 9		5 9		3 9	
/	1 1		6 1		4 F		2 F	
&			5 0		4 6		2 6	
空格	0 0		4 0		4 0		2 0	
+	3 0		4 E		4 B		2 B	
-	2 0		6 0		4 D		2 D	
\$	2 B		5 B		4 4		2 4	
.	3 B		4 B		4 E		2 E	
:	0 D		7 A		5 A		3 A	
,	1 B		6 B		4 C		2 C	
;	2 E		5 E		5 B		3 B	
#	3 D		7 B		4 3		2 3	
*	2 C		5 C		4 A		2 A	
(	1 C		4 D		4 8		2 8	
)	3 C		5 D		4 9		2 9	
%	2 D		6 C		4 5		2 5	
@			7 C		A 0		4 0	
!			7 D		4 7		2 7	
<	3 E		4 C		5 C		3 C	
=	0 B		7 E		5 D		3 D	
>	0 E		6 E		5 E		3 E	

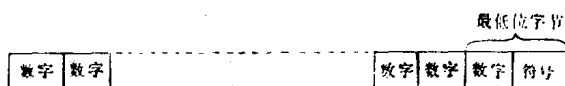
字母——数字数据可以直接用字符代码装入机器内存中，其格式如图1.2a所示。



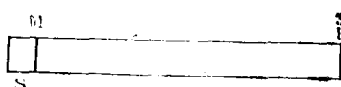
(a) 字母数字数据



(b) 有信息区的十进制数



(c) 装配式十进制数



(d) 定点十进制数



(e) 浮点十进制数

注：S表示符号位

图1.2 IBM-360/370系列机数据格式

§4 计算机内部的十进制数代码

日常生活中一般用十进制运算，送入计算机的数必须要从十进制变成二进制表示。计算机输出时也要从二进变成十进。在商业应用中，二进与十进的变换量在计算机的基本运算中占极大的比例，因此需要有适合于进行快速数制变换的数据格式。这就是十进代码。