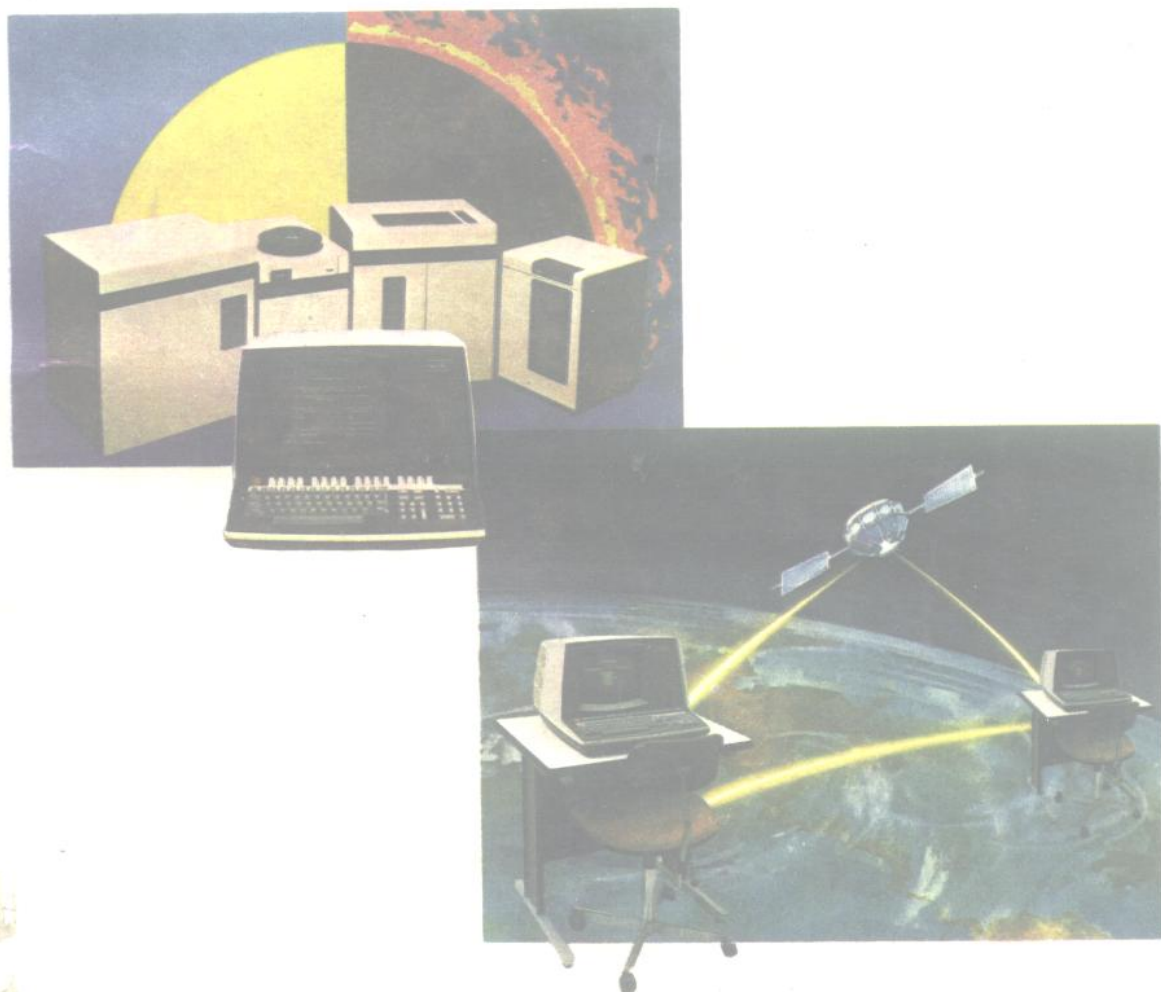


计算机情报检索

江向东 李秋山 编



机械工业出版社

计算机情报检索

江向东 李秋山 编



机 械 工 业 出 版 社

内 容 简 介

本书比较详尽地介绍了计算机情报检索系统和数据库管理系统原理、设计及应用方面的知识,主要内容有:情报检索一般知识,计算机情报检索基本原理,数据库管理系统,情报系统设计和开发,情报检索程序设计,联机检索,联机情报检索系统和软件分析,数据通讯和汉字信息的计算机处理。书中给出了作者自行设计的定题服务检索系统和微型机联机检索系统的完整流程图和详细说明,对国外九个比较著名的数据库管理系统进行了介绍,对国外八个著名的联机检索系统(DIALOG、ORBIT、ESA-IRS、BIRDS、STAIRS、ISIS、MINISIS和JOIS)的数据结构和检索命令进行了详细的分析。

本书的读者对象:计算机情报检索系统和数据库管理系统的系统分析员、系统设计员、系统维护员和程序人员;科技情报和图书部门的领导和一般工作人员;软件人员和其他科技人员;大专院校和研究生院有关专业的师生;对信息处理感兴趣的人员。本书可作为教科书或参考书,也可作为自学材料。

计算机情报检索

江向东 李秋山 编

责任编辑: 邱锦来

封面设计: 刘 代

机械工业出版社出版(北京阜成门外百万庄南里一号)
(北京市书刊出版业营业许可证出字第117号)

通县大中印刷厂印刷

新华书店北京发行所发行·新华书店经售

开本 787×1092 1/16·印张 214 16·字数 513 千字

1987年12月北京第一版·1987年12月北京第一次印刷

印数 0,001—6,070·定价: 5.35元

统一书号: 15033·6579

2180128

前 言

计算机逐渐普及,人类正步入“信息化”社会。在高度信息化的社会里,人们可在办公室和家庭中,使用终端设备或微型计算机,通过四通八达的数据通讯网络,去存贮、处理和查找自己需要的一切数据和情报,去分享人类知识宝库中取之不尽,用之不竭的信息资源。目前,计算机情报检索网络的迅速发展已使人们看到了信息化社会的美好前景。

计算机情报检索的产生和发展,不仅是科技情报战线上具有划时代意义的事件,也是人类社会的重大事件之一,在某种意义上说,计算机情报检索与数据处理奠定了信息化社会的技术基础。当前,以情报检索中心的计算机系统为核心的情报网络实际上是信息化社会的缩影。

随着科技现代化和社会信息化的进展,很多科技工作者希望使用计算机去查找他们所需要的科技情报资料,很多领导和行政管理人员渴望借助计算机的帮助去处理和检索各种文件和信息,图书情报部门的广大工作人员急切需要用计算机去存贮和加工浩如烟海的书刊资料,可以断言,将有越来越多的人希望学习和掌握这方面的知识。然而,在我国,系统地介绍计算机情报检索知识的书(包括中译本)还屈指可数。

十多年来,作者一直从事计算机情报检索系统和数据库管理系统方面的研究、设计和开发工作,并喜爱收集、涉猎、阅读和研究这方面的书刊资料,久而久之,积累了一些有价值的情报和资料,取得了一点知识和经验,现将它们整理出来,浓缩在这本书中,奉献给同行或对这门科学技术感兴趣的读者。由于作者水平有限,书中定有不当和错误之处,恳请专家学者及读者不吝赐教、及时指正。他年若有再版机会,必当吸取大家的宝贵意见,悉心修正此书。

知识性和实用性兼备,雅俗共赏是本书的写作宗旨,情报检索系统和软件、数据库管理系统和软件是本书的写作重点。作者试图通过这本书使读者对情报检索系统和数据库管理系统方面的知识有一个比较全面的了解,所以本书的内容比较广泛,既有基本概念、基本原理和理论,又有实用系统的介绍和分析(例如本书对国外九个著名的数据库管理系统和八个著名的联机情报检索系统作了不同深度的介绍和分析),既有系统分析、系统设计和系统评价的一般论述,又有检索程序的全套框图,既回顾过去的发展历史,又评论当代,展望未来,既谈西文存取,又谈汉字处理,既介绍大中型计算机情报检索系统,又介绍微型机情报检索系统。这样求全是否会造成面面俱到,泛泛而论的结果,请读者去评价。

本书共分十章,章节内容的安排原则是:由浅入深,由一般叙述到实例介绍,由数据库管理系统到情报检索系统,由系统设计到系统使用,由西文处理到汉字处理。

第一章由二作者合写,第二、三、四、五、七、八、九、十诸章由江向东编写,第六章由李秋山编写。曾民族同志对书稿作了认真而细致的审阅和修改,在此深表谢意。

由于篇幅所限,本书仅列出了近一百篇参考文献,还有更多参考文献未能一一列出,在此顺向书中列出姓名的作译者和未曾列出姓名的作译者致谢。

编者

1985年11月

目 录

第一章 情报检索的一般知识

§ 1-1 概论..... (1)	一、情报检索语言的概念和作用..... (8)
一、概述..... (1)	二、对情报检索语言的要求..... (9)
二、情报检索研究的内容..... (2)	三、情报检索语言的分类..... (10)
三、情报检索的类型..... (2)	四、情报检索语言的发展趋势..... (11)
§ 1-2 情报检索的方法和步骤..... (4)	§ 1-4 情报检索的效率..... (11)
一、分析研究课题, 确定主题范围..... (5)	§ 1-5 计算机情报检索的发展历史..... (12)
二、选择检索方法和检索工具..... (5)	§ 1-6 计算机情报检索的现状和发
三、确定检索途径和检索标识..... (7)	展方向..... (14)
四、利用检索工具查找所需情报	一、各国现状..... (14)
线索..... (8)	二、数据库..... (15)
五、查找原文、获取情报..... (8)	三、通讯网络和终端设备..... (15)
§ 1-3 情报检索语言..... (8)	四、联机情报检索服务..... (15)

第二章 计算机情报检索的基本原理

§ 2-1 序言..... (17)	二、检索条件和比较条件..... (20)
§ 2-2 情报结构..... (17)	三、加权检索..... (21)
§ 2-3 检索原理..... (18)	四、词表助检..... (21)
§ 2-4 检索技术..... (19)	五、提问分析方法..... (22)
一、逻辑检索..... (19)	

第三章 数据库管理系统

§ 3-1 数据库概念..... (23)	六、第三范式..... (41)
一、数据处理的重要性..... (23)	§ 3-4 物理结构..... (45)
二、基本数据术语..... (23)	一、物理结构与逻辑结构的区别..... (45)
三、数据库的产生和发展..... (24)	二、指示字..... (46)
四、数据库特性..... (25)	三、文件寻址..... (47)
§ 3-2 数据库结构概述..... (27)	四、检索技术..... (50)
一、数据库结构的目的..... (27)	五、倒排文件系统..... (55)
二、数据库结构的类型..... (28)	§ 3-5 数据库管理系统..... (56)
§ 3-3 逻辑结构..... (28)	一、数据库管理系统的发展..... (56)
一、实体和属性..... (28)	二、数据库管理系统的性能要求..... (58)
二、数据模式..... (29)	三、数据库语言..... (61)
三、树结构..... (33)	四、数据库管理系统存取记录的基
四、丛结构..... (34)	本流程..... (65)
五、关系数据库..... (34)	五、几个实用的数据库管理系统..... (66)

第四章 情报系统的开发

§ 4-1 情报系统的开发步骤及逻辑		一、一般系统设计的评价和论证	
结构表示方法.....	(84)	流程.....	(94)
一、情报系统的开发步骤.....	(84)	二、一般系统设计完成报告的内容.....	(94)
二、情报系统的逻辑结构表示方法.....	(84)	§ 4-7 详细系统设计.....	(95)
§ 4-2 计算机与情报系统设计.....	(87)	一、系统流程图.....	(95)
一、计算机在情报系统中的地位.....	(87)	二、输出设计.....	(95)
二、计算机化情报系统的基本组成.....	(87)	三、输入设计.....	(96)
§ 4-3 情报系统的结构.....	(89)	四、详细系统设计报告.....	(97)
一、集中式系统.....	(89)	§ 4-8 系统实施.....	(98)
二、带有若干小型计算机的集中式系统.....	(89)	一、人员培训.....	(98)
三、分布式系统.....	(89)	二、系统检验.....	(99)
四、集中式与分布式系统的比较.....	(90)	三、系统转换.....	(99)
§ 4-4 系统分析.....	(91)	§ 4-9 情报系统的维护.....	(101)
一、系统分析的重要性及系统分析员的任务.....	(91)	一、系统维护的种类.....	(101)
二、如何进行系统分析.....	(91)	二、系统维护的几个问题.....	(102)
§ 4-5 一般系统设计.....	(92)	三、提高系统的可维护性.....	(102)
一、与系统设计有关的若干问题.....	(92)	四、系统检查.....	(102)
二、一般系统设计的基本步骤.....	(93)	§ 4-10 情报系统成败的因素.....	(103)
§ 4-6 一般系统设计的评价和论证.....	(94)	一、情报系统失败的原因.....	(104)
		二、情报系统成功的条件.....	(104)

第五章 情报检索程序设计

§ 5-1 文献磁带.....	(107)	五、程序执行.....	(116)
一、概述.....	(107)	§ 5-4 情报检索程序概述.....	(117)
二、INSPEC磁带.....	(109)	一、什么是情报检索程序.....	(117)
§ 5-2 定题服务.....	(113)	二、情报检索程序的特点.....	(118)
一、概述.....	(113)	§ 5-5 磁带检索程序设计举例.....	(119)
二、所需技术条件.....	(113)	一、序言.....	(119)
§ 5-3 程序设计概述.....	(115)	二、磁带检索的基本流程.....	(119)
一、计算机语言.....	(115)	三、提问加工.....	(120)
二、流程图.....	(116)	四、生成检索词表.....	(126)
三、程序.....	(116)	五、检索处理.....	(128)
四、程序的输入、编辑和编译.....	(116)	六、打印输出.....	(131)

第六章 微型计算机情报检索系统

§ 6-1 概论.....	(133)	六、怎样选择微型计算机.....	(136)
一、微型计算机的基本概念.....	(133)	七、微机在国内外图书情报部门的应用情况.....	(136)
二、微型计算机的分类.....	(134)	§ 6-2 一个实用的微机联机情报检索系统介绍.....	(138)
三、微型计算机的组成体系.....	(135)	一、概论.....	(138)
四、微型计算机的特点.....	(135)		
五、微型计算机的应用范围.....	(135)		

二、系统功能.....	(139)
三、系统特点.....	(139)
四、系统环境.....	(139)
五、资料档.....	(139)
六、L 77 80情报检索系统数据库.....	(142)
七、系统功能与程序框图.....	(144)

八、几个问题的讨论.....	(152)
§ 6-3 微机定题情报检索系统.....	(157)
一、导言.....	(157)
二、检索方式.....	(157)
三、程序结构.....	(163)

第七章 联机检索

§ 7-1 导言.....	(168)	二、文献数据库.....	(173)
一、联机检索的基本原理.....	(168)	三、事实数据库.....	(173)
二、联机检索的优缺点.....	(168)	四、源文献类型.....	(175)
§ 7-2 联机检索的发展历史.....	(170)	五、数据库评价.....	(175)
§ 7-3 终端设备.....	(171)	§ 7-5 联机检索操作.....	(177)
§ 7-4 联机数据库.....	(172)	一、联机检索前的准备工作.....	(177)
一、概况.....	(172)	二、联机检索的基本过程.....	(179)

第八章 联机情报检索系统与软件分析

§ 8-1 DIALOG.....	(184)	五、检索功能.....	(236)
一、概况.....	(184)	六、排序和打印功能.....	(237)
二、检索功能.....	(185)	§ 8-6 STAIRS.....	(238)
§ 8-2 ORBIT.....	(193)	一、概况.....	(238)
一、概况.....	(193)	二、数据结构.....	(239)
二、操作命令.....	(194)	三、数据库的建立和维护.....	(242)
三、检索技术.....	(197)	四、联机对话系统 AQUARIUS.....	(244)
§ 8-3 ESA-IRS/QUEST.....	(199)	五、STAIRS的优缺点.....	(247)
一、概况.....	(199)	§ 8-7 JOIS.....	(248)
二、数据结构及数据压缩.....	(202)	一、概况.....	(248)
三、数据库创建.....	(207)	二、JOIS-I软件结构.....	(248)
四、数据库维护.....	(210)	三、JOIS-I文件.....	(251)
五、联机检索系统.....	(211)	四、JOIS-I命令及功能.....	(251)
§ 8-4 BIRDS.....	(219)	五、JOIS-II命令及功能.....	(254)
一、概况.....	(219)	§ 8-8 MINISIS.....	(254)
二、数据结构.....	(220)	一、概况.....	(254)
三、数据库的创立和维护.....	(225)	二、功能分析.....	(255)
四、查询子系统.....	(228)	三、系统结构.....	(256)
§ 8-5 CDS/ISIS.....	(231)	四、数据结构.....	(258)
一、概况.....	(231)	五、数据库创立和维护.....	(263)
二、系统的主要特点.....	(232)	六、关系的投影和连接操作.....	(264)
三、数据结构.....	(233)	七、联机检索.....	(266)
四、数据库维护.....	(235)		

第九章 数据通讯

§ 9-1 序言.....	(280)	§ 9-3 数据传送技术.....	(281)
§ 9-2 数据通讯系统的构成.....	(280)	一、数据传送的基本概念.....	(281)

二、数据传输的基本技术.....	(282)	三、前端处理机.....	(289)
§ 9-4 数据通讯信道.....	(283)	§ 9-8 典型的前端处理机配置.....	(290)
一、信道的种类.....	(283)	§ 9-9 数据通讯网络协议和分层结构.....	(292)
二、信道频级.....	(284)	§ 9-10 国外数据通讯网络的现状及发展方向.....	(293)
§ 9-5 调制解调器.....	(285)	一、美国.....	(293)
一、调制解调器的工作原理.....	(285)	二、欧洲共同体.....	(293)
二、调制解调器的类型.....	(285)	三、英国.....	(293)
三、调制技术.....	(286)	四、法国.....	(293)
§ 9-6 多路转换器和集线器.....	(287)	五、联邦德国.....	(293)
一、多路转换器.....	(287)	六、日本.....	(294)
二、集线器.....	(287)	七、新式数据通讯网络.....	(294)
§ 9-7 可编程序的通讯处理机及其应用.....	(288)	八、数据通讯网络的发展趋向.....	(295)
一、可编程序的通讯处理机.....	(288)		
二、信息转接.....	(289)		

第十章 汉字信息的计算机处理

§ 10-1 前言.....	(296)	五、汉字输出方式和设备.....	(308)
§ 10-2 汉字信息处理技术的发展和现状.....	(296)	六、汉字信息处理软件系统.....	(310)
一、汉字信息处理技术的发展历史.....	(296)	§ 10-4 汉字自动排版系统.....	(313)
二、汉字信息处理技术的现状.....	(298)	一、汉字自动排版系统的组成.....	(313)
§ 10-3 汉字信息处理系统的几个主要问题.....	(299)	二、自动排版系统的工作流程.....	(314)
一、前言.....	(299)	三、蒙纳中文激光照排系统简介.....	(316)
二、系统用字.....	(300)	§ 10-5 汉字情报检索系统.....	(318)
三、汉字的输入编码、输入方式和设备.....	(300)	一、汉字情报检索系统的基本特点.....	(318)
四、汉字字库.....	(307)	二、汉字情报检索系统的开发方法.....	(319)
		三、将MINISIS改造成汉字检索系统的方法.....	(320)

第一章 情报检索的一般知识

§ 1-1 概 论

一、概述

情报检索 (Information Retrieval) 一词是1950年 Calvin N. Moore在一次国际数学家会议上提出来的, 迄今为止, 该词并没有公认一致的明确定义, 虽然众说不一, 但大部分说法相差不多。例如, 苏联切尔内认为, 情报检索是从大量文献中查找含有用户提问中指定特征的文献或含有用户所需事实的文献的过程, 英国维克利解释说, 情报检索是指从汇集的文献情报中找出特定用户在特定时间所需情报的操作过程。

实际上, 情报检索包括情报存贮和检索两个方面, 在检索之前, 必须将情报收集起来, 按科学方法进行整理, 并按一定规律存贮起来, 形成书本式检索工具或计算机可读数据库。在检索时, 通过手工方法, 或者使用计算机, 按照存贮情报所依据的规律, 将用户所需求的情报找出来。由此可知, 在情报检索过程中, 人、情报和计算机或其他检索工具是情报检索系统的三个基本要素, 将用户所需情报的特征与书本上或数据库中所含情报的特征进行比较, 将比较一致的情报取出来则是情报检索的基本过程 (参见图 1-1)。

随着社会的信息化, 人们对Information一词的理解和认识都比原来的“情报”广泛得多, 它是指包括科技情报、经济情报、管理情报乃至生活情报等在内的“信息”。我国广大的科技情报工作者仍习惯地采用意义比较狭窄的“情报”和“情报检索”的叫法。我们认为, 按更确切、更科学的理解, 采用“信息”和“信息检索”的叫法更好一些。

当前, 在新的产业革命中, 以计算机技术为代表的新的信息技术正在推动科学技术、工业农业, 以至人们生活各个方面的迅猛发展, 也在引起科技情报工作或“情报检索”的技术变革。我们相信, “信息检索”将比传统的“情报检索”更能得到人们的理解。

现代的情报检索不仅包括科技情报工作中的文献检索, 还包括数据检索和事实检索, 而且数据和事实检索的发展很快, 应用越来越广。例如, 档案、病历、清单、产品、设备、人员、物资的存贮和检索实际上也是一种广义的情报检索, 如果将它们称为信息检索更为确切。

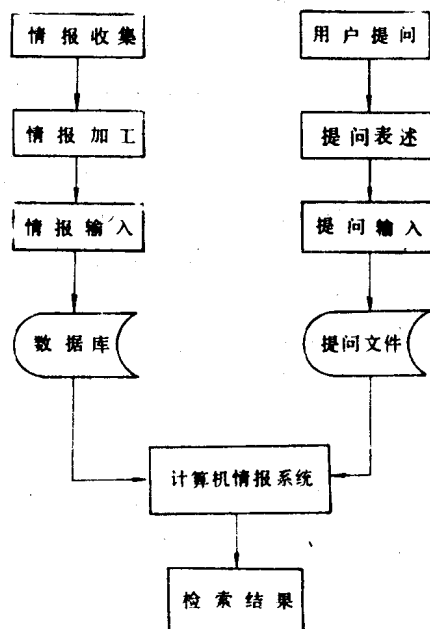


图 1-1 情报检索的基本过程

本书仍沿用情报检索一词，并通过科技文献的存贮、处理和查找来介绍以计算机作为检索工具的情报检索的基本原理和情报检索系统。

情报检索是科技情报工作的重要环节。随着科学技术和工农业生产的迅速发展，每年产生的文献资料和科学数据也迅速增加，日本有人称之为“情报爆炸”。另一方面，由于情报资料来源多，内容交叉等原因，使得查找情报的工作变得十分复杂和困难，如果不采用现代化的情报检索技术和手段，用户要获得自己所需的情报确实如大海捞针。

最近十几年，计算机技术在信息处理方面的应用越来越普遍，情报检索已发展成为计算机信息处理的重要组成部分和分支学科。当前，情报检索已经成为计算机科学、情报学、科学学和管理科学等学科共同发展、互相渗透的一门新兴的科学技术。

二、情报检索研究的内容

计算机在情报工作中的应用，不仅给情报检索带来了巨大变革，而且也带来了一系列新问题，需要人们研究解决。就文献检索而言，它依据的主要基础是图书馆学、目录学和计算机科学。情报检索与信息科学和管理科学也有密切的关系。无论是手工检索或计算机检索，都会涉及到情报检索语言、情报检索系统、情报检索策略和情报检索服务等方面的问题，这都是情报检索研究的内容。

三、情报检索的类型

可以从不同的角度对情报检索进行分类，例如，按照被检索的内容，可分为文献检索、数值检索和事实检索三类；按被检索情报的年代，有定题检索和追溯检索两类；按检索手段有手工检索和计算机检索两类。下面对这些类别的情报检索作简略的介绍。

（一）文献检索

文献检索的对象是文献资料，这是最典型的一种情报检索，国外一些大中型的情报检索系统所建的数据库大部分都是文献型的数据库，我们可以将文献检索理解为情报检索的同义词。

文献数据库所存贮的文献项目一般与书本式文献一样，有篇名、作者、文章出处、出版年月，文章摘要，还有一些能够表达文章主题内容的标识项，如主题词、分类码等。由于磁盘容量所限，绝大多数文献数据库都不存贮文章全文，有些也不存贮文摘。

书本式文献是手工检索的重要工具，人们通过书本式文献可以获得需求情报的线索，即获得所需文章的出处，然后根据这些线索便可找到原文。书本式文献的种类很多，如果靠手工方法一本一本地进行翻阅，那就需要花费科技人员的大量精力和时间，而且要获得较高的查全率是比较困难的。

现代化的情报检索系统是以计算机为核心的系统，其检索速度是手工检索无法比拟的。人们通常将使用计算机来存贮、处理和查找情报的技术和方法简称为计算机情报检索，以便与手工检索相区别。使用计算机进行情报存贮和检索，在西方已日益普及，所以，情报检索一词常常是计算机情报检索的代名词或简称，情报检索系统实际上也是指计算机情报检索系统，而不是指手工检索系统。

计算机情报检索的基本原理及系统是本书的主题，将在第二章至第十章中介绍。

（二）数值检索

文献一般是以文学形式表示的，文献检索的对象是文字，而数值检索的对象是以数字表示的数值，如生产指标，统计数据、物价、物质材料的特性、化学分子式、数学公式等数据

库都属于数值型数据库,对这些数据库的检索可称为数值检索。文献检索的结果一般并不能解答用户要解决的任何问题,它只是提供用户查找原文的线索,即使找到了原文也只是作为参考,不一定有现成的答案。数值检索的结果可供用户直接使用,如钢铁产量、国民产值、电视机价格、黄金的密度、聚氯乙烯的分子结构,等等。数值数据库必须有高度的准确性,所以,数值的搜集、加工和输入必须非常细心,鉴定数据是一件非常重要但又十分复杂的工作。

数值数据库的发展速度已超过了文献数据库的发展速度,这是因为数值检索对于制定国民经济计划,工程设计,人民的日常生活,都有非常重要的意义。例如,建筑数值数据库,既可以存贮各种建筑材料的型号、强度、刚度和其他物理、化学特性,也可以存贮各种建材产品的型号、规格和价格;物理数据库可以向用户提供物质的密度、比热、沸点、熔点、拉力和压力等参数;热力学数据库可存贮各种物质的热力学特性和计算公式。用户检索这些数据库,可以获得他们所需的数据、数字或数值,而不必花费大量时间去翻阅各种厚厚的手册。

(三) 事实检索

如果我们要了解与某一事物或某一事件有关的情况(如事件发生的时间、地点和过程)可以检索事实型数据库,这种数据库是在数值型数据库的基础上发展起来的。例如,如果我们输入“我国哪座山最高”这样的提问,也可以查询“世界上最大的一次空运事故”的有关情况。

国外的一些大型检索系统,如美国洛克希德公司的DIALOG,美国系统开发公司的ORBIT和欧洲航天局的ESA-IRS(这三大系统的名称在以后的章节中将许多次提及到,它们的详细情况将在第八章中介绍),不但建有大批文献数据库,而且也建立了若干个数值数据库和事实数据库供用户检索。

(四) 定题检索和追溯检索

定题检索和追溯检索实际上是代表两种不同的情报服务方式,这两种情报服务方式主要是指计算机情报检索而言的,当然,也适用于手工检索。

定题检索或定题服务方式是美国学者H. P. 卢恩首先提出来的,其英文名称是Selective Dissemination of Information,国内外书刊通常将这种服务方式简称为SDI。SDI的字面含义是选择性情报提供,提供方法是,情报提供单位按一定周期(通常是每月一次)和用户的要求,从所存贮的最新情报资料中选择用户感兴趣的部分,并将检索结果打印出来,送给用户(参见图1-2)。

面向文献磁带的检索系统只能向用户提供定题服务,所以又称定题服务检索系统,这种检索系统的程序设计及服务方式将在第五章中介绍。现代的一些大型检索系统都是面向磁盘存取的联机检索系统,这种系统虽以提供追溯检索为目的,但也具有定题检索的功能。

追溯检索的英语是Retrospective Search,这种服务方式的特点是,一次检索可以获得数年,甚至十几年内与某一课

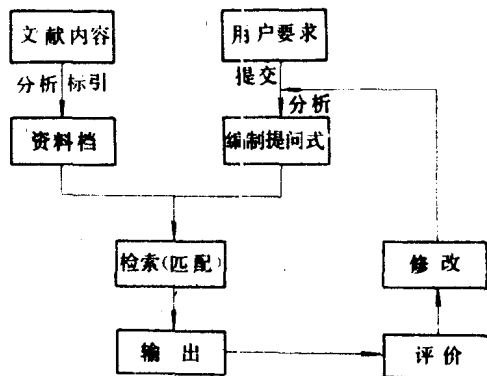


图 1-2 SDI流程图

题有关的情报。现代的一些联机数据往往存贮有十几年以内的文献,很适合于追溯检索。因为追溯检索都采用联机会话方式,所以联机检索和追溯检索两词具有同样的含义。联机检索的有关知识在第七章中介绍,联机检索系统和软件方面的知识在第八章中叙述。

一般说来,无论采用什么样的检索方式,都要求情报检索系统具有较高的查全率和查准率,以便能较好地满足用户的要求。现代的联机检索系统大都具有全文检索功能,几乎所有的文献项都可以作为检索对象,而且,在检索过程中,用户可随时修改检索策略,所以,只要检索策略得当,检索结果是比较理想的。

定题检索都采用批式处理方法,用户不能介入检索过程,不能预知检索结果,一旦取得检索结果,即使发现结果不满意,也难以及时修改检索式。尤其是文献磁带检索系统,很难实现对文摘和标题两字段的检索,所以,定题服务方式的查准率和查全率不一定理想。但是,这种检索系统对计算机的要求不高,即使小型计算机也能胜任,而且程序编制工作不太困难,所需费用也不多,凡具有计算机的单位都可以建立定题检索系统。在我国,由于通讯技术和磁盘容量的限制,联机检索尚未普及,而众多的计算机系统具备开展定题服务的基础和条件,看来,定题服务大有推广普及的必要。还应当指出,这种服务方式能实现主动服务,定期地向用户提供情报,使用户能及时地掌握有关领域、有关专业、或有关课题的最新情报,即使在联机检索网络四通八达的国家,也有不少用户定期地从情报检索中心获得定题检索的结果。

(五) 情报检索网络

要在一个地区、一个国家、或世界范围内实现联机检索,使广大的用户都能在单位内或附近地方检索本地区、本国或国外的数据库,联机检索网络是不可缺少的条件。西方一些发达国家,都建有四通八达的数据通讯网络,一些大型情报检索中心,将他们的计算机与数据通讯网联接起来,使得用户能够通过终端设备与通讯线路检索情报中心的数据库,这就是我们常说的情报检索网络或数据通讯网络。例如,现在在北京可以直检欧洲航天局ESA-IRS,美国DIALOG和ORBIT,英国INFOLINE等四大情报系统的数据库。关于数据通讯网络方面的知识留待第九章介绍。

§ 1-2 情报检索的方法和步骤

情报检索是通过检索工具而得以实现的。检索工具是一个广义的概念,可以从很多角度进行分类,从方式上看,有手工检索和机械检索两大类。机械检索按原理又可分机械式穿孔卡片、光电检索系统、计算机检索系统等。手工检索又有卡片式和书本式之分,从内容看,又有索引、文摘、全文等详简程度的不同。情报检索,无论通过什么样的检索工具,都是为了查找和选取所需的情报,查找过程固然视所用检索工具而不同,但一般说来,可以划分为以下五个步骤:

1. 分析和研究课题,明确欲查情报的要求和范围。
2. 选择检索方法和检索工具。
3. 确定检索途径和检索标识。
4. 利用检索工具查找所需情报(即资料)线索。
5. 获得原文。

下面分别谈谈每个步骤的具体内容。

一、分析研究课题，确定主题范围

用户或情报检索人员在使用检索工具进行查找之前，必须对课题进行详细的分析和研究，弄清课题所需情报资料的范围。对于代替用户进行检索的情报检索人员来说，弄清用户提问的真实含义是非常重要的。如果使用计算机作为检索工具，用户不一定有机会自己上机检索，常常要请计算机部门的检索人员代为检索，检索人员如不了解用户的真正意图和确切要求，检索结果就很难满足用户的要求。久而久之，用户将对检索系统产生不信任感，而减弱他们对情报检索的兴趣和信心。最理想的情况是由熟悉检索工具，了解检索策略和检索方法的用户自己检索，或者由熟悉专业的专职检索人员代为检索，这样，他们就能应付检索过程中出现的问题，灵活而机动地获得满意的检索结果。

为了尽可能获得好的或较好的检索结果，用户或情报检索人员必须学会如何正确地表达自己对检索主题的要求。实践表明，有些人不大善于恰当地表达被检索课题的主题，致使漏检和误检的现象屡见不鲜。在影响检索结果的诸因素中，检索主题要求是否恰当是最重要的因素，对检索效果的好坏具有决定性作用。在确定检索主题时，应注意以下两点：

1. 检索主题概念要具体 检索主题所涉及的概念要具体，不要笼统和抽象。对主题的叙述要简明扼要，抓住重点。

2. 检索主题范围要适当 科技人员一般是针对自己的研究项目或生产任务而提出检索要求的，他们要求获得的情报往往比较专，有些属边缘学科或新学科，所以命中文献常常比较少，甚至没有。碰到这种情况，要适当放宽主题范围。有些发展历史较长，文献量很多的专业，一个课题往往会找出成千上万篇文献，此时，要适当缩小主题范围。

计算机检索系统要求用户以检索表达式的方式送入用户对所需情报的主题要求，如何正确地书写或输入检索表达式，在第二章中介绍，在其他章节中也要涉及。对于广大的用户和情报需求者来说，必须了解、学习和掌握这方面的知识。

我国的一些计算机情报检索部门，常常举办训练班，介绍手工检索和计算机检索的方法，並设有专职检索员，协助或代替用户确定检索主题，编写检索式，填写用户提问单和上机检索，检索结果通过邮寄方法寄给用户。如果采用联机检索方法，用户可自己上机，或者旁观检索人员的操作，检索结果立等可待。

二、选择检索方法和检索工具

无论在国内还是在国外，常用的检索方法有两种，一种是手工检索方法，另一种是计算机检索方法。在西方一些国家，计算机检索比较普及，很多用户乐于使用它，手工检索方法仍然有人使用，但不是主要的。在我国，计算机检索尚未普及，手工检索仍然处于主导地位，即使将来计算机检索普及了，手工检索仍会在某些方面、某些部门、某种情况下发挥它应有的作用。

我们认为，在目前情况下，我国的广大科技工作者应尽可能通过设立在北京、上海、广州、福州、长沙、重庆、西安、成都和沈阳的60余台终端设备，以联机会话方式去检索自己所需的情报资料和科技信息。通过国际通讯网络可以检索欧美地区所建的350多个不同种类的数据库（这些数据库的中英文名称，收藏文献年代及检索系统名称请参阅第八章附录I），这350库的总文献量不少于1亿5千万条，各行各业的科技工作者都可能从其中获得他们所需的情报。

我国已有一些科技情报部门建立了计算机情报检索系统，利用引进的文献磁带（参见第

五章)为各行各业的用户提供定题检索服务。北京文献服务处和机械工业部科学技术情报研究所都已各自建立了可供联机检索,面向全国用户的多种文献数据库,例如北京文献服务处建有美国政府报告(GRA)数据库(拥有100万篇文献记录),机械部情报所建有科学文摘(INSPEC)数据库(200万篇文献)、工程索引(COMPENDEX)数据库(100万篇左右)、机械工程(ISMEC)数据库(约17万篇文献)和金属文摘(METADEx)数据库(约50万篇),科技工作者都可以通过信函方式要求这两个单位提供追溯检索服务。此外,还有一些单位正在建立联机检索数据库,积极创造向广大用户提供联机检索服务的条件。

为了提高文献的查全率,除了使用计算机以外,最好辅以手工检索。手工检索工具是一种传统的检索工具,主要有文摘、简介、题录、索引和卡片,由这些检索工具构成的检索系统称为手工检索系统。目前,世界上手工检索工具的种类和数量愈来愈多,许多读者不知道有些什么检索工具,也不知道如何使用它们。我国有些情报图书单位出版了很多介绍检索工具使用方法的参考书,读者可从中获得手工检索知识,为有效地查找情报创造条件。

按出版形式分,手工检索工具有期刊、单卷、附录和卡片四种;按收录范围分,有综合性的、专业性的和专题性的三种类型;按著录形式分,有题录、简介和文摘三种形式。利用这些检索工具查找文献之前,首先要熟悉各种检索工具中反映文献主题内容所采用的不同标识。标识主要有两类:分类和标题,它们是查找文献的主要依据。不少检索工具收录的文献线索,都按一定分类系统编排,使同类内容的文献处于同一类名之下,根据类目即可查到有关文献线索。有的检索工具收录的文献线索,用号码作为分类标识,按照号码顺序编排,要查找有关文献线索,先要了解所需文献的类号,然后依号查找。如果文献是按标题的字顺排列的,查找这样的检索工具就像查字典一样。

利用手工检索工具也有两种方式:第一,读者在情报人员的指导和协助下,利用卡片目录或书本式文摘索引等查找文献;第二,委托情报人员代查文献。手工检索几乎不需要特殊设备,查找方法简单、灵活、检索费用比较低,但是,大量翻阅检索工具十分麻烦,很费时间。目前,手工检索方式仍然是我国最基本的检索方式。

为了迅速而准确地查到所需的资料,除了要熟悉检索工具以外,还要掌握一定的检索方法。检索方法一般有以下三种:

1. 追溯法 此法的主要特点是以文献后面所附的参考文献为线索,逐一追踪查找,一步一步地扩大可查找的范围。这种方法适用于手工检索工具不全或没有检索工具的情况下,其缺点是,检索效率不高,漏检率和误检率较高。科学引文索引就是在这样方法基础上发展起来的一种新式检索工具。

2. 搜索法 搜索法又分为顺查法和倒查法两种。顺查法是由远及近的查找方法,根据检索工具,从课题发生的年代起,由远而近的一直查下来。查找时,剔除与课题关系不大的线索及文献,找出所需的文献。此法查全率较高,但查找起来颇费时日。

倒查法与顺查法相反,是一种由近及远的查找方法。先查近期资料有其优点,因为近期资料不仅反映了科技发展的最新成果和最新水平,而且一般都引用早期有价值的文献,从而可从中了解到该课题的早期发展状况。

3. 分段法 分段法又称循环法,是前述两种方法的综合,其特点是,先利用检索工具查得一批文献资料,再利用这些文献资料所附的参考文献追溯查找,两种方法分期分段地穿插进行,循环下去,直到找到满意的文献为止。这种检索方法可以克服检索工具不全的困难,

保证连续地查得所需年限内的文献资料。

选择检索工具也很重要。首先要选择质量较好、使用方便的检索工具。对检索工具的一般要求是,收录专业面广,分类方式齐全,文献数量多,报导速度快,文摘简明而准确,编排清晰而醒目,并附有多种索引。有条件选择时,可先用综合性检索工具,然后再辅以专业性检索工具。

三、确定检索途径和检索标识

任何情报资料都具有外观和内容两方面的特征。外观特征是指文献上标明的项目,如资料名称(篇名,刊名,书名,题名等),作者姓名,文献序号(如科技报告编号,技术标准号、专利号等),文种,发表年月,出版地点等;内容特征指文献所研究的内容属于什么学科分类,所探讨的对象属于什么范畴,文献中涉及哪些主题词等等。检索工具虽然种类繁多,但不外是按照文献的外表特征或内容特征,安排相应的检索途径。这些途径有:

1. 书名途径 根据书刊资料名称来查找的检索工具有“图书书名目录”,“期刊刊名目录”,“科技资料索引”等。这类目录索引一般按书刊资料名称字顺编排与查找。由于科技资料的名称一般较长,难于记忆,加之按名称字顺编排易造成同类文献分散,不能满足族性检索要求等问题,所以现在一般检索刊物都不采用这种途径。但是在计算机情报检索系统中,上述问题得到了较好的解决,情报检索系统提供书名检索途径。

2. 著者途径 根据著者姓名来查找是常用的途径之一。著者包括个人著者和团体著者(机构、公司名称等)两种,使用的检索工具有“著者索引”、“机构索引”等。这类索引按著者姓名的字顺编排和查找,由于编辑简单,国外许多检索工具都有这种索引。国内外从事科学技术研究工作的个人或单位,都各有所长,同一著者发表的文章,专业范围大多相同。于是,在同一著者姓名下,往往集中了内容相近或有着内在联系的文献,在一定程度上能够适应族性检索的要求,但不能作为系统查找文献资料的主要途径。著者检索途径,国外非常普遍,我国也开始重视这种索引。各国由于风俗习惯和语言文字的差异,对姓名的写法各式各样。欧美各国书写姓名一般是名在前,姓在后,我国正好相反。利用著者途径检索时,要注意著者姓名的排列规定。设计情报检索系统时,既要尽可能照顾风俗习惯,又要规范化和标准化。

3. 序号途径 有些检索工具是按文献资料的专用号码的大小顺序排列的,这类检索工具有“报告号索引”,“合同号索引”,“入藏号索引”,“专利号索引”,“专利对照号索引”,“登记号索引”等等。这类索引编制简易,便于查找,但必须了解有关文献资料的编号规律,因而这类索引的利用率受到很大限制,这一途径不能作为主要的检索途径。

4. 分类途径 有些检索工具是按某种分类系统而编排的,属于此类的检索工具有“分类目录”,“分类索引”等。分类检索系统比较能够体现学科系统性、反映事物的派生、隶属与平行的关系,便于从学科门类的角度来查找,能够较好地满足族性检索的要求。但是,分类表往往落后于科学技术的发展,往往不能正确处理新兴学科和边缘学科,造成查找不便。再者,按分类途径检索必须了解学科分门别类的体系结构,在将概念变换为分类号的过程中,常易发生差错,造成漏检或误检,影响检索效果。但是分类检索现在仍是一种重要的检索途径。

5. 主题途径 主题途径是以代表文献资料内容实质的主题词作为查找文献的一种途径。主题词(包括叙词,关键词等)按字顺编排,查找方便。检索时先确定主题词,再按主

题词的字顺查找,即可获得某一主题的文献。采用这种途径能较好地满足特性检索要求,所以,它适用于查找范围具体,内容专深的课题资料。国外检索工具大部分编有主题索引,主题途径是一种比较方便,检索效率较高的检索途径。这一检索途径具有很多优点,如用文字作标题,表达概念准确而灵活,不论主题专指程度有多高,都可用相应的主题词标引出来,而不受类目设置和级位的限制;可随时容纳新的主题词,新兴学科和边缘学科的文献资料的标引和查找也不困难;能将分散在各个学科领域中的资料集中在一个主题词下,便于进行综合研究。

我们认为,在实际检索工作中,可以主题检索途径为主,以分类、著者、序号、书名等途径为辅,取长补短,相互配合,以便取得较好的检索效果。

选定检索途径并准确确定检索标识(如主题词,分类号等)之后即可正式开始检索。按主题词法编排的检索工具,一般都提供主题词表,供准确地选择主题词之用。如美国《工程索引》的“工程主题词表”,《化学文摘》的“索引指南”,都是指导检索者选择主题词的辅助工具。按分类法编排的检索工具,在文摘正文的前面或后面附有“目次表”,它实际上就是一个包括着大类和小类的类目表,可供分类法检索时参考。

四、利用检索工具查找所需情报线索

在查找文献资料时,要尽量利用各种有利的条件,从最简捷的途径入手,例如,主题内容具体的问题可以从主题途径查寻,主题范围广泛的问题可以从分类途径查找,若已知文献作者,可以从作者途径查寻;若已知专利号,可以根据专利号索引查找,此外,还应尽可能利用累积索引。

着手查找之前,最好把准备用作检索标识的主题词、分类号、著者姓名等列成一个表,依次查找,并在查找过程中,随时加以修正补充。当查找效果不够理想时,应考虑所用的检索标识是否正确,发现问题后及时进行调整。

五、查找原文,获取情报

根据检索工具获得文献线索后,还必须根据所获线索去查找原文。原文到什么地方去查找呢,这是用户很感兴趣的一个问题。有些读者或用户,通过计算机或手工检索工具找到了一些或一大堆文献线索后,发现文摘或索引并不能满足他们的要求,因而要求获得原文,他们常因找不到原文而深感苦恼,有的甚至对查找文献失去了兴趣。这里反映了两个问题,一是收藏原文的图书情报单位未能提供优良的原文检索工具,如馆藏目录或联合目录,二是读者手头缺少这样的工具。当然,有些资料国内未引进,读者是找不到的。要解决原文难找的问题,图书情报部门是要作出很大努力的。

国外一些大型情报检索系统,向用户提供原文的复制品,较好地解决了这个问题。我国的一些计算机情报检索系统,也应仿效国外的做法,着手解决原文提供问题,否则,用户的数量和检索效果将会受到不良影响。

§ 1-3 情报检索语言

一、情报检索语言的概念和作用

情报检索语言是根据情报检索的需要而创制的一种人工语言,简称检索语言,又称标引语言或情报语言。检索语言是情报检索系统存贮与检索所使用的共同语言。存贮和检索是检

索系统对立矛盾的双方,把这一对矛盾沟通和统一起来,在存贮人员与检索人员之间起到桥梁作用的工具,就是检索语言。

具体地说,存贮文献时,文献的内外特征必须按照一定的语言加以描述。检索文献时,情报的提问也必须按照一定的语言加以表述。两者之间必须采用共同的语言,否则,提问与存贮无法比较,两者是否匹配(相符)也无从谈起。所以,检索语言实质上是标引与检索之间的约定语言,它的作用如下:

- (1) 保证不同标引人员表征文献一致;
- (2) 保证提问语言和表征文献的语言一致;
- (3) 保证检索者按不同需要检索文献时都能获得最高的查全率和查准率。

情报检索语言作为情报检索的人为语言,它能够简单明白而又专指地表达文献及检索课题的主题概念,排除自然语言存在的多词一义、一词多义、词义含糊、词语与概念非一一对应等现象,从而减少检索与标引之间的不一致性或误差,换句话说,提高情报检索的查全率和查准率。

情报检索语言的主要功能大致可归纳如下:

- (1) 对文献的情报内容(以及外部特征)加以标引;
- (2) 对内容相同及相关的情报加以集中或揭示其相关性;
- (3) 对大量情报加以系统化或组织化;
- (4) 便于将标引用语和检索用语进行一致性比较。

情报检索语言由词汇和语法组成。词汇是指登录在类表、词表中的全部标识。一个标识(分类号,检索词,代码)就是它的一个词语,而分类表,词表则是它的词典。语法是指如何运用标识(单个标识,或几个标识的组合)来正确表达文献内容和情报需要,以有效地实现情报检索的一整套规则。

二、对情报检索语言的要求

情报检索语言的质量及其使用正确与否,对情报检索效率有着重大影响。下面谈谈对情报检索语言的若干基本要求:

1. 保证较高的查全率和查准率 或者换个说法,漏检率和误检率要能控制在允许的范围内。

严格地说,收录的文献用某种情报检索语言标引并归纳入检索工具或检索系统之后,在需要时应能将其中含有用户指定情报的文献全部检出。事实上,很难达到全部检出的要求,通常只能达到某个百分比。

查全率主要与一种情报检索语言是否能比较全面地显示概念之间的等级关系和相关关系、以及在它的词汇之间是否能排除同义现象的功能有关。

查准率是指检出的全部文献与真正满足需要的文献的比例。事实上,在情报检索中,检出的文献也很难全部都与某项情报需求有关。所以查准率也只能达到某个百分比,换句话说,一定的误检率是应当允许的。查准率主要与某种检索语言的标识(单个标识或若干标识的组合)是否能达到较高的专指度,在它的词汇之间是否能排除多义现象和同形异义现象、以及它是否能正确地显示概念之间的关系等功能有关。

2. 满足多种检索要求 社会上对情报需要是多种多样的,而且是动态变化的,即使是同一个人也会提出不同类型的检索要求。典型的检索要求大至有下列几种: