

WenXianXinXiZiDongHuaJianSuo

文献信息自动化检索

● 胡凤振 梁妙凤 编著

● 河南科学技术出版社



文献信息自动化检索

胡凤振 梁妙凤 编著

河南科学技术出版社

内 容 提 要

本书是学习情报、图书及档案等文献信息自动化检索系统的基础知识教材。书中详细地介绍了文献信息自动化检索的服务方式、系统设计和实例,以及文献信息的收集、加工、存贮、检索的方法与步骤,还有供实现电子计算机自动检索编程序用的工作流程图。另外,还介绍了汉字信息处理及学习本书所需要的基础知识。

本书可作为大专院校情报、图书和档案管理专业的教科书,也可供从事上述专业自动化管理的科研人员、工程技术人员和其他有关人员参考。

文献信息自动化检索

胡凤振 梁妙凤 编著

责任编辑 刘嘉

河南科学技术出版社出版

河南省新乡市印刷厂印刷

河南省新华书店发行

850×1168毫米大32开本6,875印张 159 千字

1986年8月第1版 1986年8月第1次印刷

印数: 1—4,200册

统一书号15245.65 定价1.40元

前 言

随着科学技术的迅速发展,各种文献信息的数量与日俱增。为了充分发挥情报、图书及档案等文献信息的重要作用,如何加速文献信息自动化管理的问题,已广泛地引起了人们的注意。

我国情报、图书及档案等文献信息的自动化管理,特别是自动化检索起步较晚,人才及这方面的文献资料也较少。根据目前社会发展的需要,我们编写了《文献信息自动化检索》一书,较详细地介绍了情报、图书及档案等文献信息的收集、加工、存贮、查找的方法及步骤,并详细地介绍了供实现电子计算机自动检索编程序用的工作流程图,以及自动化检索系统设计的方法、步骤和设计实例,还介绍了汉字信息的处理及学习本书所需要的基础技术知识等。

本书是在多年教学讲义的基础上整理而成的。在编写过程中特别注意到内容系统,说理深入浅出,文字浅显易懂。本书既可作为大专院校情报、图书及档案等管理专业的教科书或参考书,又可供从事情报、图书及档案工作的科研人员、工程技术人员和其他有关工作人员参考或自学。

本书在编写过程中曾得到南京大学王永成、郑州航空工业管理学院刘沁海等老师的大力支持,在此表示衷心感谢。

由于编者水平有限,错误之处衷心希望读者批评指正。

编 者

目 录

第一章 概论.....	(1)
第一节 信息的特点、种类及各种信息间的相互关系 (1)	
一、信息的概念.....	(1)
二、信息的特点.....	(2)
三、信息的种类.....	(3)
四、各种信息间的相互关系.....	(7)
第二节 研究文献信息的意义.....	(9)
第三节 文献信息自动化检索的概念和意义.....	(12)
一、概念.....	(12)
二、意义.....	(13)
第四节 文献信息自动化检索的服务方式.....	(16)
第二章 文献信息自动化检索技术基础.....	(20)
第一节 离散数学简介.....	(20)
一、特征及内容.....	(20)
二、集合论简介.....	(21)
三、图论简介.....	(28)
四、关系与映射.....	(33)
五、数理逻辑简介.....	(36)
第二节 微型计算机简介.....	(38)
一、微处理器与微计算机概论.....	(38)
二、微处理器的结构.....	(45)

三、微处理器与微计算机的应用.....	(50)
第三章 文献信息的管理及自动检索系统的设计.....	(54)
第一节 文献信息的管理.....	(54)
一、手工管理.....	(54)
二、现代化管理.....	(55)
第二节 文献信息自动化检索系统的设计.....	(56)
一、系统分析阶段.....	(56)
二、系统设计阶段.....	(59)
第三节 文献信息自动化检索系统的设计实例.....	(62)
一、系统分析.....	(62)
二、系统设计.....	(63)
第四章 文献信息的存贮.....	(69)
第一节 存贮器简介.....	(69)
一、内存贮器.....	(69)
二、外存贮器.....	(73)
第二节 文献信息的存贮.....	(78)
一、存贮方法.....	(79)
二、存贮格式的转换.....	(99)
三、压缩存贮.....	(101)
四、词表简介.....	(103)
五、文档的结构.....	(105)
六、数据库简介.....	(109)
第五章 文献记录的加工及排序技术.....	(118)
第一节 文献记录的加工.....	(118)
一、对原文献记录进行加工.....	(118)
二、编制标题关键词索引.....	(119)

第二节 排序.....	(136)
一、内排序.....	(137)
二、外排序.....	(139)
第六章 文献信息的检索系统与检索方法.....	(141)
第一节 概述.....	(141)
一、分类.....	(141)
二、软件体系.....	(142)
第二节 检索步骤.....	(146)
一、给出提问.....	(146)
二、核验提问.....	(150)
三、加工提问.....	(153)
四、查找满足用户要求的文献记录.....	(155)
第三节 检索的方法.....	(155)
一、顺排检索方法.....	(155)
二、倒排检索方法.....	(170)
第七章 联机文献检索与计算机网络.....	(181)
第一节 联机文献检索简介.....	(184)
一、联机与脱机文献检索的不同点.....	(184)
二、联机文献检索的终端使用方法.....	(187)
三、联机检索的条件.....	(189)
第二节 计算机网络简介.....	(190)
一、计算机网络的组成.....	(190)
二、终端设备.....	(194)
三、计算机网络的功能.....	(197)
第八章 汉字信息处理简介.....	(198)
第一节 汉字的特点及汉字信息处理的方向.....	(199)

一、汉字的一般特点.....	(199)
二、汉字信息处理的方向.....	(200)
第二节 汉字的输入与输出.....	(204)
一、汉字的输入.....	(204)
二、汉字的输出.....	(206)
第三节 汉字信息处理的条件.....	(209)
一、硬件系统.....	(209)
二、软件系统.....	(211)

第一章 概 论

第一节 信息的特点、种类及各种 信息间的相互关系

一、信息的概念

信息与人类社会的历史一样长久。信息是一种极其宝贵的重要资源，它与人们日常生活息息相关，犹如空气和水。信息流的畅通是物流畅通的前提，是一切经济、技术、行政活动等必不可少的重要条件。现代社会正处在“信息化”时代；在较先进的国家里，从事信息工作的人已达50%。一些经济学家指出，生产率的提高，大部分是靠收集和掌握新的信息而得到。故人们当今把信息、能源和材料称为现代科学技术的三大支柱。信息工业已经成为世界上的第一大工业。

信息是在人类历史的不断发展中逐渐被人们认识、重视的。自人类产生了语言和文字后，人们为了表达自己的思想，则用语言或文字进行交流，从而也就产生了信息。随着社会不断前进，现代科学技术的迅速发展，特别是发明了电子计算机以后，各种各样的信息都可以使用计算机进行高速、高效率的处理，这样就加快了信息的流通，促使经济效益的提高、人们生活的改善、企业及行政管理的现代化及科学技术的高速发展。

信息就是一切物质运动的形态表现和内容描述的反映。对接

收者来说，预先不知道的报道，并能起借鉴、参考的事实和思想的总和，就叫做信息。从计算机处理信息的角度看，信息就是对数据的解释，或者说，信息就是数据所包含的意义。因此信息不仅包括数字，也包括文字、语言、图象及符号等。所以，凡是文献、情报、档案及图书等，通通属于信息的范畴。

二、信息的特点

由于信息与物质生产是不可分割的，所以，信息与物质的关系引起了人们的重视，并着重研究了信息这一特殊资源的性质及特点。

能源和材料资源是有形的，而信息是无形的资源。尽管信息的载体各种各样，并且是有形的，但并不能说信息本身是有形的。

一般说来信息有如下特点：

(1) 信息不是一成不变的。信息之所以可以发挥作用，是由于人们对它进行分析、记忆、思考、归纳及运用的结果。所以信息对于人们来说既是输入物也是输出物。再者，信息如同一切生物一样有着生死规律。信息的“生”是指信息能产生价值的作用；“死”是指其价值作用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。现有的信息不可能包含一切事物，即不管怎样广泛收集、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，但经过人们的分析研究，可扩充到不同的应用领域中去。

(3) 信息是可压缩的。信息资源为了便于应用，可以被压缩、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公式等等。

(4) 信息具有替代能力。信息可以代替人力、物力、财力，有了信息就有了一切。只要我们考察分析一下国内外的一些企业和单位，就可以清楚的证明这一点：哪些企业和单位经济效益高，竞争能力强，就是由于它们收集和掌握的信息多而准确，充分发挥了信息的替代能力所致。

(5) 信息可以迅速传递。随着科学技术的迅速发展，尤其电子技术、缩微技术及卫星通讯技术的高速发展，信息得以迅速的传递，它的传递速度大大超过了物质的运输速度。

(6) 信息是可以共享的。信息一般不能作转手交易，只能作分享交易。其它物质在卖给别人，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。所以一般来讲，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。也正由于信息具有这种易于扩散的特点，才能达到共享之目的。

信息所具有的特点有利于人们收集和掌握，并能广泛地应用到各个领域，而充分发挥信息的巨大作用。

三、信息的种类

随着信息资源的发展，它已深入到人们生活的各个领域，因此信息的种类繁多，例如有商品信息、经济信息、军事信息、政治信息、科技信息、文化信息等等。信息从使用的角度又可分为两大类：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，如各种情报及各种文献资料等；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，如一种新式武器，一次事故等。本书研究的是直接信息，主要包括以下几种。

1. 文献信息

文献信息是人类在生产和生活过程中积累起来的经验和知识的记录，它主要是用文字、缩微、机读及声象等形式记载下来的信息。它包括图书信息、档案信息、期刊信息及会议信息等。

一般说来它有如下特点：

(1) 文献信息的范围广、种类多、数量迅速增加。因为文献信息是记载人类历史发展进程中各方面的资料，是人类精神财富的一部分。它汇集着世代劳动者认识世界，改造世界的结晶，即许多成功的经验和失败的教训。因此，按其性质、特点、形式和用途的不同，可分成图书、档案、期刊、报告、政府出版物、会议备忘录、专利、标准、产品样本及学术论文等等文献信息。随着科学技术的迅速发展，新的科学不断出现，作为反映新技术、新材料、新工艺等等科研成果的文献信息的数量就急剧地增加。

(2) 文献信息的形式多、文字种类多。由于文献信息的数量急剧增加，记载文献信息的形式也越来越多，如有印刷型（包括铅印、油印、石印、胶印等）、缩微型（包括缩微胶卷、缩微胶片、缩微卡片等）、机读型（如磁带、磁盘等）、声象型（如电影片、唱片、录音带、录象带、幻灯片等）。此外，出版文献信息所使用的文字种类，也越来越多，据报道，世界上现在用于出版期刊信息的文字种类就多达六、七十种。

(3) 文献信息分布分散并交叉重复。由于科学技术的迅速发展，各学科之间相互交叉，互相渗透，所以文献信息的分布也就极为分散。例如有时一篇文献信息可以利用不同的载体形式，采用不同种的文字，发表在不同的地方。

(4) 文献信息的更新及使用期越来越短。由于新技术、新产品迅速出现，不断地代替旧产品、旧技术。因此，反映到文献

信息上,就使得文献信息的使用时间在逐渐缩短。

总而言之,文献信息作为人类实践活动的记录,它包括的范围广、数量多,既分散又交叉;提供利用的周期也越来越短。

2. 图书信息

图书信息是人们实践活动的概括和总结,是向人们传递知识的文献信息。它包括农业图书、工业图书、医学图书及小说等。

一般图书信息有如下特点:

(1) 图书信息的内容成熟、系统、全面。因为图书信息是专门著者经过千思万虑地推敲、选择、鉴别而融会贯通写出来的,因此图书信息是比较系统地向人们传递某一方面的文献信息。

(2) 图书信息一般在时间上没有严格的要求。图书信息所报道的文献信息,比起期刊、论文、学术报告及特种文献信息等都较晚;越古老的图书越不好采集,也越加珍贵,也就成为采集的重要对象了。

(3) 图书信息提供用户利用一般是图书原件或影印本。因此,用户如果要想对某些问题获得较全面系统的知识,或对不熟悉的问题进行初步了解,借鉴和参考图书信息则是一个有效的办法。

总而言之,图书信息是一些专门著作的集合,范围很广,是向人们传递较系统全面的知识。对人们的文化及科学技术水平的不断提高,起着很大的作用。

3. 档案信息

档案信息是记述和反映人们在各自单位从事社会科学及自然科学研究、行政领导、生产技术和基本建设等实践活动中完成的特定现行效能,可作为真实的历史记录,具有保存价值;它是按

照一定归档制度集中保存起来的文献信息。它包括文书档案及科技档案等。

一般说来，档案信息具有如下特点：

(1) 档案信息必须是在某个单位形成的，记述和反映本单位在各项实践活动中的真实记录。

(2) 档案信息必须是完成了一定的现行效能的记录，是某个单位实践活动的历史凭证，而不是正在实践和使用着的文献信息。

(3) 档案信息是具有保存价值的文献信息。如果不是某个单位的真实历史写照，档案信息也就失去了凭证作用，这样也就没有保存价值。

(4) 档案信息必须按照一定的归档制度经过整理是具有专业性、完整性的文献信息。

(5) 档案信息必须是集中保管。为了避免重复劳动，更好地节约人力、物力、财力，充分发挥档案信息的作用，对用户提供利用及发挥档案信息的经济效益都是有利的。

总而言之，档案信息是指人们在各自的工作单位里，在从事不同的实践活动中形成的，能够反映实践活动单位的历史面貌，并能够起法律作用的信息集合。

4. 情报信息

情报信息是针对一定对象传递的有用的最新信息。它包括军事情报、政治情报、经济情报及科技情报等等。

一般说来，情报信息具有如下特点：

(1) 情报信息是有特定对象的。因为有些信息，对某些人来讲，是已知的或无用的，这样在他们今后的实践活动中，也就不起什么作用了。因而这些信息对于已知的人们来说，就不称为

情报了。反之，这些信息对于未知的人们来讲，是最新的信息，并在今后的实践活动中，能够起到继承、借鉴或参考的作用。那么，这些信息对于这些人们来讲，也就是情报了。

(2) 情报信息必须是经过传递了的文献信息。情报信息主要来源于文献信息，而文献信息包括档案、图书、期刊、报告等等，对于形成档案、图书、期刊、报告的单位或个人来讲，已不是最新信息，只有把档案、图书、期刊及报告等经过传递，才能够被人们所知道。所以，一旦被人们采纳利用了的文献信息，不再是档案、图书、期刊及报告，而变成了情报信息。

(3) 情报信息必须具有一定的使用价值。情报信息的使用价值是指对现实及未来的科学研究和生产建设等能够起继承、借鉴或参考作用。否则，人们收集情报信息也就没有必要了。

总而言之，情报信息是向人们传递有用的最新信息，不论对现实或未来所进行开发、研究和采取行动，都将成为最新的、最有用的激发性信息。

四、各种信息间的相互关系

文献、图书、档案、情报与信息间的关系可以用图 1—1 表示。

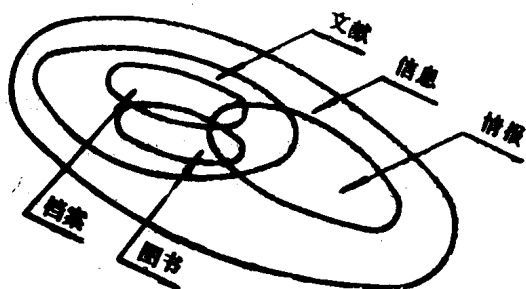


图 1—1 文献信息概念结构范畴图

由图 1 - 1 可以看出, 尽管它们的概念不同, 但是, 它们相互间范畴结构关系非常密切, 尤其档案、图书都同属于文献的范畴。而文献与情报又都同属于信息的范畴, 并且两者间可以相互转化。实践证明, 文献信息与情报信息间也确实存在着密切的关系, 因为文献信息是情报信息的主要载体; 情报信息主要是通过文献信息的传递进行交流的。文献信息是情报信息的重要来源之一。在情报工作中要有针对性地广泛搜集文献信息, 并且要进行文献工作 (如收集、分类、编目、保管、报道、借阅等等)。所以文献工作是情报工作的重要组成部分, 它不但是情报的基础, 而且渗透到情报工作各个环节中。因此, 情报学的研究必须借助文献学的研究成果。

由于档案与图书同属于文献的范畴, 根据情报与文献间的密切关系, 当然就可以看出档案、图书与情报间同样也存在着密切的关系, 尤其从图 1 - 1 可以看出档案、图书与情报的概念间存在着相关关系, 也就是交的关系, 因此相互间同样也是可以互相转化的。实际上, 在二十世纪五十年代以前, 世界各国科技情报工作与图书馆工作都是交织在一起的, 两者没有自成体系。随着科学技术的不断发展, 在工业发达的国家中, 科技情报工作的重要性也就越来越明显, 促使情报工作逐渐从传统的图书馆工作中派生出来, 形成自己的独立工作体系。但是, 图书馆工作和情报工作的关系一直是十分密切的, 两者的共性都是为社会的科学技术发展和文化进步服务, 都是向读者提供他们所需要的文献信息, 服务的对象也是相同的, 两者的工作手段都需要实现现代化。同时, 情报学在形成及发展过程中, 广泛应用了图书馆学中目录学的理论和方法, 并有所发展。当前图书馆学也在研究情报工作中的某些理论和方法, 以丰富其自身的内容, 使图书馆工作

能发挥情报工作的作用。

档案与图书一样,同样与情报间关系非常密切,可以互相转化。凡是形成的档案,它的复制品被情报部门搜集利用,则它就成为这个情报部门的情报资料。对情报部门来说,所形成的情报研究报告不仅是档案,而且是能提供服务的情报资料。由此可见情报工作和档案工作有很多共同之处:一是两项工作都是为实践需要提供服务的;二是两项工作的工作环节及内容都是相同的,即都是搜集、管理和提供利用的工作;三是两者提供利用的形式有些也是相似的,如编辑出版汇集、文摘、目录、索引及简介等。

从发展趋势看,档案、情报、图书、文献、新闻及文书等工作,将会统一起来处理。尤其随着计算机、通讯、网络、缩微、复印及数据库等项技术的高速发展及应用,以上各种工作,至少在信息自动化检索方面是一样的。所以本书介绍的文献信息自动化检索,包括档案信息、情报信息及图书信息等方面,它们的检索方法及步骤都是相同的。

第二节 研究文献信息的意义

文献信息是人们从事各项实践活动的记录,是人类社会宝贵的精神财富,它汇集着人类认识世界、改造世界的结晶,记载着许多成功的经验和失败的教训。随着科学技术的高速发展,各行各业要想立于不败之地,保持强有力的竞争力,就必须继承,借鉴或参考文献信息。对个人来说,也离不开各式各样的文献信息。所以,研究文献信息的意义是非常现实和深远的。

(1)文献信息是人类生存、发展的必需。人们生存不是目的,但生存确实是人们能够对人类社会有所贡献的必要前提。因此,