

档案管理与计算机

辅 导 提 要

孙 淑 扬 编 著

一 九 八 七 年 十 月

档案管理与计算机

中国人民大学档案学院

武汉市档案学会

目 录

第一章	档案管理与计算机.....	(1)
第二章	电子计算机概论.....	(3)
第三章	电子计算机基本结构.....	(11)
第四章	电子计算机软件基础.....	(27)
第五章	电子计算机程序设计.....	(43)
第六章	档案的前处理工作.....	(55)
第七章	档案数据的存贮技术.....	(63)
第八章	档案数据库.....	(82)
第九章	机读档案的检索.....	(94)
第十章	汉字处理系统.....	(112)
第十一章	电子计算机档案管理系统的研制.....	(126)

第一章 档案管理 with 计算机

本章是绪论，主要内容叙述了电子计算机在档案工作中的应用及其对档案工作的促进。同时概述了本课程的性质特点和学习目的要求。

〔本章学习要点〕

一、电子计算机在档案工作中的主要应用

电子计算机在档案工作中的应用可以分为两个方面，一是对档案机读目录的处理，其二是档案日常管理工作的自动化。

计算机对档案机读目录的处理，首先对档案文献进行著录标引，得到二次档案信息，将二次档案信息输入计算机，形成档案的机读目录，计算机通过对机读目录进行运算处理，完成以下管理工作：

1. 自动编目 编目工作是档案管理部门开展各项业务工作的基础和必要条件，同时，又是一项很繁琐的工作。计算机可以按照用户要求，实现多种方法编目。

2. 统计工作 档案的统计工作是通过对其数量的观察研究，以指标数字的形式，揭示档案及档案工作的某些规律性，计算机以其极强的运算能力，实现各种统计活动，包括对档案内容本身的各种统计和档案利用情况的统计。

3. 检索工作 电子计算机提供多途径的检索方法对机

读目录进行检索，其速度和质量都远远超过手工检索。

计算机用于档案馆室日常管理工作，主要有：档案的借阅管理，归还管理，缩微制品的检索，库房管理，温湿度自动调节，防火报警自动控制等。

二、本课程性质 本课程是档案管理科学与计算机科学相互交叉渗透而形成的边缘科学，就其所学内容来看，偏重于自然科学范畴。

三、本课程的学习方法 由于《档案管理与计算机》是一门新课，现在正处在发展阶段，一些技术问题和理论问题还有待于进一步的研究、发展、完备，又因为这门课程涉及计算机科学的很多领域，所以学习起来有一定的难度，为此，在学习中提醒同学们注意以下几点：

1. 要打破计算机神秘论的思想，学习这门课的目的之一，就是要掌握计算机的基本原理和使用方法，打破其神秘论。这一点也应当是学习这门课程首先要树立的思想。

2. 学习中重点掌握三基，即基本理论、基本概念、基本方法。及时复习，稳扎稳打，循序渐进。在教科书中，有时为了帮助学生理解一个概念，往往从多方面、较深入地阐述，内容偏多，而录音授课是基本教材的提炼与深化，我们只要求学生掌握讲授的内容。

3. 理论与实际相结合，配合课上讲授内容，课下参观一些计算机管理系统和计算机机房是十分必要的，有条件的学生应当作到这一环节，从而巩固所学知识，使理论和实际相结合。

第二章 电子计算机概论

本章概括叙述电子计算机分类、发展、特点及应用的一般情况，并着重讲述与据数处理有关的基本知识。

〔学习要点〕

一、电子计算机的发展演化

1946年第一台电子计算机问世，它的名字叫做埃尼阿克（ENIAC）。埃尼阿克是个庞然大物，整个机器包括一万八千个电子管，七千个电阻，一万个电容和六千个继电器。机器占地面积一百七十平方米，重量大约三十多吨。它的运算速度每秒钟五千次，这个运算速度是普通人计算能力的十万倍，在当时是很了不起的进步。

自从埃尼阿克问世以来，至今已有四十年的历史，在这四十年中电子计算机迅速发展，已经经历了四代的演变。对于电子计算机发展时代的划分标准，在学术界尚不大统一，比较通用的标准是按照它所采用的电子元器件分代。以电子元器件分代的同时兼顾计算机的体系结构、使用方式、软件等方面的条件，将计算机的发展划分为以下四个时期：

第一代计算机，大约从1947年至1957年，这一代计算机所使用的元器件是电子管（真空管）。

第二代计算机，大约从1958年至1964年，这一

代计算机所使用的电子元器件是晶体管。晶体管与真空管相比，寿命和可靠性都有所提高，能量消耗、体积和成本都有所降低，因此很快取代了第一代计算机。

第三代计算机，大约从1965年至1970年，这一代计算机的重要标志是采用了集成电路。同时在软件方面有了操作系统，出现了终端和网络。这一代计算机的性能/价格比较有了很大提高。

第四代计算机，大约从1970年至八十年代，其重要标志是采用了大规模集成电路。大规模集成电路的应用促进了计算机向巨型化和微型化的发展。运算速度超过亿次的巨型计算机的出现和体积小、使用灵活的微型机的出现，为计算机的广泛应用创造了更好的条件。

日前，日本等国家正在进行第五代电子计算机的研制，第五代计算机应当是超大规模集成电路的计算机，同时其软件要能够理解自然语言，具有逻辑推理演绎功能，即智能型的计算机，近年来第五代电子计算机的研究工作有了较大的进展。

二、电子计算机的特点

电子计算机之所以发展飞速，是因为它具有很多突出的特点：

1. 运算速度快 一般计算机运算速度为每秒钟几千次或几万次，巨型计算机的运算速度可以达到每秒钟亿次以上，这个运算速度是人和其他任何运算工具所不能比拟的。

2. 运算精度高。

3. 具有“记忆”信息的能力 电子计算机的存贮器具

有很大的存贮容量，在档案管理工作，计算机可以存贮成千上万张著录卡片，以至于存贮大、中型综合性档案馆全部馆藏文件的目录。

4. 具有逻辑判断能力 计算机不仅能进行算术运算，而且能进行逻辑运算，具有逻辑判断能力。利用这个功能，计算机才能胜任事务管理工作。

5. 具有自动运行工作的特点 电子计算机能够按照人们事先编好的程序，自动地进行运算、处理工作，并且自动地将处理结果打印出来。由于电子计算机的工作是自动进行，所以有人把计算机档案管理系统称为档案自动管理系统 (Archives Automation System)。

三、电子计算机的主要应用

电子计算机在人们的生产和生活中得到广泛的应用，归纳起来有四个主要方面：

1. 用于科学计算

2. 用于数据处理和信息加工 如财务管理，物资管理，企业管理，档案、情报、图书的管理以及办公自动化、信息管理系统等。

3. 用于自动控制 主要指生产过程、管理过程的自动控制，自动调节。

4. 用于计算机辅助设计 (CAD) 计算机用来辅助工程设计、机械设计以至于轻纺、服装设计等工作，可以提高设计质量和效率，收到显著的经济效果。

此外，电子计算机用于人工智能的研究工作，近年来也有很大进展。

四、数据的概念

数据是基本事实的记录。在计算机科学中，凡是能够输入到计算机中，并且被计算机进行处理的对象统称为数据。因此数据的概念是很广泛的。

数值型数据（数值化了的数据）是指数字。如：1，2，3，…8，9，0，1952，8848等。

字符型数据，指字母和符号组成的数据，如：字母A，B，C，…X，Y，Z；各种符号+，-，×，/，!，%，&，（）；字符串IBM—PC（计算机型号），I090301（文件分类号），02.23.57（档号），Beiing（北京）等。

各种表达式，如： $(3 + 2) \times 10$ （算术表达式），A $\wedge$ C（逻辑表达式），C>D（关系表达式）等。

文字型数据，如China（英文：中国），中央档案馆（汉字）。以及世界各国文字和我国少数民族的文字。

图象型数据。计算机处理的各类图形、图象、图表以致于照片、影象，输入计算机，都成为图象型数据。

声音型数据，包括能够输入计算机进行处理、存贮并输出的各种声音，如音乐、讲话等。

在计算机档案管理系统中，经过著录标引的二次档案信息输入计算机构成档案的机读目录，机读目录包括档号、题名、责任者、日期、主题词等项，由足够多的条目组成，称之为档案数据。

五、数据处理

数据处理是对数据进行收集、存贮、加工、传递等一系

列活动的总和。数据处理的目的就是从大量的杂乱无章的信息中提取出对人们有用的信息（数据）来。

自古以来，伴随着人们的社会活动和生产活动就存在着许多数据处理的问题，需要随时将彼此有关联的数据进行统一处理。在计算机诞生之前是手工数据处理阶段，处理数据信息的效率低、质量差。电子计算机诞生大约十年之后，其应用领域由科技计算扩展到数据处理。电子计算机数据处理的主要优点是对数据进行自动处理，处理速度快、质量高。

计算机对档案数据的处理包括对档案数据的收集、存储、整理、分类、编目、统计、检索输出等一系列的工作。

数据处理的特点：

1. 数据处理加工的对象大多数是非数值型的数据，如字符型或文字型数据，数据量大，数据结构复杂。以档案数据为例，一个计算机档案管理系统所管理的数据往往是拥有上千条、上万条档案条目的机读目录。由于要处理巨量的汉字，所以档案管理系统的输入、输出、存储数据量都是很大的。

2. 运算量少而简单。数据处理的数据量很大，而运算量少，而且比较简单，大多是分类、归总、比较、判断、选择等处理。

3. 数据处理用于管理工作，要求报表输出功能强，时间性强。

六、数据处理系统

在计算机数据处理时，人和机器构成一个完成数据处理功能的整体，这个整体称为数据处理系统。

常见的数据处理系统有以下类型：

1. 批处理系统

批处理系统的作业方式是批处理，即，将一批用户的要求，分别写在提问单上，集中起来组成一组，按照系统规定的格式一起送入计算机，机器自动地一个作业一个作业地进行处理，直到一批作业全部处理完毕，再自动地将处理结果打印出来，由系统工作人员整理后分发给各个用户。早期的图书、情报检索多采用批处理。往往是租用某个计算机，每周一次进行处理，在此之前由工作人员用电话将用户检索要求收集集中起来，通过一次上机运行，检索各用户要求，然后打印报告，分发给各用户。采用批处理方式，计算机吞吐量较大，使用效率较高，成本较低。其缺点是用户不能及时得到处理结果；处理时用户不与计算机系统接触，如果机器处理过程中发现用户提问有错误，不能够中途修改，只能待一批作业全部处理完了，再行修改，修改之后随下一批作业进行处理，这样会造成人力和机时的浪费。

批处理适用于以下一些情况的检索：（1）应答时间要求不严格；（2）处理过程中不需要人干预；（3）输入输出数据量很大。

2. 实时处理系统

实时处理方式的主要特点是计算机对于外来的处理要求，以足够快的速度进行处理，并在一定时间内作出响应，立即将处理结果交给用户。在这种处理方式下，用户的提问不积压，处理及时，得到结果也及时，可以克服批处理的一些缺点。

3. 联机集中处理系统

在联机处理方式中，用户坐在计算机终端前面直接与计算机对话，终端与主机相连，人和机器以问答方式，直接提出检索要求，直接得到答案。联机处理是比批处理更进一步的检索方式，在这种处理方式下，从准备数据、编制程序，直到运行程序、最后得到检索结果，用户紧密地与计算机相联系，可以随时观察、掌握计算机的工作过程，发现问题，及时修改，从而使批处理的一些弊病得到根本改善。

4. 分布式处理系统

在分布式系统中，若干台结构上互相独立的计算机分布在远离主机的不同地区，它们在逻辑上、物理上互相连接，机器之间可以相互通信。数据在什么地方产生，就在什么地方处理，计算机之间数据、硬件、软件资源即是分布的，又是共享的。这种系统可靠性高，灵活性大，经济性、扩充性都较好。

〔名 词 解 释〕

一、电子数字计算机 电子计算机的一种，它是以数字形式的量值在机器内部进行存贮和运算的电子计算机。是目前应用最广泛的电子计算机。

二、集成电路

1. 中小规模集成电路 通过半导体集成技术，把许多逻辑电路集中做到一小块硅片上的基本电子元件称为集成电路，其集成度在100个门电路以下者，称为中小规模集成电路。

2. 大规模集成电路 集成度超过100个门电路或者

10000个以上晶体管元件的集成电路。

3. 超大规模集成电路 集成度超过一万个门电路或者十万个以上晶体管元件的半导体集成电路。

三、数据 基本事实的记录，在计算机中凡是能够输入计算机，并被计算机所处理的对象，统称为数据。

四、数据处理 对数据进行收集、存贮、加工、传递等一系列活动的总和。其目的是从大量的杂乱无章的数据中提取出对人们有用的信息来。

五、批处理 将计算机的用户要求，集中起来成批进行处理的作业方式。是早期文献信息检索系统采用的主要处理方式。

六、实时处理 参见“学习要点”六、2。

七、联机处理 在数据处理工作中，终端与中央处理机相连，用户直接将数据输入到中央处理机进行处理，处理结果又可以直接传给用户。这种工作方式称为联机处理。

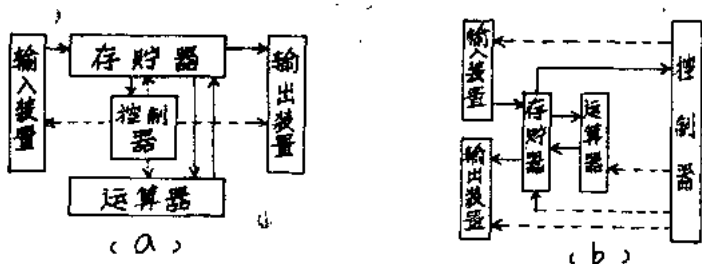
第三章 电子计算机基本结构

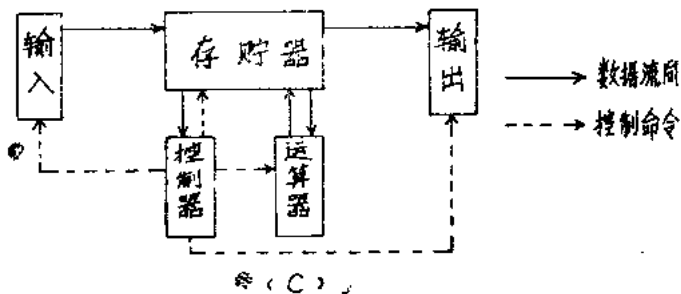
本章讲述了电子计算机的基本组成，各组成部分的主要功能及各部分之间的信息传送。并分别介绍各个部分的基本原理和常用设备。

〔学习要点〕

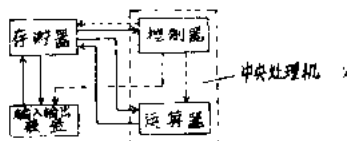
一、电子计算机的基本结构示意图

通过计算机结构图，可以示意地表示出计算机的基本组成，以及各部分之间的数据、信息传递关系，因此结构图是学习本章的基础。对于计算机结构图不仅会读、会画，并且能够识别各种变幻形式。下面列举几种不同形式的结构图。图一中三个结构图是由五个方框组成：控制器、存储器、运算器、输入装置和输出装置。图二结构图将输入装置和输出装置合为一个方框来表示。





图一 电子计算机结构图 (一)

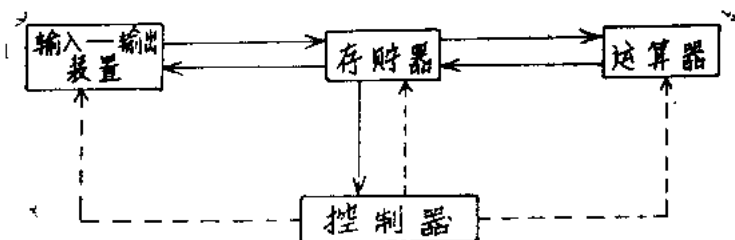


(见后页)

(a)

控制命令——→ (b)
数据流向——→

图二 电子计算机结构图 (二)



图二 (b)

计算机结构图所表示的模式，是自从电子计算机问世以来，绝大多数计算机的基本组成模式，是由大数学家冯·诺

依曼提出的，故而称为冯·诺依曼结构。这个结构图说明以下问题：

1. 电子计算机由五个基本部分组成，即控制器、运算器、存贮器、输入装置和输出装置。

2. 计算机的工作过程：当计算机进行运算或者数据处理时，首先由输入装置将数据、程序、用户提问要求输入计算机，输入的数据存放在存贮器中，当需要进行运算时，数据由存贮器传递到运算器，同时，数据信息(包括程序和提问要求)要从存贮器传递给控制器，以便控制器指挥各部分工作。运算器运算处理后的中间结果和最终结果送回存贮器，按照控制器的指令，在适当的时候由存贮器将运算结果送到输出装置，输出给操作员。图一、图二中的数据流向和控制命令指向表明了计算机中数据和命令信息的传递过程。

在一些计算机上，设有控制台，控制台上装有各种开关、按钮和指示灯，操作员(或用户)通过控制台，可以启动计算机和观察机器各部分的工作是否正常。

二、电子计算机各部份的功能

1. 控制器 控制器相当于计算机的大脑或神经中枢，它的作用是控制整个计算机自动地、协调一致地进行工作。其功能的具体表现如下：

(1) 在适当的时机，自动将程序、原始数据输入到内存贮器和外存贮器中。

(2) 控制内、外存贮器与各部件交换信息。

(3) 控制运算器和内存贮器及有关部分自动而协调地执行程序中规定的运算过程。

(4) 控制运算结果从计算机内部送至输出设备执行输出, 以打印或者图象等形式表示出来。

控制器在向各执行部件发出命令之前, 首先要从存贮器中取出计算机操作人员事先编好的程序, 然后按照程序向各执行部件发出控制命令。这些控制命令通称为“指令”。计算机的指令很多, 主要有算术运算指令、逻辑运算指令、程序控制指令、输入输出指令、数据传送指令及特殊指令等类型。一台计算机的全部指令构成这台计算机的指令系统。控制器通过指令系统来完成其控制功能。

2. 运算器 运算器的功能是对数据进行算术运算和逻辑运算。算术运算是指按照算术运算规则进行的加、减、乘、除、乘方、开方、求绝对值等运算。逻辑运算指数据的比较、选择、分类、移位、转移、逻辑与、逻辑或、逻辑非等非算术性的运算, 其运算对象主要是非数值型的数据。在计算机解题时, 首先将要解决的问题化为一系列的算术运算或者逻辑运算, 运算器不断地得到由存贮器提供的数据, 进行快速运算, 并且把运算的中间结果和最终结果送回存贮器保存起来。运算器的整个工作过程是在控制器的统一指挥下, 按照程序所编排的先后次序自动地进行。

通常把运算器和控制器合起来称为中央处理单元(中央处理机), 简称CPU, 表示这两个部件是计算机的中心部件, 其性能好坏对整个计算机的性能起着决定性的影响。

3. 存贮器 存贮器是计算机的记忆部件, 其功能是用来存放原始数据、处理这些数据的程序、运算的中间结果和最终结果。

4. 输入装置 输入装置的作用是将外部世界的数