

数据处理 概论

魏晴宇 王 阳 主编
张道芳

中国统计出版社

TP274
30

数据处理概论

魏晴宇 张道芳 王 阳 主编

中国统计出版社
一九九四年四月

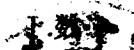
(京) 新登字 041 号

图书在版编目 (CIP) 数据

数据处理概论 / 魏曙宇等主编

-北京: 中国统计出版社, 1994.4

ISBN 7-5037-1517-0



Ⅳ. 电子计算机—数据库管理系统—教材

(IV.TP311.14)

中国统计出版社出版

(北京三里河月坛南街 38 号 100826)

河北容城印刷厂印刷

*

850×1168 毫米 32 开本 12 印张 33.4 万字

1994 年 7 月第 1 版 1994 年 7 月容城第 1 次印刷

印数: 1-5 500 册

*

ISBN 750371517-0 / C.870

定价: 9.90 元

前 言

数据处理是计算机应用的一个重要领域。利用计算机数据处理技术进行信息的采集、加工和分析是现代化信息管理发展的必然趋势。

本书以数据处理为对象，详尽地阐述了计算机应用的基础知识及其操作技术。在注重理论的完整性和科学性的同时，突出了使用和操作计算机的方法和技能，从而为广大的信息管理人员提供了一本理论与实践并举的教科书。

本书共包括四篇十三章。其中，第一篇（第1~2章）简单地介绍了计算机的若干基础知识，涉及到计算机的基本组成和工作原理、计算机的软件系统，特别是DOS操作系统；第二篇（第3章）是有关文字信息处理的技术，以WordStar为例，形象地讲述了文字处理系统运行的基本步骤及其各个功能模块的使用；第三篇（第4~9章）是有关计算机程序设计的技术，通过BASIC这一具体的语言讲述了计算机程序设计的基本算法，包括顺序、分支和循环等。第四篇（第10~13章）则全面说明了利用FoxBASE通过人机会话的方式进行结构型数据管理的方法。从而为进一步建立特定的管理信息系统打下扎实的专业基础。

本书在编印过程中，我们得到了国家统计局统计干部培训中心石占前等同志的大力支持，在此表示衷心的感谢。

因成书时间紧迫及编者水平所限，错漏之处在所难免，敬请广大读者批评指正。

编 者

1994年7月

目 录

第一篇 基础知识

第一章 计算机系统简介	(1)
1.1 数据和信息	(1)
1.2 电子计算机的特点及其发展历史	(8)
1.3 计算机的基本结构和工作原理	(16)
1.4 数据在计算机内的表示和存储	(26)
1.5 指令系统的简介	(36)
1.6 软件系统	(43)
1.7 计算机程序语言	(49)
1.8 操作系统介绍	(56)
习题	(61)
第二章 微型机的使用	(62)
2.1 对微型机的补充说明	(62)
2.2 磁盘操作系统 DOS	(68)
2.3 文件和环境	(73)
2.4 DOS 的常用命令	(77)
2.5 文件的组织和管理	(82)
2.6 CCDOS 和汉字的输入	(87)
习题	(89)

第二篇 文字信息处理

第三章 文字处理软件 WordStar	(90)
3.1 概述	(90)
3.2 WS 的启动和退出	(92)
3.3 简单编辑	(95)

3.4	高级编辑命令.....	(103)
3.5	打印及相关控制命令.....	(115)
3.6	不同文件间的复制.....	(126)
	习题	(127)

第三篇 数值型数据处理 (算法和基本 BASIC 介绍)

第四章	程序设计的基本思想	(128)
4.1	算法的概念.....	(128)
4.2	三种基本的控制结构.....	(133)
4.3	算法的流程描述.....	(139)
4.4	关于算法和程序设计思想的讨论.....	(143)
	习题	(146)
第五章	BASIC 的一些基本概念	(147)
5.1	BASIC 语言简介	(147)
5.2	常量和变量.....	(150)
5.3	标准函数.....	(153)
5.4	表达式.....	(156)
	习题	(158)
第六章	简单程序设计	(159)
6.1	赋值与数据的输入.....	(159)
6.2	数据的输出.....	(167)
6.3	几个辅助性的语句.....	(171)
6.4	简单程序设计举例.....	(173)
	习题	(176)
第七章	分支	(179)
7.1	转向语句(GOTO 语句).....	(179)
7.2	条件的表示方法.....	(180)

7.3 条件语句(IF 语句).....	(183)
7.4 程序设计.....	(189)
习题	(197)
第八章 循环	(200)
8.1 步长型循环(FOR-NEXT 语句)	(200)
8.2 条件型循环(WHILE-WEND 循环)	(206)
8.3 多重循环.....	(209)
8.4 程序举例.....	(213)
习题	(220)
第九章 函数和子程序	(222)
9.1 函数.....	(222)
9.2 子程序.....	(225)
9.3 ON-GOSUB 语句	(234)
习题	(236)

第四篇 结构型数据处理

(汉字-FoxBASE⁺ V2.10 基础)

导言 关于结构型数据	(238)
第十章 数据库系统的概念	(241)
10.1 数据库系统的基本概念	(241)
10.2 汉字 FoxBASE ⁺ V2.10 的特点和性能	(244)
10.3 汉字 FoxBASE ⁺ 的运行环境和一些约定	(247)
习题	(248)
第十一章 数据库的建立和修改	(249)
11.1 建立和修改数据库文件的结构	(249)
11.2 输入数据	(268)
11.3 列数据库清单	(270)
11.4 库文件的修改	(274)

习题	(283)
第十二章 数据的简单处理和加工	(285)
12.1 对库文件中数据的查询和检索	(285)
12.2 数据的累加和统计	(290)
12.3 打印报表	(303)
12.4 函数	(311)
习题	(322)
第十三章 对数据进一步处理及库文件之间的联结	(324)
13.1 复制数据库文件	(324)
13.2 排序	(330)
13.3 同类数据合计	(331)
13.4 索引文件	(334)
13.5 数据库的联结	(342)
13.6 更名和删除库文件	(350)
习题	(352)

附 录

附录一 WordStar 命令一览表	(353)
附录二 怎样启动和使用 BASIC	(357)
附录三 BASIC 的保留字清单	(361)
附录四 常用的 BASIC 标准函数	(362)
附录五 FoxBASE ⁺ V2.10 全屏幕编辑键及 控制键的功能	(364)
附录六 FoxBASE ⁺ V2.10 命令集	(366)

第一篇 基础知识

第一章 计算机系统简介

本章讨论数据和信息的概念，简单介绍电子计算机的基本知识(包括软、硬件系统)，及其基本的工作原理。

1·1 数据和信息

1·1·1 数据和信息的概念

近年来，信息这个词已广泛用于生活的各个领域，例如经济信息、商品信息、市场信息、科技信息…这样的词，我们经常都可以听到，但到底什么是信息呢？在人们的理解上并不是一致的。它通常是作为新闻、消息、情报、数据、知识等等的一种比较含糊的统称。虽然人们对信息的利用很早就已开始，而且我们每天都在用语言、文字、信件、图象甚至手势、表情等等来传播和交换信息，利用所得到的信息。但是迄今为止对它的本质仍然不十分清楚。人们从不同的角度对信息给予了不同的解释和定义。例如：

“信息，消息。”(辞源)

“信息，就是谈论的事情，新闻和知识。”(牛津字典)

“信息就是观察或研究过程中获得的数据，新闻和知识。”(韦氏字典)

“信息是所观察事物的知识。”(日语广辞苑)

“信息就是我们在适应外部世界，并且使这种适应反作用于外部世界的过程中，同外部世界交换的内容的名称。”(美国·维纳)

“信息概念即使不作为一种哲学范畴，也可以使反映，认识形象这样一些哲学范畴具体化，不仅可以揭示这些哲学范畴的质的特点，也可以揭示其量的特点。”(原苏联·A·别尔格)

“信息是原材料，知识是思维对信息加工的产物。”(法国·L·布里渊)

.....

可以看到，人们在认识上是不统一的，但尽管没有公认的、统一的定义，上面的种种说法却可以帮助我们不同的侧面来理解这个词，了解它的某些特征。根据数据处理中所要考虑的情况和问题，我们首先给数据下一个定义，然后再讨论信息这个词的含意。

数据也可以从不同的角度给予定义。这里我们将数据定义为：对于客体属性的记录。根据这个定义，数据这个概念包括了两方面的含意：

第一、它是客体属性的反映。这是数据的内容，当数据作为客体属性的反映时，它是属性名和属性值的统一体。例如，有一张人员登记表，如表 1·1。表中的每个人都是客体，姓名、性别、年龄等等是属性名，而李平、张小明、男、女、32、财务科等等则是属性的值。一个属性名可以联系着多个属性值，但一个属性值则必须联系着一个属性名。数据离开了属性名就失去了其所反映的客体属性，因而使属性值失去了意义。例如，单独的一个数据“25”或“财务科”，什么也不能说明。反之，数据没有属性值则不能反映客体的具体特性，因而也就失去了其作为数据的内容。从表 1·1 还可以看到，数据具有定量和定性的分别，例如，年龄、入厂时间是反映客体定量属性的，而性别、职务等则是反映客体定性属性的。

表 1·1

姓名	性别	年龄	入厂时间	所属单位	职务	...
李 平	男	32	1973	财务科	科长	
张小明	女	26	1986	财务科	科员	
王光华	男	29	1980	物资科	科员	
.....						
.. ..						

第二、数据具有载体和表示形式。数据作为客体属性的记录，必须有一定的物理载体，它可以记在纸上，以纸作为载体。但也可以采用别的载体，例如磁盘、录音带、录像带等等。

数据在表示形式上也是多种多样的，最常用的是语言符号，例如当我们用符号来表示 25 这个数时，我们可以用“25”来表示，也可以用“贰拾五”来表示，还可以用“twenty-five”来表示等等。作为数据表示形式的符号又称为数据的符号载体。

数据的表示形式与内容有一定的联系，通常我们习惯于将定量的属性值用数字符号来表示，如 32 等。对于定性的属性值则采用文字符号，如女、北京等，但这也不是绝对的，在一定的条件下，定性的属性也可以用数字来表示它的值。例如，我们可以约定用数字 0 表示男，用数字 1 表示女。在我们的实际生活中，对于地区使用邮政编码，对于产品名称、型号等有时也用数字表示。同样，定量的属性有时也可以采用文字符号来描述，例如我们用老、中、青来描述年龄。

什么是信息呢？我们将信息理解为：数据经过加工后得到的，对于某个目的说来有用的知识。

为了能更好地理解，我们举几个生活中的例子。

医生在看病时要询问病人的情况，了解病人的生活习性和病史、测体温、量血压、听诊、化验、…，这一系列的工作都可以认为这是在采集数据。在采集了这些数据以后，经过医生的考虑，在头脑中对这些数据进行处理，最后得出结论：病人大概是患了什么病，是身体的某个部份出了问题。就是说，医生已经将他采集来的数据加工成了对于病人说来有用的知识，即信息。

一位县长对于本县某个农民的一个耕地今年种了什么农作物，产量如何，这样的数据可能没有多少兴趣，因为这对他的工作说来没有什么用。可是这种数据经过各级统计人员的采集和整理，分类汇总等等加工，变成了统计报表。报表中提供了这样的信息：本县今年种了多少亩水稻，多少亩小麦，多少亩蔬菜，各自的产量多少，和去年相比情况如何等等。这些数据对一位县长说来，成了必须掌握和了解的

东西，因为这对他的工作有用。所以对这位县长说来，这些加工后得到的数据就是信息。

当会计做帐时，必须有各种发票和单据，这些发票和单据对于会计说来，就是原始数据。会计将它们按照一定的规定和处理顺序加工，做成了为各种不同需要服务的帐目和报表，用以向各有关部门提供各类信息。

所以简单说来，数据处理工作也就是将数据加工转换为信息的工作，如图 1.1 所示。

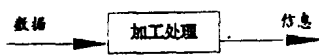


图 1.1 数据和信息的关系

但是在理解数据和信息这两个概念时应该注意，前面的说法不能绝对化。当我们把数据定义为客体属性的记录时，什么是“客体属性”？我们没有说明。在谈到信息的概念时，什么是“某个目的”？我们也没有说明。因此，对于某个指定的目的说来，有些知识是经过加工后得到的信息，但对于另外一个目的说来，这些知识又可能成为“客体属性的记录”，即数据。所以数据和信息这两个概念可以视为一对范畴，它们相互对立、相互联系、并且在一定的条件下相互转化。

在前面所举的例中，我们将医生对于病人病情的诊断视为对数据的采集和加工过程，将医生诊断的结果视为对病人有用的知识(信息)。但是当卫生部门进行医学统计，分析某地区的疾病情况，各种疾病的发病率、治愈率等等时，对于每个病人的诊断结果又被视为“客体属性的记录”，再从中加工出新的信息。

同样，一个县的农业统计年报，对于县的农业管理部门说来，是经过统计等等加工后得到的重要信息，但是当进行全省的农业统计汇总时，每个县的农业统计年报又被视为“客体属性的记录”，需要再从中加工出对全省农业发展有用的信息来。即使就一个县说来，如果我

们的目的是对它的农业发展进行长期规划或预测，那么过去若干年的农业统计年报也只能作为一批基本数据。

因此在数据处理工作中，常常会出现图 1·2 所示的情形。为了某个目的，我们采集了有关的数据，这些数据经过方法 1 的加工处理，得到相应的信息。这些信息对于另外一个目的说来，又是数据的一个组成部份，它们和其它的数据一块，再经过方法 2 的加工处理，得到新的信息。而这些信息对于另外一个目的说来，又可能成为一部份基本数据，即需要加工的原材料等等。

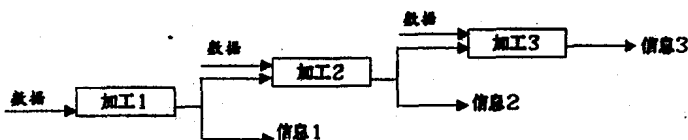


图 1.2 数据的多次加工

由于这个原因，在讨论和说明问题时，如果没有明确目的是什么，也不一定要将数据和信息这两个概念分得那么清楚（实际上，不明确目的时是说不清楚的）。我们常常将数据处理工作又称为信息处理工作，这不会造成什么误解和困难，但在实际进行一项数据处理工作时，分清这两者却是必要的。数据是被加工的对象，是本次加工处理的输入内容，而信息则是加工后得到的结果，是本次处理的输出内容。

1·1·2 数据处理工作的意义

上面的说明只是对于数据处理工作的一种比较抽象的概括，实际的数据处理工作当然并不这样简单。我们在前面的例子中提到了医生、统计人员和会计人员，这些工作谁也不会认为是简单的事情。

在数据处理工作中，常常会涉及许多问题，而且不是轻易可以解决的。

概括地讲，一项数据处理任务包括下列的几个步骤：

(1)采集.由于客体的属性是多种多样的，从不同的角度出发，为

了不同的目的，可以提出不同的要求。因而在数据的采集工作中，一个重要的问题是：将客体的哪些属性作为有用的数据来采集。如果将范围定得太大，看来似乎很全面实际上却增加了数据采集工作的困难。例如有的数据要付出很大的代价才可能采集到，有的甚至根本无法采集到；反之，如果我们将数据采集的范围定得太小就可能无法加工出某些有用的信息。实际上，这个问题在相当大的程度上决定着整个数据处理工作的质量，应该给予足够的重视。

(2)校验.在数据处理工作的每个环节上都存在着校验的问题，但这里所说的则是作为单独一个环节的校验工作。为了保证数据的质量，采集到的数据必须用某种方法进行认真的校验，不能允许错误的数据大量地混杂在其中，在计算机的数据处理工作中有一句名言：“如果输入的是垃圾，那么输出的也是垃圾”。这就是说，当我们输入的是一些含有重大错误的数据时，对它们进行加工处理的结果绝不可能得到高质量的信息。数据的校验工作表面看来似乎不太困难，实际上也不容易。例如，对成千上万的，甚至几十万数据进行数据的校验，就是一件相当复杂的事情，没有一定的科学方法是很难作好的。

(3)传输.传输是数据处理工作中不可缺少的一环，因为数据处理工作的各个环节并不一定是在同一个地点进行。数据的采集工作可能是分散在各个不同的地方，而对数据的加工则是在某个确定的地方，数据的使用又可能是在另一个地方。所以数据需要经过传输，送到指定的地方去。数据的传输工作的效果如何也将影响到信息的质量。这里所说的效果包括两方面的意思：一方面是准确性，即保证在传输过程中不致出错，使采集到的数据受到损失。这一点容易理解，例如文件和书信传递中的丢失问题，无线电通讯中的误码问题等等。尽管我们可以保证在传输中基本上不出错，但要完全消灭错误却是一件非常困难的事情。数据传输中效果的另一方面是指实时性，即要保证数据能及时传输到。不能及时提供准确的数据也就不可能及时得到所需的信息。我们可以设想一下，如果一个厂长知道的只是一个月前市场的销售和原材料供应情况，他有没有可能组织和安排好生产？同样，一

个商店的经理只知道上个月的销售和货源情况，他能作好生意吗？

(4)加工.即将数据加工转换为有用的信息。这当然是关键的一环，因为数据并不能自动地转换为信息。前面我们用医生看病的例子来说明数据和信息的关系，可以想到，同样是这一批数据，如果给一个没有医学知识的人时，很可能什么也加工不出来。与此类似，一批股票市场情况的数据，对于不懂经济的人说来，是看不出所以然的。应该注意的是，对于数据的加工不仅需要付出能量和时间，而且还要具有相当的知识，没有这些是作不到或作不好的，加工工作的水平决定着对同一批数据所能取得的信息的水平。我们还应该看到，数据加工这个词的含义是非常广的，通常我们将对数值型数据的各种计算，对非数值型数据的拼接、取子串、转换等等都视为对数据的加工。实际远远不止如此，例如，文章的翻译、文件的起草编辑、数字和图形的转换、对问题的分析、预测、推理等等，都属于数据加工的范围。

(5)存储.由于数据的采集和传输都需要时间，这就使得数据处理工作表现为一个持续的过程，在加工工作中不仅要用到当前的数据，而且也要用到过去一段时间的数据，用到从别的途径得来的数据。加工后得到的信息也需要保存。所以采用一定的方法，用一定的物理介质来保存原来的数据和加工后得到的结果是必要的。至于数据存储中应该注意的事项，安全问题和数据丢失的危害性等等，这是都比较容易想到和容易理解的。

(6)输出.加工后得到的信息应该及时提供给使用人员，否则可能失去意义。

数据处理工作并不是现在才有的，在人类的历史上早已出现，只不过由于历史条件的限制，人们还不能从更广泛、更一般的意义上来认识它，不将它视为数据处理工作而已。例如，会计、统计、新闻、档案等等工作中，就有相当大一部份内容是属于数据处理性质的，只是因为加工的数据对象和加工方法不同，这才分门别类，各自根据自身的特点形成各自的学科。随着人类社会的发展，人们对于信息的需求在不断发展，在数量上、质量上和時間上都提出了更高的要求。特

别是近几十年来，由于生产的迅速发展，生产规模的不断扩大，生产效率的提高，使所需要的信息量急剧增加。要求的条件也越来越高。形成了所谓“信息爆炸”的局面。现在人们在生产和社会活动中所需要信息，已经很难单纯依靠个人的自然智力来应付。举例说，一个有上千人工厂的厂长，他不可能全面而详细地掌握全厂的生产情况、财务收支情况、产品销售情况、原材料供应情况、市场变化情况、工人思想情况…，他只能采取集体领导、分工负责的方法，依靠全厂各级管理干部和工人。但即使如此，仍有许多问题不能解决，其中的一个重要问题就是：在我们相当大一部份企业中，现在基本上仍是采用手工式的数据处理工方式，这种不适应现代化、高效率生产的管理方法所必然带来的一部份后果就是：单据多、票证多、报表多、数据的不统一和某些错误、数据加工不及时、工作中急需的信息无法及时得到。数据处理工作中这种被动的局面造成了大量有形、无形的损失，它迫使人们去寻找改变这种局面的方法和技术手段，以进一步促进生产的发展。这就提出了数据处理工作如何自动化的问题。因而也就出现了数据处理这个学科，以及在各个领域中利用计算机进行数据（信息）处理的要求。我们现在所讨论的数据处理，也就指如何利用计算机技术来完成数据处理工作中的各种要求。这是我们学习数据处理这门课程的基本目的。

1.2 电子计算机的特点及其发展历史

1.2.1 电子计算机的特点

如果确切一点，也许我们将电子计算机称为“信息处理机”更恰当一些，因为它是一种能够按照人们预先规定好的指令，去自动、快速地完成数据处理工作（包括科技方面的数值计算）的设备。它的出现是二十世纪科学技术上的一项重大成果。对于计算机的掌握和使用，是今天任何一个人都应该作到的。

在总体上，我们可以将电子计算机分为两大类，一类是数字计算

机，另一类是模拟计算机。数字计算机是直接对各种符号进行加工处理，例如，对文件进行编辑修改，对数进行加、减、乘、除等各种演算等等。它可以说是算盘这一类计算工具的发展。模拟计算机不是直接对数字进行运算，而是用一些物理量，例如电压、长度等来表示被处理的数字，它可以说是计算尺这一类计算工具的发展。对于某些科学和工程上的问题说来，模拟计算机使用起来更方便、更简单。但是模拟计算机的计算精度不可能很高，一般只能达到千分之一左右的精度。所以对于数据处理和计算精度要求更高的问题说来，模拟计算机是不合适的。数字计算机解题的精度要高得多，并且具有很大的通用性，所以在数据处理工作中使用的都是数字计算机。以后我们谈到计算机时，也都是指的数字计算机这一类。

数字计算机具有以下的一些特点：

(1)数值计算和逻辑运算的功能。数值计算是指对数据进行加、减、乘、除等运算。逻辑运算是指对数据进行比较、判断、选择等。利用计算机的逻辑运算功能，我们可以对数据进行逻辑检查、对比、分类分组、并按照规定要求进行挑选等等。这就给数据的加工处理工作带来了很大方便，可以将数据的计算以及整理数据、分类汇总、检索查找、报表加工等工作都交给计算机去处理。如果进一步考虑，我们还可以想到，数值型数据是用一定的符号（信号）来表示，非数值型数据也是用某些符号（信号）来表示，只是加工的方法不同而已。所以求档资料的起草、修改和编辑等等都可以在计算机上进行。甚至图纸的设计和绘制、作曲和演奏等等也可以用计算机来作辅助工具。

(2)“记忆”的功能。计算机可以将大量的基本数据、处理过程中得到的中间数据，以及处理的最终结果，按照人们的命令保存起来，以便在任何需要的时候再提取出来使用。这个功能很类似于大脑的记忆功能，在计算机科学的术语中，它称为对数据的存储和提取。现在的计算机可以存储几百万到上亿的数据，这就给将大量的资料、报表等都存储到计算机内提供了可能。而且从计算机内提取数据所用的时间是以秒甚至微秒（等于一百万分之一秒）来计算，这是任何人工保存