

档案自动化基础

耿立大 邱晓威 孙 钢 编著

档案出版社



档案自动化基础

耿立大 邱晓威 孙 钢 编著

档案出版社

一九八七年

档案自动化基础

狄立大 邱晓威 孙 钢 编著

※

档案出版社出版

(北京市西城区丰盛胡同21号)

新华书店北京发行所发行

国防科工委印刷厂印装

※

开本: 787×1092毫米 1/32 印张: 11.5 字数: 256千字

1987年9月第一版 1987年9月第一次印刷

印数: 1—10000册

统一书号: 7283·100 定价: 2.75元

ISBN7—80019—043—9/G·30

前 言

本书是向档案工作者介绍档案自动化方面基础知识的一本入门读物。由于档案自动化涉及到计算机技术、信息处理技术和档案管理技术等广泛的领域，本书将选择一些比较重要的问题和基本知识作深入浅出的介绍。我们认为，在初次接触档案自动化课题时，应首先对档案自动化的主要方面有较全面的了解，这将有助于确定深入研究和工作的方向。因此，本书内容侧重知识性和普及性。同时，为了适应档案自动化工作的实际需要，本书还介绍了一些实现档案自动化的具体方法。愿本书的出版能对档案工作者学习和了解档案自动化有所帮助。

本书作者分别来自计算机界和档案界，尽管我们做了很多的协调和努力，但对档案自动化这个新颖而复杂的问题，仍会有考虑不周和叙述不准的地方。希望广大读者给予批评指正。我们相信，随着档案自动化工作的发展，档案自动化理论必将得到进一步的补充和完善。

编著者

1986年9月

目 录

第一章 概述	(1)
第一节 档案自动化的概念.....	(1)
第二节 档案工作与档案自动化.....	(3)
第三节 档案自动化的应用.....	(9)
第四节 实现档案自动化的条件.....	(15)
第五节 档案自动化的理论研究.....	(20)
第二章 与档案自动化关系密切的计算机应用	(26)
第一节 计算机和信息化社会.....	(27)
第二节 计算机应用概况.....	(28)
第三节 社会科学情报资料工作自动化.....	(32)
第四节 图书馆自动化.....	(37)
第五节 办公室自动化.....	(42)
第六节 管理信息系统.....	(49)
第七节 语言文字处理和排版印刷自动化.....	(53)
第三章 电子计算机简介	(57)
第一节 计算机的发展和特点.....	(57)
第二节 计算机硬件系统和档案自动化的 要求.....	(65)
第三节 计算机存贮器.....	(69)
第四节 计算机输入输出设备.....	(76)

第五节	计算机网络·····	(82)
第六节	汉字信息处理系统·····	(90)
第七节	微型计算机·····	(92)
第四章	计算机软件知识·····	(95)
第一节	计算机系统软件·····	(96)
第二节	计算机解题过程和程序设计·····	(102)
第三节	计算机语言概述·····	(108)
第四节	FORTRAN 语言·····	(110)
第五节	BASIC 语言·····	(113)
第六节	COBOL 语言·····	(119)
第七节	微型机软件·····	(123)
第五章	档案目录的计算机存贮方式·····	(134)
第一节	档案目录存入计算机的形式·····	(134)
第二节	数据结构简介·····	(142)
第三节	档案目录的文档管理方式·····	(155)
第四节	《文献目录信息交换用磁带格式》 简介·····	(159)
第五节	档案目录的数据库管理方式·····	(165)
第六章	档案自动检索系统的基本算法·····	(170)
第一节	算法设计基本知识·····	(170)
第二节	常用数据处理算法介绍·····	(173)
第三节	自动检索算法介绍·····	(201)
第七章	档案自动检索系统的设计与改进·····	(246)
第一节	系统工程简介·····	(246)
第二节	档案自动检索系统的实现步骤·····	(249)
第三节	档案自动检索系统的设计·····	(262)

第四节	档案自动检索系统的改进·····	(280)
第五节	档案自动检索系统的类型和发展 阶段·····	(283)
第八章	档案自动检索系统的数据准备·····	(290)
第一节	档案检索系统与档案检索语言·····	(290)
第二节	档案著录·····	(297)
第三节	档案分类法和档案分类标引·····	(304)
第四节	档案主题词表和档案主题标引·····	(313)
第五节	档案著录与标引的标准化与自动化·····	(328)
第九章	国外档案自动化概况和机读档案·····	(341)
第一节	国外档案自动化概况·····	(341)
第二节	机读档案·····	(353)
参考文献	·····	(360)

第一章 概 述

第一节 档案自动化的概念

一、档案自动化的含义

档案自动化 (Archival Automation) 是档案管理工作自动化的简称, 其内容可以广义理解为“计算机在档案工作中的应用”。目前档案自动化一词已在世界上广为流行。国际档案理事会设有“档案自动化委员会”, 定期举行档案自动化方面的国际会议。过去曾有过“档案的电子数据处理”、

“档案自动信息处理”等提法, 但是目前使用最多的称呼则是“档案自动化”。我们参照国际档案理事会的叫法, 用“档案自动化”一词来表示使用电子计算机对档案进行自动化管理。这里的“自动”和“自动化”都不是指无人或完全自动的意思, 而是指由计算机辅助人进行档案管理工作。由于档案工作的各个环节中, 档案的编目和检索部分最易于实现自动化, 因此“档案自动化”又主要是指档案编目和检索自动化。为在本书范围内不致发生概念上的混乱, 本书也使用档案自动化的提法。

二、档案自动化与信息处理技术的关系

档案自动化的技术基础是电子计算机技术。目前人们已

经认识到电子计算机就是信息处理机。档案自动化是与信息处理自动化密切相关的。想要实现档案自动化，就应当对信息和信息处理技术进行研究。

信息的形式是多种多样的，我们将档案、图书、情报资料等通称为文献信息。档案馆、图书馆、情报资料中心可以看作信息库或信息源。对这部分信息的重要性，人们是非常清楚的。有人统计过，一个人如果读完大学，就意味着生命的 $\frac{1}{4}$ 已用于同书本打交道。一个研究人员查阅资料的时间往往占全部工作时间的 $\frac{1}{3}$ 。60 年代美国和日本进行的统计表明，一些化学研究人员业务工作的时间分配如下：查阅文献资料占 50.9%，研究实验占 32.1%，写报告占 9.3%，计划和思考占 7.7%。可以想见，一个历史学家用于查阅史料的时间比例会更大。这些事实和统计告诉我们，文献信息对于社会的发展是不可缺少的。档案作为人们社会活动的历史记录，还具有“凭证”的重要作用，无论现在和将来，档案都是具有研究、借鉴和考证价值的重要信息。

“信息库”或者“信息源”的提法说明了这样一个事实，即历史悠久的档案馆、图书馆和情报资料中心的工作方式和工作重点，正从注重文献收集、保管的向心工作方式转到一种注重信息发送的离心方式上来。档案自动化为这种转变提供了技术上的支持。

档案的主要作用是它可以作为历史的说明和凭证。“历史说明”作用可以依赖于只保留内容信息的计算机文字处理技术得以实现，而“历史凭证”作用则需要采用既保留内容信息又保留形态信息的计算机图形存贮和处理技术来实现。其实，无论是文字信息还是图形信息，都是以编码的形式在

计算机中进行存贮和加工的，只是后者占用较大的存贮空间。因此对于计算机来说，信息就是一些编码，这些编码的具体含义只有人才能够理解。

信息处理技术的理论基础是信息论。信息论简单地说，是利用数学的方法，研究信息的计量、传送、变换和存储的一门学问。现代通讯技术、计算机技术、自动控制等都是信息论作为基础理论。以上这些技术是信息处理技术的主要支柱。它们使信息的采集、加工、传递进入了自动化阶段。档案自动化正是在信息处理自动化的基础上发展起来的。信息处理自动化代表着新的技术和手段，而档案自动化则是这些新技术在档案管理中的具体应用。因此从本质上说，研究或实现档案自动化，就是研究和认识档案与信息处理自动化新技术的关系。为了正确地理解这个关系，可以从两个方面来分析和认识。一方面是对档案工作的现状进行分析，并对其发展趋势进行预测。在此基础上充分认识档案工作对自动化信息处理新技术的需求程度。另一个方面是，对现有的信息处理新技术进行考察、分析、论证、实验，在此基础上正确地评价这些新技术在档案工作中的作用和采用这些新技术所应具备的条件以及采用后对档案工作产生的影响。下面几节中将对上述问题作一些概括性的分析和介绍。

第二节 档案工作与档案自动化

一、档案工作对自动化新技术的需求

人们都有这样一个经验，即更愿意去大一些的档案馆、

图书馆去借阅所需的文献资料，因为这样做可以有更大的把握满足自己的需求。人们都这样想、这样做，就使得综合性强、馆藏量大的馆中保管的档案资料和书刊被利用的机会增加，效益增大。于是该馆的地位也就越加重要。这就是所谓的“价值体积效应”。在这个“价值体积效应”规律的支配下，档案馆、图书馆、情报资料中心都积极地从事收集和采购工作，以不断地丰富自己的馆藏。这样做的结果，使得各个馆的馆藏量迅速膨胀。但是从提供利用来说，馆藏量越大则查找起来越困难。因此，必须注意到馆藏量的增加给整理、保管和提供利用带来的一系列问题。所以馆藏的迅速膨胀问题如果处理不好也会带来灾难性的后果。解决这个问题的途径在哪里呢？除了增加人员编制和扩大库房以外，就是及时采用先进的管理技术和方法。这种情况就反映了档案工作对自动化技术的需求。从发展趋势来说，当前我国档案事业正处在一个大的发展转变之中。这种转变主要表现为：由以管理党政档案、纸质档案为主，转为全面管理各种门类和各种形式的档案；由以传统的手工管理为主进到以应用电子计算机等现代技术为主的多层次技术结构的管理技术和方法；由以机关、团体、企业、事业的档案室工作为主进到以各级档案馆工作为主；由以一般地接待查阅档案为主进到以广泛地开发传递档案信息为主。

这种发展趋势使档案工作者面临着新的问题。然而与这个令人振奋的新趋势相对比，档案部门的现状却是不容乐观的。由于十年动乱的影响以及其他原因，多数档案馆基础工作薄弱，程度不同地存在着检索方法落后、检索工具不全、文件保密与开放界限不清、借阅手续繁琐等问题。这些问题

阻碍着档案工作的顺利发展，但是如果我们积极地研究和引进新技术，可以促进这些问题的解决，扭转被动局面。

二、信息处理新技术的适用性和对档案工作的影响

与档案工作有关的信息处理技术正在迅速发展之中。究竟哪一种新技术可以采用，应当经过充分的论证和试验来确定。这是因为档案是不能再生的，它的保管期限应尽量长，一旦损毁造成的损失是无法弥补的。所以必须选用那些适应这些要求的成熟技术，不可盲目追求先进性。另外，某一种新技术总有它的局限性，甚至常常会对档案工作提出新的要求，带来很大影响，因此更加需要有准备地、逐步地、局部地采用这些新技术，以便于协调各种关系，发现和解决遇到的问题。

对档案有深刻影响的新技术主要有缩微技术和计算机技术。缩微技术使得档案保存的空间缩小数十倍以上，保存时间延长数十倍。而计算机技术则对档案信息的处理以至对档案学的理论发展产生深刻的影响。这些影响在以下几个方面将有所体现：

1. 档案载体的改变。随着办公自动化的实现，档案信息的载体从传统的纸张或胶片变为磁带、磁盘、光盘等。载体的改变，将直接影响到档案工作的各个环节。

2. 传递方式的改变。从手翻眼看和物理载体的传递到使用电子自动化手段超越时间和空间的限制来检索和加工、传递档案信息。

3. 对档案学理论发展的推动。

在自动化之前的很长时期中，由于各类文献的管理主要

依靠人脑记忆和部分手工参考工具。所以，人们非常注重各类文献的“个性”，并且创立和完善了各类文献管理学科，例如档案学、情报学、图书馆学等等。这些学科分别对各自的文献特点做了详尽而周密的研究和考察，确定了一套有效的管理方法和一套相应的理论依据。这些学科的价值是人所共知的，但是随着整个社会的信息化和各类信息的社会化，这些不同类文献的“共性”将越来越引起人们的重视。在整个自动化信息处理系统中，各类文献信息的区别和界限将变得模糊起来。一方面某类文献信息如果不符合统一的标准就难以进入自动化系统；另一方面，一类文献适合于自动化的管理方法也必将被另一类文献所采用。所以对于各类文献共性的研究不可避免地提到日程上来。

在档案管理的现代化趋势中，我们已感觉到有大量的概念、方法和理论涌到档案管理的过程中来。由于档案毕竟有自己的特点和规律，所以对于各种“外来”的概念、方法和理论必须进行消化、整理和解释才能有效地吸收和使用，使之与自身融为一体。在这个过程中，简单地照搬和保守地拒绝等态度都是不可取的。

只有档案学理论的发展，才能使这种“个性”与“共性”的矛盾得到解决。换句话说，只有注重研究并解决这对矛盾，才能使档案学理论的发展跟上现代化步伐。因此，可以说电子计算机等新技术的采用是档案学理论发展的新的推动力量之一。

三、档案自动化的简况

1946年电子计算机问世以后，开始主要是用于科学计

算。从50年代起，人们逐渐认识到计算机的作用并不限于“数值计算”。例如，用计算机可以翻译外文、查找文献、进行经济管理等。随着计算机的普及应用，60年代初，国外一些档案工作者在计算机专业人员的配合下，开始在档案工作中试用计算机。由于档案工作的特殊性以及当时计算机（特别是外部设备）技术水平的限制，这些试验应用并不总是成功的。其中有些系统由于不能完全适用于档案工作而被淘汰，有些系统则经过不断改进而演变为更加适用的新系统。同时，档案自动化的发展也带来了一系列引起争议的档案工作方法和档案学理论问题，要求档案工作者和有关专家深入研讨。

70年代，档案自动化工作在欧美各国开展较多，其中美国、加拿大、英国、法国以及苏联、东德、罗马尼亚和波兰等国都已建立起若干个档案自动化系统，有些至今还在使用。此外，在很多国家的档案教学中都设有“计算机和档案自动化”课程。档案自动化的研究和实践也很活跃，许多刊物上不断地发表有关档案自动化的论文和报导。

随着微型计算机的出现和迅速普及，进入80年代以来，计算机在档案工作中已有了更多的应用。由于计算机价格的大幅度下降，功能的不断增强，今后它将更大规模地应用于档案工作。计算机和档案自动化也将成为每个档案工作者必须了解的基础业务知识。

当然，必须说明，计算机尽管优点很多，用途很大，但仍然要由人创造一定的条件，它才能正常而有效地工作。另外，采用电子计算机需要较大的投资，所以使用计算机前首先要论证和规划。但是，总的说来，档案自动化所带来的好

处是显著的,对档案工作的改善是明显的,也就是说,为使用计算机而付出的代价将成倍地得到报偿。这已经被一些国家档案自动化的实践所证明。可以说,以采用计算机为主要技术手段的档案自动化是不可逆转的发展趋势。另外,也应当看到,档案自动化面临的问题尽管与图书馆自动化、情报资料工作自动化有某些共同点,但是档案自动化毕竟有自己的特殊问题。这些特殊问题使档案自动化工作变得更加复杂,这是我们研制档案自动化系统必须予以充分重视的。

影响到自动化的档案和档案工作的特点归纳起来有如下一些:

1. 档案的馆藏在各馆之间绝大部分是不重复的。各档案馆之间的分工,按时间(历史时期)、空间(部门、机构和地区)划分得非常明确。除了某些出版物(资料)外,可交换的部分和重复的内容极少。

2. 档案的利用方式以大批量的专题查阅为主,并且要求查阅原文,仅仅查阅目录和文摘不能满足要求。

3. 大部分档案的利用需要办理审批手续,并且由档案管理人员查找。

以上列出的第一条是档案馆与图书馆、情报资料中心的主要差别。一个档案馆依靠自己的力量建立一个具有一定规模的档案数据库必然需要较长时间。由于各馆之间保管内容不重复,互相交换业务又很少,就使得建立自动化处理档案数据库的工作不能互相支持和代替。一般一个人一年只能著录标引2万张机读目录卡片。而那些需要采用计算机的档案馆都至少有数十万件或数百万件馆藏文件(或案卷),可见建库工作要花费相当的时间和人力。由此还带来其他一些困

难。比如，著录工作的标准化需要编制主题词表，而编制对所有档案馆或各个历史时期都适用的主题词表是非常困难的。此外，还有一些其他技术问题也起着阻碍的作用。例如汉字处理技术。图书馆、情报资料中心可以先引进国外的机读目录，建立西文书刊、资料数据库开展工作，而档案馆的中文档案数据库则必须自己建立等等。总之，实现档案自动化要付出更多的努力，克服更多的困难。

第三节 档案自动化的应用

档案自动化包括档案工作中计算机应用的各种方式。目前限于电子计算机的功能和档案工作本身的特点，档案自动化仅能应用在档案工作的某些环节。下面介绍几种主要的档案自动化应用项目。

一、档案自动编目系统

档案编目工作是档案馆的基础工作之一。完整的档案目录是整个档案馆工作“有序性”的体现。这种“有序性”一方面体现了档案实体的存放结构，另一方面体现了档案主题内容的查找途径。描述档案的实体存放结构，需要编制馆藏指南、全宗介绍、案卷目录和卷内文件目录。为了提供档案主题内容的查找途径，则必须编制各种分类目录、主题目录、专题目录、作者索引、时间索引等等。

在实际工作中，很多档案馆缺乏完备的目录系统，编目工作在广度、深度、种类和标准化等方面都还存在着一定问题。造成这种情况，并不是由于不重视这项工作，而是由于

编目工作量太大，难以全面完成。

电子计算机几乎可以按任一著录项目进行编目，编目结果可以自动打印或用计算机控制，自动进行排版印刷。这样使得整个编目过程速度快、周期短、效率高、项目全，不仅节省大量的人力，而且将极大地改进档案馆的检索工具，促进档案馆的基础整理工作并有助于馆际档案目录信息的交流。

档案编目系统是档案自动化的最重要和最基本的系统之一，也是应用最广和最先实用化的档案自动化系统。

通常，编目系统和检索系统作为两个相互独立的子系统而连结成一个档案自动化系统。一般的作法是，这两个子系统可以同时设计，然后分步实现。可以先调试运行编目系统，再调试运行检索系统。

传统的档案检索工具有三种，即以档案全宗为对象的检索工具（如案卷目录、卷内文件目录、全宗介绍、全宗文件目录等），以全部馆藏为对象的检索工具（如分类目录、档案馆指南等），以专题或责任者为对象的检索工具（如专题目录卡片、专题介绍、作者目录等）。其中以全宗为对象的各种检索工具着重反映的是档案之间的历史联系。只有专题目录和分类目录在一定程度上反映了档案的内容特征。

电子计算机可以针对单份文件进行主题检索，从而打破全宗的限制，因为同一主题的档案或资料可能分散在不同的全宗中，从而给专题检索带来一定的困难。同样，电子计算机可自动编出主题目录和专题目录，并可以超越全宗的界限进行统一编目。这是计算机编目的一个优越性。为了实现自动编目，需要把反映档案主题内容的信息按照著录标引的规则填入机读目录卡片（或表格），然后输入电子计算机，而由