电子计算机档案检索

（上　　册）

邱 晓 威 编著

兵 器 部 档 案 馆 编 印

一 九 八 五 年 七 月

前　　言

档案自动化是一个较新的课题，它涉及电子计算机技术，信息系统、直至档案学理论等许多学科和专门技术。档案自动化的理论还有待于在实践基础上的研究和总结。可以说目前正处在一个实际中迫切需要解决，而理论研究上较为欠缺的时期。本讲义的编写者参与了某些档案自动化实验性系统的设计工作以及参加了与自动化有关的国家标准的制定和起草工作，对档案自动化的理论作了初步探讨，有义务向其他档案工作者提供这方面的经验和知识。

档案自动化知识涉及面广，在本书深入讨论每一种专门技术和理论是很难做到的。为适应档案工作者的实际水平，本讲义有意避开一些较难的数学知识和较专门的计算机技术问题，尽量以概念的说明代替繁琐的论证。对于不承担软件设计的档案工作者来说，讲义中的某些具体算法可以略去不读，但是对于一般的粗略流程图，则建议读通。

本讲义的编写过程中，从王永成、戴璞、孙钢、姜世华等同志处得到很大的帮助和指导，在此一并致谢。

由于时间仓促，编者水平所限，有疏漏不妥之处欢迎指正。

邱晓威

一九八四年十月

内　容　提　要

本书是档案电子检索课程的专业教材。书中对档案自动化进行了分析，概括性地讲解和论述了档案自动检索涉及的各类问题。内容共分十章，分别对有关的计算机知识、文献信息的存贮和查找技术、检索系统的设计和与档案工作有关的标准化等问题进行了介绍。考虑到档案工作者急需，本书后面附有一套自动检索算方法　图，可直接用于编制编目和检索程序。

本书可作为档案工作者学习档案自动化入门读物，亦可作为档案检索应用软件设计人员和档案标引人员的参考手册。

目　　次

上册

下　　册

第一章　概述

§1—1　信息与社会的发展

一、信息——三大资源之一

社会的经济、政治、科学、文化等活动与信息有着密切的关系。信息与能源、材料被人们统称为三大资源。

1、信息资源的性质

信息是指对接受者来说预先不知道的报道。信息又是所有容易获得或不易获得的事实和思想的总合，並可在某个时刻供人参考。

知识是人们对于事实和思想的提炼和总结，即在实践活动中获得的对客观事物的认识。知识通常根据实践的不同门类划分成许多学科。

智慧是对知识的综合，是由某些学科的知识升华而成但又超越学科界的理论。可以说信息编织了知识，而知识又升华为智慧。

从使用的角度来分，信息又可以分为二类，一类是直接信息，一类是隐含信息。直接信息是指接收者可以直接使用和吸收的信息，例如各种资料，史料文献各种数据和报导等。隐含信息则是指包含于某一事物中的信息，需要经过分析和考证才可揭示出来。例如一种新式武器，一件发生的事故等等。

现代社会正处于“信息化”的过程之中，在较先进的国家里，从事信息工作的人已达到百分之五十。而且一些经济学家指出，生产率

的提高大部分是靠新知识或新信息。由于信息与物质生产已不可分割，所以信息与物质的关系引起了人们很大的重视。並着重研究了信息这一特殊资源的性质。

能源和材料资源是有形的，而信息是无形的资源。有时信息的载体是有形的，但並不能说信息本身有形。

信息的性质可归纳如下：

(1)、信息是活的。第一，信息之所以可以发挥作用，是由于人脑对它进行观察、记忆、思考、归纳。所以信息是人脑的输入物和输出物。第二，如同生物有生有死一样，信息也有“生”和“死”。“生”是指信息的产生和其价值的存在；“死”是指其价值的消失或信息的失效。

(2)、信息是可扩充的。尽管一部分信息会随着时间流逝而失效，但大部分信息在使用过程中是不断扩充的。这种扩充性体现在现有的信息不可能包含一切事物，另一方面不管怎样广泛收集、深入考察也不可能将反映某一事物的信息完整无缺地得到。

(3)、信息是可浓缩的。信息资源为了便于应用可以被浓缩、综合和归纳。例如许多事物信息反映的规律可总结成一条定理，或者将许多数据信息之间的关系压缩成一个方程式等等。

(4)、信息具有替代能力。信息可以代替资本、工人或其它有形物质。只要考察一下工厂自动化、办公自动化的工作情况就可以认识到

信息的替代能力。

(5)、信息可以以光速传递。电子通讯和通讯技术使得信息按光速进行传递，显然信息的传递速度大大超过了物质的运输速度。

(6)、信息是扩散性的。一般来讲，普通物质的保管和封锁可以做得很严密，而对于信息的保密就困难得多。信息的扩散过程，通常也就是其发挥作用的过程。

(7)、信息是可以分享的。信息一般不能作转手交易，只能作分享交易。其它物质资源在买给别人、或送给别人后，自己就不再享有它了。而信息传递给别人后，本人则仍然享有它。

关于信息资源的特性，人们会不断有新的发现，並更加广泛而合理地使用它。

2、信息资源的作用

信息资源的影响目前已经极为广泛，並且可以预见对将来的影响更加巨大。有人已经提出了“信息社会”的各种模式。

在信息社会中，图书、情报资料、档案是信息资源的一部分，可以统称为文献信息。图书馆、情报资料馆、档案馆可以看做信息库，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，一个人如果上完大学，就意味着生命的四分之一用于与书本打交道。一个科技工作者查阅资料的时间往往占全部工作时间的三分之一。六十年代美国和日本进行的统计，一些化学研究人员业务工作时间分配

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%。一个历史学家用于查阅史料的时间比例则更大。这些事实和统计告诉我们文献信息对于社会的发展是举足轻重的。还可以举出许多例子来证明档案资料给政府决策、以及生产、生活带来的效益。档案做为人们社会活动的历史记录，无论现时和将来都是人们必须学习、研究、借鉴的重要信息。

“信息库”，或者“信息源”这样的提法说明了这样一个事实，即历史悠久的档案馆，图书馆和较晚出现的情报资料馆的工作方式和重点，正从文献收集、保管的向心过程转到一种信息发送的离心方式上来。以中央档案馆为例，党中央在起草《关于建国以来若干历史问题的决议》时查调了中央档案三万多卷（件）。近五年来中央档案馆为各方面提供了二十三万卷（件）档案，为１９６６年前提供档案的１·５倍。

在人们的社会活动中，由于不能及时获得有关信息，或不注重信息的使用而造成损失的事例是非常多的。例如，５０年代中期，美国曾经耗资２０万美元动员了大量人力来研究符号逻辑在计算机电路计算方面的应用。事过之后，才得知苏联数学家隆茨１９５０年就做了这方面的工作，他的成果早已赫然在目地存放在图书馆里了。

另一个例子是希腊工程师克利斯多非独自发现的改进粒子加速器的技术。他把结果通知美国加州大学的核实验室后，信件淹没在浩如

烟海的资料之中，杳如黄鹤。同时，各国科学家都大动干戈研究这项已经被发现的技术。几年后，档案管理人员才在尘封的卷宗中发掘出这封信和这位天才的科学家。

二、信息技术的进步

很早以前人们就以各种方式传递着各种信息。但是限于笨重的载体和落后的传递方式，信息的交流很不方便。随着科学技术的进步，信息技术也发展起来，並且在理论上形成了专门的学科——信息论。

信息论简单地说是利用数学方法，研究信息的计量、传送、变换和存储的一门学科。现代通讯技术、计算技术、自动控制，甚至遗传学都是以信息论作为基础理论。以上这些技术是信息技术的主要支柱。它们使信息的采集、加工、传递进入了自动化阶段。延续数千年在竹、帛、纸上刀刻手写的传统方式已经或正在被自动化处理所取代。而这不过是电子计算机出现后不到四十年发生的变化。信息处理的自动化已发展到了什么规模呢？全面回答这个问题非常困难。下面只以与档案直接有关的文献处理来说明。

进入到８０年代仅美国、加拿大、欧洲、日本就已有与书目有关的数据库、自然语言数据库５２８个。美国最大的电子计算机化的图书馆网俄亥俄州学院图书馆中心已拥有３０００台远程终端，一个中心数据库，含有书目记录７００万条，並以每周２１００条记录的速度增长。该系统向遍及美国５０个州的２２００个成员馆提供服务，