

QING BAO JIAN SUO

情报检索

王和珍 编著

上海交通大学出版社

计算机软件与应用丛书



计算机软件与应用丛书

情报检索

王和珍 编著

徐家福 审

上海交通大学出版社

内 容 简 介

本书主要介绍情报检索自动化技术及有关方法。全书共分九章：情报检索自动化导论、系统的开发、文档、文献的自动标引、主题词表、检索技术、微型机情报检索系统、汉字检索中的若干问题及国际联机情报检索。

本书可作为大专院校计算机专业、图书情报专业的教材或教学参考书，也可作为广大管理干部、科技干部和其他工程技术人员的学习参考书。对于从事图书馆或其他情报部门工作的人员，可利用书中介绍的算法和框图来设计本单位的情报检索自动化系统。

情 报 检 索

上海交通大学出版社出版

(淮海中路 1984 弄 19 号)

新华书店上海发行所发行

常州村前印刷厂印装

开本787×1092毫米 1/32 印张12.75 字数284,000

1987年12月第1版 1987年12月第1次印刷

印数：1—8000

标准书号：ISBN7-313-00104-S/TP39

定价：2.10 元

前 言

情报是一种财富、一种资源、一种具有社会、政治、经济价值的知识商品，是一个国家、一个社会赖以生存和发展的基础。在科学技术高度发展的今天，一切都离不开情报，情报的积累、传递和利用已成为革新和变革的直接动力。这与采用现代化科学技术来探查、开发和利用自然资源具有同样重要的意义。情报数量与日俱增，正以爆炸般的速度增长着，每约3年翻一番，各种各样的情报不断产生、增加，在人类的生产和社会活动中发挥着越来越重要的作用。电脑的发明、应用使情报的加工、处理获得有力的工具；电信和电脑的结合大大缩短了时间和空间的距离，使情报的收集、加工、存储、传输、检索和利用融为一体。缩微技术能把大量的情报记录于很小的载体面积上，有利于情报的交流和使用。这些新技术提供了对各种情报的实时处理和传递，提供了日新月异的良好服务，使得人类的知识和智慧有了突飞猛进，使得人类在信息化社会中获得前所未有的活动范围和创造能力。给社会生产力带来新的飞跃、给科学技术带来新的发展、给人类生活带来新的变化。

情报管理自动化是利用电脑系统地、科学地管理浩瀚的情报资料；情报检索自动化就是借助于电脑从存储的众多信息中检出所需情报的过程。本书介绍了这一领域的基本知识、实现技术和发展动向，强调实际、着眼应用。撰写本书的

目的是普及情报检索自动化的知识,推动我国情报检索自动化的发展。

本书第一、二两章扼要叙述了情报检索自动化的基本知识,第三、四、五、六章介绍了情报检索的主要技术,包括:文档、文献的自动标引、主题词表、检索技术;第七章着重讨论了实现微型机情报检索的必要性、可能性和微型机情报检索系统的具体构造方法;最后两章涉及汉字检索技术和国际联机情报检索系统的使用。

本书既可作为大专院校计算机专业、图书情报专业的教材或教学参考书;又可作为广大领导干部、管理干部和科技管理人员的入门书籍;对于从事图书馆或其他情报部门的工作人员,可利用书中的实现算法及框图设计本单位的情报检索自动化系统。书中介绍的设计思想和实现算法能方便地移植于人事管理、科技管理、设备管理、仓库管理等各个方面的计算机应用领域。

在编写过程中得到中国技术情报学会情报检索专业委员会主任曾民族研究员的大力帮助,并得到徐家福教授的审读,在此,致以衷心的感谢。

编者 1986年12月

《计算机软件与应用丛书》

(第一辑)

计算机数据安全原理	周锡龄	著
情报检索	王和珍	编著
数据库设计	施伯乐等	编著
程序设计方法学	朱三元	编著
微型计算机应用例解	白英彩等	编著

《计算机软件与应用丛书》编委会名单

顾问：刘振元 陈祥禄 朱雅轩

名誉顾问编：徐家福 杨芙清

主编：朱三元

编委：(按姓氏笔划排列)

白英彩 李玉茜 刘建明 何守才

陈涵生 陈娜芬 周锡龄 施伯乐

徐民祥 曹东启

目 录

第一章 情报检索自动化导论

§ 1.1 情报与情报管理自动化·····	(1)
§ 1.1.1 情报的重要性·····	(1)
§ 1.1.2 情报发展的现状·····	(3)
§ 1.1.3 图书情报工作的自动化·····	(4)
§ 1.2 情报检索自动化的发展·····	(8)
§ 1.2.1 脱机批处理情报检索阶段·····	(8)
§ 1.2.2 联机检索阶段·····	(9)
§ 1.2.3 计算机情报网络阶段·····	(10)
§ 1.3 我国计算机情报检索的现状和研究课 题·····	(12)
§ 1.3.1 我国计算机情报检索的现状·····	(12)
§ 1.3.2 今后的研究课题·····	(15)

第二章 系统的开发

§ 2.1 开发一个系统的必备条件·····	(18)
§ 2.1.1 情报检索系统的硬件·····	(18)
§ 2.1.2 情报检索系统的软件·····	(20)
§ 2.1.3 情报检索系统中的数据库·····	(20)
§ 2.2 系统开发的步骤·····	(24)
§ 2.2.1 系统分析·····	(24)

§ 2.2.2 系统设计·····	(25)
§ 2.2.3 程序设计·····	(29)
§ 2.2.4 运行·····	(30)
§ 2.2.5 评价·····	(30)

第三章 文档

§ 3.1 文档的设计·····	(32)
§ 3.1.1 设计原则·····	(32)
§ 3.1.2 应考虑的有关事项·····	(33)
§ 3.1.3 文档设计举例·····	(38)
§ 3.2 市售文档·····	(41)
§ 3.2.1 市售文档出现的背景·····	(41)
§ 3.2.2 磁带的记录形式·····	(44)
§ 3.2.3 ISO-2709国际标准简介·····	(48)
§ 3.2.4 西文文献磁带生成程序·····	(48)
§ 3.3 中华人民共和国国家标准《文献目录 信息交换用磁带格式》GB 2901-82···	(50)
§ 3.3.1 应用范围·····	(53)
§ 3.3.2 定义·····	(53)
§ 3.3.3 磁带规格·····	(54)
§ 3.3.4 目录记录格式·····	(54)
§ 3.3.5 有关文件·····	(61)
§ 3.3.6 编制原则·····	(62)
§ 3.3.7 汉字文献磁带生成程序·····	(64)

第四章 文献的自动标引

§ 4.1 概论·····	(67)
---------------	--------

§ 4.1.1 自动标引的必要性·····	(67)
§ 4.1.2 自动标引方式·····	(67)
§ 4.2 索引的自动编制·····	(68)
§ 4.3 词组作关键词语的标引法·····	(78)
§ 4.3.1 关键词组法标引系统的结构·····	(79)
§ 4.3.2 关键词组法标引系统的评价·····	(88)
§ 4.3.3 系统实例·····	(89)

第五章 主题词表

§ 5.1 主题词表简介·····	(90)
§ 5.1.1 词汇规范化的必要性·····	(90)
§ 5.1.2 主题词表的定义·····	(93)
§ 5.2 主题词表的构造·····	(94)
§ 5.2.1 构造主题词表的原则·····	(94)
§ 5.2.2 收词来源·····	(95)
§ 5.2.3 选词原则和习惯·····	(95)
§ 5.2.4 词汇类型·····	(97)
§ 5.2.5 主题词表参照系统·····	(99)
§ 5.2.6 主题词表的结构·····	(104)
§ 5.3 计算机编制主题词表·····	(112)
§ 5.3.1 机器可读主题词表的建立·····	(112)
§ 5.3.2 机器可读主题词表 MVT 的维护·····	(116)
§ 5.3.3 机器可读主题词表 MVT 的作用·····	(117)
§ 5.4 汉语主题词表·····	(123)
§ 5.4.1 汉语主题词表简介·····	(123)
§ 5.4.2 汉语主题词的组配·····	(124)

§ 5.4.3 汉语主题词表的机内表示·····	(125)
§ 5.5 主题词表生成系统实例·····	(129)
§ 5.5.1 建表原则·····	(129)
§ 5.5.2 系统实现·····	(129)

第六章 检索技术

§ 6.1 检索系统·····	(144)
§ 6.1.1 检索类型·····	(144)
§ 6.1.2 检索步骤·····	(145)
§ 6.2 顺排文档的检索·····	(152)
§ 6.2.1 提问档的建立·····	(153)
§ 6.2.2 顺排文档的检索处理·····	(163)
§ 6.2.3 对菊池敏典算法的改进·····	(175)
§ 6.3 倒排文档的检索·····	(180)
§ 6.3.1 倒排文档的建立·····	(180)
§ 6.3.2 提问的编辑·····	(182)
§ 6.3.3 倒排文档的检索过程·····	(201)
§ 6.3.4 工作区的使用策略·····	(203)
§ 6.3.5 文档适用的提问逻辑式的展开方式及 查找方式·····	(205)
§ 6.3.6 检索语言简介·····	(208)
§ 6.4 联机检索·····	(212)
§ 6.4.1 联机检索所需的条件·····	(212)
§ 6.4.2 联机检索系统特有的功能·····	(213)
§ 6.4.3 联机检索举例·····	(215)
§ 6.4.4 网络环境下的情报检索系统的设想·····	(226)

第七章 微型机情报检索系统

- § 7.1 在微型机上实现情报检索的可能性…(233)
 - § 7.1.1 微型机系统的组成……………(233)
 - § 7.1.2 用微型机实现情报检索的可能性……(236)
 - § 7.1.3 图书情报部门与微型机……………(239)
- § 7.2 微型机在情报检索系统中的应用……(241)
 - § 7.2.1 作为大、中型联机情报检索系统的
“智能终端”……………(241)
 - § 7.2.2 作为联机数据库间的转换接口和信息
格式间的转换设备……………(250)
 - § 7.2.3 作为独立的情报检索系统……………(251)
- § 7.3 微型机在图书情报资料管理中的应
用……………(291)
 - § 7.3.1 微型机在图书流通管理中的应用……(291)
 - § 7.3.2 微型机在书刊采购、编目中的应用…(300)
 - § 7.3.3 利用微型机编制索引、题录、文摘等
刊物……………(301)
 - § 7.3.4 微型机在电视资料显示系统中的应
用……………(310)
- § 7.4 我国发展微型机情报检索的道路……(311)

第八章 汉字检索中的若干问题

- § 8.1 汉字处理问题及现状……………(315)
 - § 8.1.1 汉字编码……………(315)
 - § 8.1.2 汉字输入……………(318)

§ 8.1.3 汉字输出·····	(319)
§ 8.1.4 汉字处理标准化·····	(319)
§ 8.2 汉语文献的自动分词和自动标引·····	(320)
§ 8.2.1 问题的提出·····	(320)
§ 8.2.2 发展概况·····	(321)
§ 8.2.3 方案介绍·····	(321)
§ 8.3 其他研究课题及存在问题·····	(328)
§ 8.3.1 计算机汉字编辑排版·····	(328)
§ 8.3.2 机器翻译问题·····	(329)
§ 8.3.3 中文建库问题·····	(330)

第九章 国际联机情报检索

§ 9.1 国际联机情报检索简介·····	(331)
§ 9.1.1 什么叫国际联机情报检索·····	(331)
§ 9.1.2 国际联机情报检索系统的功能·····	(332)
§ 9.1.3 国际联机情报检索系统的作用·····	(332)
§ 9.1.4 国际联机检索系统简介·····	(333)
§ 9.1.5 我国开展国际联机检索的概况·····	(333)
§ 9.2 国际上的四大联机情报检索系统·····	(334)
§ 9.2.1 DIALOG系统·····	(334)
§ 9.2.2 ORBIT系统·····	(337)
§ 9.2.3 ESA-IRS联机情报检索系统·····	(339)
§ 9.2.4 MEDLINE 联机情报检索系统·····	(342)
§ 9.3 国际联机检索系统的基本检索方法·····	(343)
§ 9.3.1 联机检索的基本步骤和方法·····	(344)
§ 9.3.2 检索指令·····	(346)

§ 9.3.3 实例·····	(362)
§ 9.3.4 使用自由词全文查找的检索方法·····	(363)
附录 A 图书目录系统用 008 字段结构说明(参考件)	
·····	(366)
附录 B 图书目录系统用的数据区内容(参考件)···	(371)
附录 C 情报检索系统用的数据区内容(参考件)···	(375)
附录 D 建议采用的其他有关标准或资料(参考件)	
·····	(387)
附录 E 常用术语解释·····	(388)
参考资料·····	(392)

第一章 情报检索自动化导论

§ 1.1 情报与情报管理自动化

§ 1.1.1 情报的重要性

人们在社会生活中息息相通，人类的实践经验和积累的知识代代相传，从这个意义上说，有了信息的传播和交流，人类才能不断进步，人类社会才能不断发展。当代，科学技术高度发展，各种科学知识急剧增长，人类智力的开发出现了新的飞跃。从一百亿光年之遥的外空世界到物质内部的基本粒子，以至更细微的物质层次，都在人类研究与观测之列。所有这一切都离不开“情报”。

情报、材料、能源被称为当代社会的三大支柱。就某种意义上来说，情报较之材料、能源对社会的发展具有更为重要的作用。例如，日本是个资源贫乏的国家，但是，战后日本的经济增长速度很快，在短短的三十多年间，一跃成为仅次于美、苏的世界第三经济大国。甚至在某些生产科技领域还超过了美、苏两国，究其原因，主要是他们善于开发，利用情报资源。战后日本的石油化工、电子等新兴工业和机械、冶金等传统工业的新技术，几乎全盘来自国外，都是利用情报搞成的；有些最新技术，如数控机床、转子发动机

等，则是在别国已经接近完成的情况下，通过情报工作，在别人的技术基础上搞成的。据报道，自1945年起，日本花了4亿多美元购买了3500多项专利，经过吸收、提高后，取得的效益相当于70亿美元。60年代他们几乎用了1/3的科研费用购买专利和许可证。反之一个国家即使拥有丰富的物质资源，但若不注意利用情报资源，其经济增长速度不可能很快。因为无论是科学研究还是生产发展，只有通过情报，才能做到在他人的最新成就的基础上去进行，避免重复劳动，有选择地引进某些必需的适用技术，有选择地确定最具实际意义的科研项目。据统计，如果科研费用是1，则中间试验费用是10，实现工业化生产所需投资则是100~200，而情报费用只占科研费的2%~5%。实践证明，通过很少的投资得来的情报往往能获得几十倍~几百倍的经济效果。由于情报不灵，造成科研重复的浪费现象是很严重的。例如，60年代初期的美国兵工系统，由于科研重复，每年损失达10~12亿美元，占同期兵工研制费用的1/10。

目前，我国科研工作中的重复劳动较多，据杨振宁在1979年我国物理学会上的讲话，当时我国的科研项目有40%和国外同行重复。这种情况的最主要的原因是情报不灵。因此，搞好情报工作，特别是关于正在进行中的科研项目的情报交流活动，对于加速科研进程，促进生产发展具有十分重要的意义。

现在有种流行的说法：“情报就是财富”、“情报就是人类的第二资源”。从美国1860~1980年这120年间四个最活跃的行业——农业、工业、服务业和情报业的劳动力的变化，即可看出美国越来越重视情报工作(见表1.1)。

表 1.1 近120年来美国四大行业劳动力变化表

行业	年份	1860	1880	1900	1920	1940	1950	1960	1970	1980
	百分比									
农 业		41	44	35	33	15	11	6	4	3
工 业		37	26	28	34	37	39	35	30	22
服 务 业		16	23	25	17	23	19	18	21	29
情 报 业		6	7	12	16	25	31	41	45	46

§ 1.1.2 情报发展的现状

19世纪,世界上的科技文献还十分有限,就期刊而言,只有100多种,到20世纪60年代,科技期刊种类达到1万种,到1980年已超过了5万种,翻了几番。每年都有大量的新发明和新专利公诸于世,每年发表的科技论文在500万篇以上。专利登记约70~100万件,文献量以每年7%~8%的速度增长,平均10~15年就翻一番。某些尖端科学,如原子能,计算机技术等,大约每2~3年就增长一倍。

一种新的发现会导致出版物数量的增加。如1960年红宝石激光器这一课题的论文约有20篇;1961年关于氩-氟激光器的论文约有100篇;1962年有关固态激光器的论文有325篇;1963年镱砷激光器、脉冲激光器和Q开关的论文有700篇;1964年离子激光器的论文有1000篇;1965年有关 N_2-CO_2 高效激光器的论文有1200篇。

随着科学技术的不断发展,分支学科越来越多,边缘科学不断出现,有人估计达1000多种。各学科之间交叉渗透现象也越来越普遍,与某课题有关的技术文献往往散登在几十

种，乃至上百种刊物上。

国外有统计材料表明，科研成果每增加一倍，情报量就要增加几倍；生产翻一番，情报量则要增加四倍。

目前全世界拥有的有效文献量总量达1亿件以上。并以每年几百万件的数量在增加。

文献的另一部分还有非公开发表的科技报告、会议文献、学位论文等。除纸质印刷品外，还有各种视听资料（电视、电报、电话、广播等所传播的情报），计算机阅读资料（磁盘、磁带、光盘、磁泡等）以及各种缩微品。

科研人员在着手从事某一课题的研究时，首先必须查阅大量的国内外有关技术文献，以了解国内外同行在这方面已有的研究成果、目前的水平、存在的问题和发展动向，以便拟定切实可行的科研方案。在研究过程中也同样需要不断地查阅技术文献。然而，面对这浩如烟海的技术文献，要迅速查到针对性强，有参考价值的资料，没有现代化的技术装备，使人们感到如大海捞针。其结果必然是：花费大量的查阅技术文献的时间（有人估计占用科研时间的1/3到1/2），但难以查准、查全，导致科研工作的重复、延迟和严重的浪费。

§ 1.1.3 图书情报工作的自动化

科技情报部门如何迅速提供针对性强的情报资料，帮助科研人员缩短研究周期，避免或少走弯路，这是科学技术发展的重要条件之一。科技情报本身也是一种资源，用现代化的科学技术来装备科技情报工作与采用现代化的科学技术来寻找、开发和利用自然资源具有同样重要的意义。

电子计算机的出现和发展，为科技情报工作现代化提供