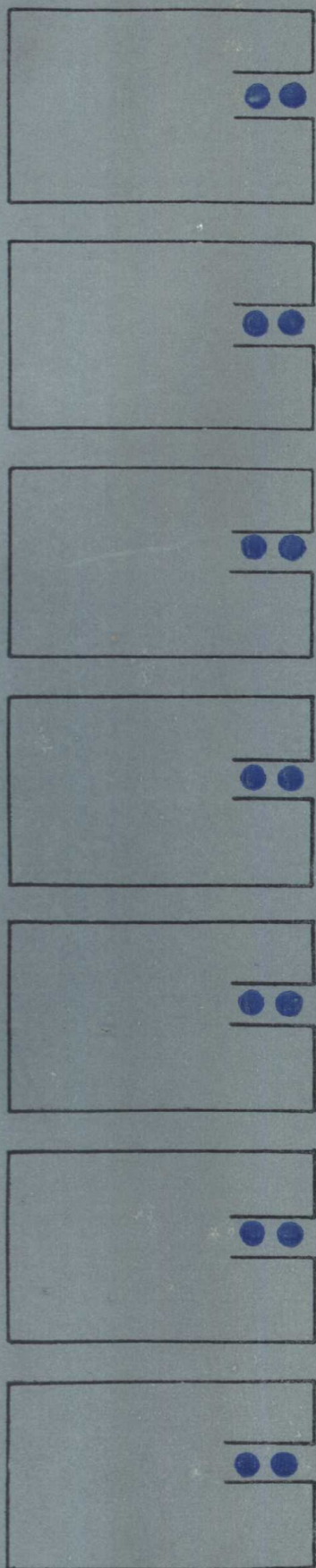


JISUAN JI
QING BAO JIAN SUO

袁鹤龄 编著

计算机
情报检索

湖南大学出版社





计算机情报检索

袁鹤龄 编著



湖南大学出版社

781203

内 容 简 介

本书概要地介绍了情报检索的基本概念、工作内容和发 展概况,系统地介绍了计算机情报检索系统的环境要求、检索方法、服务方式、存储和检索的数据结构以及系统设计技术等,并列举了各类计算机情报检索系统的应用实例。对联机检索也进行了详细的介绍。

本书可供计算机应用、图书情报、档案等专业作为教材或教学参考书,也可供有关图书馆工作人员、科技情报人员、大专院校师生和广大工程技术人员阅读。

计 算 机 情 报 检 索

袁鹤龄 编著



湖南大学出版社出版发行

(长沙市岳麓山)

湖南省新华书店经销 湖南大学印刷厂印装



787×1092毫米 16开本 15.375印张 355千字

1987年4月第一版 1987年4月第一次印刷

印数: 0001—5000册

ISBN 7—314—00113—8/Z·8

统一书号: 17412·7 定价: 2.60元

前 言

信息社会是大量生产知识的社会,各种文献资料浩如烟海。在日益信息化的进程中,如何在浩瀚无际的书山文海中有效地查找所需要的文献信息,便成为急待解决的问题。计算机情报检索作为一门新兴的交叉学科由此应运而生,它借助于计算机科学、情报科学等相关学科的理论和方法迅速地发展起来,现已成为情报工作现代化的重要手段。鉴于目前国内有关计算机情报检索方面的教材和教学参考书很少,作者根据1983年通过的高等学校本科计算机及应用专业《计算机情报检索》教学大纲,在经多次讲授的《计算机情报检索》讲义的基础上,整理改编成本书,以满足读者的需要。

全书共分七章。第一章是情报检索导论,介绍情报检索的基本概念,工作内容和发展概况,使读者对计算机情报检索有一个总的轮廓性的了解。第二章是联机情报检索。是为适应国际联机检索在我国迅速开展的需要,专门一章介绍了联机检索的有关内容。特别是对著名的 DLIALOG 国际联机检索系统的检索指令、检索方法和检索实例进行了介绍。第三章介绍情报检索系统对计算机系统的资源要求。第四章介绍计算机情报检索系统。第五章列举了用于情报存储与检索的数据结构。第六章用了较大的篇幅介绍计算机情报检索系统的设计。第七章是应用实例,主要介绍我国自建的几个中小型情报检索系统,特别是汉字情报检索系统;也介绍了几个著名的国际联机检索系统。书后附录包括了三大国际联机情报检索系统的主要文档等内容。

本书在取材上,既重视传统内容的介绍,也注意最新成果的吸取。力争从不同侧面反映计算机情报检索领域内的最新发展,特别是我国情报检索方面的发展。例如,在顺排成批检索中,既介绍了日本的菊池敏典方法,也介绍了我国学者对菊池算法的改进;在联机检索中,既介绍国际联机情报检索的过程与方法,也介绍我国研制的联机检索的预处理软件等。

本书可供高等学校计算机及应用、图书情报、档案管理等专业学生作为教材或教学参考书,也可供图书馆工作人员、科技情报人员、大专院校学生和广大工程技术人员阅读。

在本书编写过程中,湖南国际经济学院付院长傅予行副教授、湖南大学杨润生、邱光谊教授、何子凡副教授给予了热情鼓励与支持。国防科技大学孟庆余教授对全书进行了认真的审阅。郑章飞同志协助收集了有关资料,在此一并表示衷心的感谢。

由于水平有限,错误和不当之处,恳请批评指正。

编著者

1985年12月

《计算机情报检索》目录

第一章 情报检索导论

§1.1 什么是情报检索.....	(1)
§1.2 文献资料档.....	(4)
§1.3 提问逻辑与表示.....	(11)
§1.4 情报的检索.....	(16)
§1.5 情报检索效率的评价.....	(18)
§1.6 计算机情报检索的发展.....	(21)

第二章 联机情报检索

§2.1 联机情报检索.....	(26)
§2.2 联机检索与计算机网络.....	(27)
§2.3 DIALOG 国际联机检索系统及其检索指令.....	(31)
§2.4 联机检索的一些方法.....	(40)
§2.5 DIALOG 系统国际联机检索实例.....	(49)

第三章 情报检索与计算机系统

§3.1 情报检索对计算机系统硬件的要求.....	(57)
§3.2 情报检索对计算机软件的要求.....	(60)

第四章 计算机情报检索系统

§4.1 计算机情报检索系统概述.....	(66)
§4.2 情报检索系统的检索方法.....	(67)
§4.3 情报检索系统的服务方式.....	(69)

第五章 用于存储与检索的数据结构

§5.1 预备知识：数据的基本组织形式.....	(78)
§5.2 概述.....	(79)
§5.3 排序树结构.....	(81)
§5.4 采用字符树的词典存储.....	(86)
§5.5 某些排序算法.....	(87)
§5.6 倒排文件的结构.....	(90)
§5.7 散列存储.....	(93)
§5.8 堆栈、队列及其它存储结构.....	(96)

第六章 情报检索系统的设计

§6.1 情报检索系统设计的一般步骤.....	(97)
§6.2 词表与标引.....	(108)
§6.3 资料档的组织与设计.....	(119)
§6.4 情报检索程序的设计举例.....	(124)

第七章 计算机情报检索系统应用实例

§7.1 SJTU 科技情报检索系统	(159)
§7.2 DJS—100 系列小型计算机上的汉字情报检索系统	(164)
§7.3 ZKQ/A 汉字情报自动处理系统.....	(172)
§7.4 微型机汉字情报检索系统.....	(177)
§7.5 MIDBMS—HN 医学情报编目、检索系统.....	(188)
§7.6 微机局部网络上的中文情报检索系统.....	(200)
§7.7 著名国际联机情报检索系统简介.....	(204)
§7.8 计算机电话查号系统.....	(209)
§7.9 刑事情报计算机汉字检索系统.....	(212)

附录 1 计算机情报检索服务登记表及其实例.....	(220)
附录 2 三大国际联机情报检索系统数据库介绍.....	(223)
附录 3 中华人民共和国国家标准 GB 2901—82 《文献目录信息交换用磁带格式》	(238)
参考文献	(238)

第一章 情报检索导论

§ 1.1 什么是情报检索

1.1.1 情报与情报检索

情报一词来源于拉丁文“information”，是消息传递的意思。目前，国内外情报界对情报一词有多种定义。概括起来，主要有两个方面。一种是以知识作为同义词叙述，认为情报是传播中的知识，是“意志、决策、部署、规划、行动所需要的知识和智慧”。另一种是以信息作为同义词叙述，认为“情报是指含有最新知识的信息。”这里所说的知识，是指人们在社会实践中对客观事物的认识和积累起来的经验，而信息则是表征事物性质、关系等知识、概念的符号集合体。从内含与外延观点来考虑，情报、知识、信息都处于特定的社会环境之中，有共同点，但又具有不同的范畴。可以说，信息包含了知识和情报，知识包含了情报。

人们所需要的情报，存在于不同的形式之中。有印刷型的书写情报，这是大量的；有记录在录音带、电影或电视录象带上的声象情报；有存贮在计算机磁带或磁盘上的机器可读情报等等。那么，究竟什么是情报？我们首先来分析情报的基本属性。情报具有知识性、传递性和效用性等基本属性：

(1) 知识性：情报是以知识为实体的，没有一定的知识内容就不能形成情报。知识性是情报这一事物的核心，是情报的基本属性之一。

(2) 传递性：情报具有动态的特征。知识要成为情报，必须经过传递。知识的传递是以信息为媒介进行的，情报以信息为载体。所以情报是知识的传递。传递性也是情报的基本属性。

(3) 效用性：情报具有客观的效用。人们创造情报的目的，在于传递、交流和利用。凡情报均有效用，对生产、科研、教育等客观实践具有继承、借鉴和参考价值。效用性也是情报的基本属性。

根据以上分析，情报一词可以定义如下：情报是针对特定对象的需求，通过信息负载与传递具有继承、借鉴与参考作用的新知识。

情报科学就是研究以知识为实体、以信息为载体的情报的来源、传递和利用的理论方法的科学。简称情报学。根据情报科学的内容与实质，情报学已发展成为理论情报学和应用情报学两大类。从情报学的属性看，它具有综合性与边缘性。构成情报科学的理论基础主要是图书馆学、技术科学、管理与工程科学和自动化情报检索。

计算机情报检索，是一门新兴的交叉学科。它是借助于计算机科学、情报科学等相关学科的理论和方法而逐步发展起来的。

检索就是查找的意思。关于情报检索的定义，从不同的角度出发，有着不同的定义。本课程给出定义如下：

所谓情报检索，通常是指情报的搜集和存储，并对已存储的情报进行检索与分发。

因此情报检索的全过程大致应包括以下几个方面的内容：

- (1) 情报的搜集和选择；
- (2) 情报的分析和加工；
- (3) 情报的存储；
- (4) 情报的检索；
- (5) 情报的提供和评价。

概括地说，情报检索主要包括情报的存储和情报的检索两部分的内容。

1.1.2 情报检索的分类

情报检索一词，使用很广泛，可以从不同角度予以分类：

一、情报检索根据实现的手段来分，可以分为传统的手工检索和计算机化的机器检索。本课题主要讨论计算机化的情报检索。

二、情报检索根据检索者、机器之间的关系以及计算机处理方式来分，有脱机处理和联机检索两种。

脱机处理方式是以磁带系统为基础，几乎全部情报都存储在磁带上，利用计算机的方式基本上是成批处理。成批处理是把用户的查找要求积累到一定时间后，统一上机处理，即用户在提出查找要求的当时，一般还不能马上得到检索结果。联机检索是以磁盘系统为基础的。是指用户通过终端和通信线路，从情报检索系统存储的文献库中检索自己所需的情报。联机检索时，用户可以随时对计算机提出查找要求，并通过终端马上得到检索结果。用户在检索过程中，根据得到的检索结果，可以随时修正自己的查找要求，以求得更好的检索结果。

三、情报检索根据检索词控制方式来分，有固定词表控制式检索和自由选词式检索。固定词表式情报检索中，标引和查找都要使用词表中指定的词，而在自由选词式情报检索中，可以采用情报中出现的原文词进行标引和查找。一般，当词表还未编成时，情报检索可采用自由选词式，待计算机积累出一部份词汇，并编成检索词表后，再采用固定词表式。当然，最好是一次编成词表，并用词表标引和查找。

四、情报检索根据检索文献库的年代来分，又可分为定题情报服务 (SDI—Selection Dissemination of Information) 与追溯检索 (RS—Retrospective Search) 两种。SDI 是对现行最新文献作有选择地专题情报传送。而 RS 是对所有的过期文献库进行检索。关于 SDI 与 RS 将在第四章详细讨论。

五、情报检索根据情报检索系统处理的情报内容来分，它又包括文献检索、数据检索和事实检索。

1. 文献检索：需要查找的对象是有关文献资料 (Document) 的情报。凡是检索某一主题、某一作者、某一地区、某一时期、某一产品等有关情报资料，以及查找某一论文的出处或某一刊物的收藏处等，都属于文献检索的范畴。

2. 数据检索：检索的结果是数据。检索的资料档是数字型数据库。文献数据库的数据主要以文本形式表示，而数字型数据库的数据则是存储那些通过观测或试验得到的

数值。这些数值通常以集合的形式予以存储，这些数值集合本身是不能检索的，但可将它们与可检索的关键词或叙词相联接。例如，一个数据库包含许多气象数据构成的一些表，每张表都与一个特定的城市和日期联系在一起。需要检索时，可以用特定的城市和日期进行检索，然后打印出与特定城市和日期的有关气象资料的表。在这种情况下，城市名称、日期都可以作为可检索的数据，而这些气象资料表则可看作是不可检索的数据。但它们与可检索数据的特定值有关。数据检索和文献检索的区别在于数据检索的结果可使用户直接使用，如数值、图表、公式或化学分子式等等。

3. 事实检索：凡是有关某一事物（事件、事实）发生的时间、地点和过程（情况）的查询属于事实检索的范畴。在这种检索系统中存储的情报是数值情报与非数值情报的混合。一般要从存储的情报中检索出必要的情报，再加以逻辑推理才能给出结论。例如，管理信息系统（MIS）的数据库中包含有企业管理所需要的资料，这些情报主要是用来作出政策决定。因此，包括的情报有人员情报、工资状况、办公信件、销售统计与预测、产品规格、库存数据和经济数据等信息。

与文献数据库相比，它所存储的情报在性质上有更多的变化，但由于这些数据的组织形式与所要求的目的直接相关，所以更容易达到检索目的。这些数据常常要比文献标题或文摘以更直接的形式存储。

本课程的计算机情报检索将重点讨论文献检索，事实上是计算机化的文献检索。这是因为文献是情报来源的非常重要的一部分，在各种情报检索系统中，文献检索最为重要，一般也比较复杂。因而研究得也最多。在国际上也有很普遍的意义。因此情报检索专家A·肯特说：“所谓情报检索是指机械化的情报检索，它与使用机器的文献检索几乎是同义的。”

然而，情报用户和情报需求也在发生变化。自1975年以来，非文献数据库检索数量和类型的发展，悄悄地赶上了文献数据库，并将进一步地增长。

1.1.3 情报检索系统的工作流程

计算机情报检索大致分为情报的存储与检索两个部分。即文献资料档的建立与查找两部分。或简称为文献的加工与检索。文献加工是将各种文献按一定的规则进行加工，将加工后得到的文献数据，通过输入设备送入计算机的内存储器，经过计算机处理后，将文献数据写入外存储器（一般为磁带或磁盘）。当进行检索时，则将需要的文献数据由外存调入内存进行查找，最后将检索出来的文献，通过输出设备输出。另外，计算机还可以把文献资料档的内容编制成索引刊物，然后印刷出版。

计算机情报检索系统的工作流程，可用图1.1简单地表示之。从图1.1可知，整个系统是以资料档为中心，作为资料档的来源，

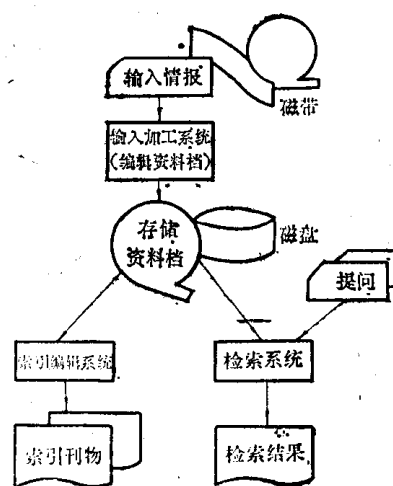


图1.1 情报检索系统工作流程

需要有情报输入加工部分；一旦形成之后，该资料档既可以作检索的依据，又可作编制索引之用。

§ 1·2 文献资料档

计算机情报检索技术主要包括情报的存储和检索两方面。情报的存储基本上是靠建立情报检索系统的单位自己来组织和建立文献资料档。文献资料档（file）又叫做文献数据库。如何合理地编制文献资料档是有效进行检索的关键。它包括 § 1·1 中情报检索流程中三项工作内容：即“情报的搜集和选择”、“情报的分析和加工”以及“情报的存储”。前两项内容，实质上是为编制文献资料做数据准备，在整个流程中，其自动化水平最低，主要是用手工方式，涉及计算机技术较少。但它们是两个极为重要的部分，是决定情报存储与检索内容的阶段，是对情报检索效果起关键作用的阶段。

1·2·1 情报的搜集

一、确定专业范围

建立一个包罗万象的文献资料档是不可能的，也不必要。一个文献资料档必须有一个专业范围。

编制文献资料档的第一步就是确定范围。情报的搜集首先要从总的方面确定所搜集情报的专业范围。选定专业范围可考虑以下四个因素：

- （1）文献资料档编制机构的任务；
- （2）用户的需求；
- （3）文献源的可获得性；
- （4）文献资料档的特色。同其他文献资料档相比，要有自己的特点，不互相重复。

选定的范围，不宜太宽，也不能太窄，应确定一个最佳范围。

二、确定情报来源

由于情报的产生涉及各个学科领域，其来源是很广泛的，包括自然科学和社会科学的各个方面。情报源主要的有期刊、会议录、科技报告、会议录、政府出版物、学位论文、技术标准、专利说明书、产品样本和科技图书等类。

情报源的确定要突出一个“全”字，应尽量包括各种出版物类型，各种载体和各文种的文献。情报源搜集得全，才能做到查全。但是，在实际搜集中，由于条件的种种限制，也由于计算机字符集的限制，只能做到相对的全。为此，必须清楚地划定收集哪些类型的资料和因条件限制不能收集哪些类型的资料，做到前后一致。这一点也要通告用户，使用户查过这个文献资料档后，就确切知道在其专业范围内已查全了哪些资料，还有哪些资料该资料档没有收集，需要进一步查找。

三、确定情报的选择

确定了情报源，甚至具体选定了收录的期刊目录，也并没有完全解决问题，还有一个情报的选择问题。

以期刊为例,有些是本专业的核心期刊,有些是相关学科的期刊,有些可能只有极少与本专业有关的文献,这些期刊都列入了收录期刊目录。因此,可以大体规定,哪些期刊全部文献收入,哪些期刊要选择部分文献收入,哪些期刊只选极个别文献收入该文献资料档。这是一项专业性很强的工作,可结合情报的分析加工同时进行。因此,在编制资料档前,要进行范围很广泛的文献筛选工作。范围广才能达到全,正确筛选才能达到准。它密切关系到文献资料档的质量。

四、情报搜集的方法与途径

情报的搜集有多种方法与途径,现列举如下:

(1) 采购(或称订购)

采购是情报部门搜集国内外情报资料经常使用的一种方法,其形式分为现购、邮购和预购三种。

(2) 交换

交换是搜集情报的有效方法之一,分国内交换和国外交换两种。国内交换关系的建立,可通过资料联系卡、派专人访问和会议协商等方式进行;国外交换关系的建立,应当按照有关部门的规定统一进行。通过交换方式取得的资料,一般针对性强。有些比较适用的科技文献资料,如非卖品的刊物、限制发行或内部使用的资料,有时不能通过订购方式获得的,但有时采用交换的方式可以获得。与国外交流的资料由于不经过书商,到达的时间比订购的一般要快一些,这对于情报的迅速传递和使用是很有利的。

(3) 索取

索取是根据报刊杂志和其他途径了解到资料的线索来源,然后通过有关方法取得所需资料。

索取的方法比较简单,不拘形式。交流会、展览会、座谈会、参观访问都是索取的场合和机会。

(4) 现场搜集

现场搜集是通过现场活动进行情报搜集的方法。

(5) 委托搜集

有些科技情报和难得的文献资料,通过以上几种方式都无法获得时,可采取委托有关部门和人员进行搜集。

(6) 情报网搜集

情报网是一种情报交流的有效工具,也是取得情报资料的重要途径。我国在全国范围内普遍建立了各种情报交流网,从中央到地方的各级情报机构,组成了国内统一的科技情报网。来自任何方面的研究成果、发明创造、技术革新和先进经验,都可通过情报网的组织,集中到达本地区、本专业的情报中心或中央科技情报部门。各情报单位根据本部门的科研和生产建设的需要,都可进行对口搜集。可充分利用这种条件进行情报搜集。

情报搜集是一项非常细致而十分严肃的工作,是情报科学的一门重要的分支学科。它要求搜集人员不仅熟悉各种途径,掌握各种方法,会用各种检索工具,而且必须深入到科研生产实践中去,了解情报用户的需求,做到有计划有针对性地搜集。

1.2.2 情报的整理

一、有关术语介绍

为了叙述的方便,下面先集中介绍几个有关情报的术语:

一次情报——由情报发生者生成的情报本身叫做一次情报或原情报。如作者依据自己的科研成果而形成的文献,诸如论文、研究报告、专利说明书、会议文献等,不论是手稿,还是印刷出版物,均称为一次情报或一次文献。一次情报有文献、资料、图画、照片等等。

二次情报——对一次情报进行加工整理而以某种形式从某些方面介绍原情报的信息叫做二次情报。象以文摘、索引、题录等形式出现的文献都是二次情报或二次文献。

著录事项——作为原情报的附属情报,被用作索引的对象,著录事项有标题项、著者项(作者姓名及著作方式等)、出版项(出版地、出版者、出版年、版次)、稽核项(页数、图表、开本、装订、定价)、附注项、提要项(内容简介)以及索取号等。

主题词——用来揭示文献内容表达概念的词,称为主题词。主题词是情报检索和图书馆学的术语。关键词是计算技术的术语。在计算机情报检索中,主题词是经过规范化的表示文献内容的关键词。为了把主题规范化,由国家统一制定各学科的“主题词典”或叫主题词表)。关键词的范围比主题词广。主题词是计算机情报检索的重要词。也是表达文献内容的重要词。在计算机情报检索系统中,主题词典供图书情报部门标引各种科技文献图书的主题词,也是提问者在编制提问单时的依据。

文摘——是从原文献中抽出主题内容,并以简练的语句写成摘要,使读者阅后能了解原一次文献的主要内容。文摘主要分为两种:一种是指示性文摘,它是以简短的内容摘要,指出原文献的主题内容,供读者判断文献是否需要。另一种是报导性文摘,这是一种不需要阅读原文献便可知道原文献要点的比较详细的内容摘要,它不仅说明了原文献的主题内容,研究对象与目的,所用方法和技术手段,而且写明了主要数据、结果和结论等内容。

标引——是把文献的主题内容和某些具有检索意义的特征,用标识符号做出多方面的反映,赋予文献索引项的一种文献加工。它是一种工作量大而技术性强的工作,涉及到专业知识,主题词表和外语水平。

主题分析——是从复杂内容的文献或检索提问中,分析出构成主题的基本概念或要素的工作。它是编写文摘、分类、标引和编写提问检索式的第一步工作。

二、情报的整理(分析与加工)

(一)主题法和分类法:

经过搜集的情报资料,必须按照科学方法加以整理,才便于存储、查找和利用。情报整理的方法主要有分类法和主题法。

所谓分类法是指以学科体系为基础,按学科分类进行编排和检索的一种方法。它能比较集中的体现学科的系统性,反映事物的层次从属关系,便于按学科进行族性检索。它的基本结构是按知识分类的次序,从一般到特殊,逐级展开学科体系。分类法通常用号码表示一个具体的类目,在号码中用园点表示类别的层次结构。分类法历史很久,在

图书资料的编目、组织和检索方面起着很大的作用。《中国图书资料分类法》、《中国图书馆图书分类法》和《中国科学院图书馆图书分类法》是我国建立图书资料，也是文献资料的主要分类依据。

所谓主题法是以语言词汇为基础，将情报资料所论述和研究对象，直接用规范化的名词术语（即主题词）表示出来的一种方法。是一种按主题概念进行编排和检索的方法。它以主题词作为情报内容标识和查找的依据。主题词由于以词汇为基础，在表达主题内容方面比较直观灵活，并且由于主题词之间有一定的语义关系，能较好地满足多元检索的要求。主题法在计算机出现以前就已被应用，但只有在计算机情报检索中主题法才占据了重要位置，并得到了迅速的发展。主题法比较适合于计算机加工处理，使计算机速度快和存储容量大的特点在情报检索中得到充分的发挥。现在全世界的计算机情报检索系统中大部份是以主题法为基础的，以分类法为基础的计算机情报检索系统比较少见。

（二）情报的分析和加工

由图 1·2 可知，情报的分析和加工，可分为加工成二次文献和加工成一次文献。二次文献的加工主要指作文摘，进行标引与分类。加工方式有手工和自动化两种方式。二次文献加工的自动化是很复杂的工作，目前还有许多技术难关没有被攻克，还在深入研究之中。目前很多工作基本上还是用人工来实现的。

我们知道，科技图书文献是非数值情报，它的情报量很大。如果情报检索按其原形存储和检索，就会造成占用存储空间异常庞大、检索时间花费很长、效率极低的恶果。因此，一般文献情报检索不是把情报的原形直接存储。也就是说不是把每篇文献资料的内容都存放到磁带或磁盘上，而是经过分析、加工和整理以后，再存储到计算机里。因此说情报的分析和加工主要指把一次文献加工成二次文献，二次文献主要是文献刊物、书目索引等检索工具。加工后的文献数据，再通过输入设备存入计算机，从而构成文献资料档。

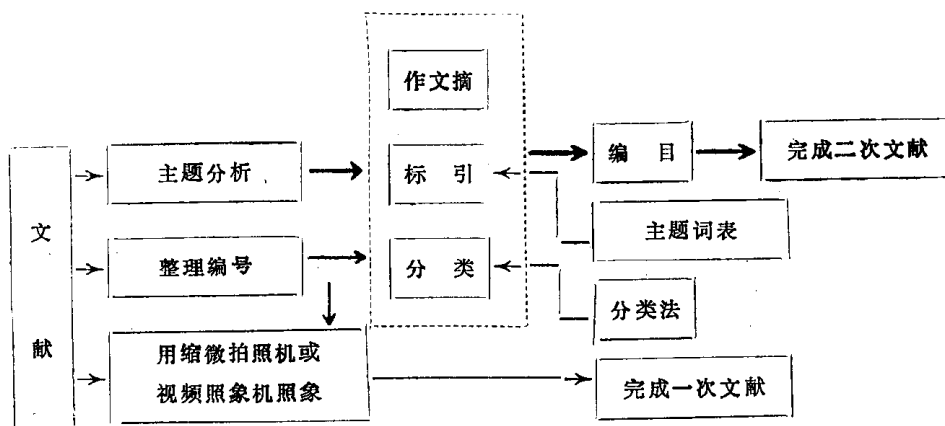


图 1·2 情报的分析和加工流程

形成二次文献的加工处理工作包括下面几个方面：

（1）抽出主题词：即选出文献中的几个主题词。

(2) 进行主题分析, 作出压缩和精练内容的文摘。文摘一般只有几百个字, 计算机有可能存储。而且还可以通过计算机排版印刷出快报, 报导本领域的最新文献。

(3) 文献标引: 按照二次文献需要登录的各种类别分别进行处理。可以登录的类别包括索取号、篇名、作者、发行年月、卷、期、页数、所载刊物名称、主题词、文摘……等等。但具体每篇文献所选择的登录项目只是其中的一部分。前面已经提到过, 上面这些工作叫做标引。

由于情报检索主要应用主题法, 所以标引就成了编制文献资料档的主要加工工作。文献资料档的质量在很大程度上依赖于标引的质量。

(4) 有的文献加工还要进行分类。分类是将文献按学科的分类加以体系化、便于族性检索(或称学科体系检索)。情报检索主要使用主题法。但分类法在计算机检索中同样有用。采用分类法的族性检索, 其好处是: 便于分割文件; 便于按类编制文摘索引; 可以缩短检索时间; 有利于进行文献调查等需要查全的工作, 避免漏检。因此不少文献资料档都设有分类字段。

1.2.3 情报的存储

加工成二次文献后, 一般还要把二次文献按一定规则, 通过计算机输入设备将其输入到计算机内, 记录在磁带或磁盘上, 供人们检索。从加工成二次文献到记录在载体上, 虽然需要很大的工作量, 但情报检索是情报一次输入以后, 可以无限次被享用。

所谓情报的存储, 通常包含了一次文献(情报)和二次文献(情报)的存储两层意思。一次文献的存储, 就是把资料按一定规则存入书架, 只要按索取号、排架号或分类号取出即可。基本上是现行图书馆的传统方法。一次文献除了书刊形式外, 还有缩微胶卷、缩微胶片、录音和录象等载体。但利用电子计算机情报检索一般只能检索到二次情报(文献), 这对于大多数用户来讲已经能够满足需要了, 因为从篇名、作者、出版年月、文摘等已经基本上了解该文献能否满足需要了。如果需要看原文献, 还得通过检出的分类号, 索取号去索取原文献。因此, 计算机情报检索存储的情报一般是指二次情报(文献)。而检索过程则是根据用户的提问要求, 从二次情报(文献)的资料档中检索出符合提问的文献。

1.2.4 文献资料档

一、文献资料档的编制

1. 文献资料档记录的内容:

文献资料档的每一个记录一般要包括每篇文献的主要特征, 即包括如下内容的“文献记献”:

- ① 文献号;
- ② 著录事项(包括篇名、作者、类别、出版日期等);
- ③ 主题词;
- ④ 文摘等。

2. 文献资料档的编制:

文献资料档的编制，同二次文献是一致的。即确定了著录项目的规则，经过对文献的选择，分析和加工，著录出一篇文献的条目（目录数据），然后通过计算机的输入和处理，将文献记录一条条的记载在磁带等载体上。这些记录的集合，就构成文献资料档。因此，在某种意义上说，文献资料档就是二次文献的磁带版。

一般在把目录数据输入计算机前，要将分析加工著录好了的目录数据，填在一张印好的统一使用的工作单上，填好后经过核对就可送到数据录入部门将其输入计算机。输入后经反复核对无误后才能正式编入文献资料档。

编制文献资料档的单位主要是文摘社。由原来手工编文摘而改为计算机编制文摘刊物，同时作为产品生产文献资料档。其它编制单位有国家机关、学会、协会、高等院校和研究机关等。

在计算机应用比较先进的国家，文献资料档的编制已变成一种专门的工业性行业。例如法国文献中心有 400 多个工作人员，其中有 250 人专门从事资料档的编制工作，200 人进行情报的选择和加工工作。填好工作单后送达数据穿孔部门，经输入建立法文的 PASCAL 文献资料档。每年收录 50 多万记录，实际上就是一个文献库的编制工厂。

二、文献资料档的应用

文献资料档最初是用于编制文摘刊物，是以副产品的身份问世的。但一经用到情报检索上，就大放异彩，引起了图书情报的联机革命。文献资料档除了用来进行检索外，还可以用其编制各种索引和专题文献目录，为图书情报机构提供咨询服务，帮助选定情报源等。用途是很广泛的，是一种重要的数据资源。

使用文献资料档进行情报检索的单位，主要有图书情报机构、商业情报检索单位，高等院校和科研机构等。文献资料档的提供目前有购买、租用和特许证三种方式。

由于编制文献资料档需要大量的人力和物力，而且今后情报的数据会越来越大。依赖一个单位独立完成资料档，不但成本高，而且是十分困难的。因此，出现了商品化的市售资料档。目前商品化的市资料档最常用的是磁带。因为磁带便于携带和保管，更重要的是价格便宜。提供这些资料档的单位大部分是经营多年的文摘刊物或索引杂志的出版单位。市售资料档比自己加工输入处理的成本低，二次情报的质量也较好，所以得到了广泛的采用。

近年来，我国也陆续引进了一些国外的市售文献资料档（一般称作文献数据库）。例如机械工业部科技情报所就引进了 INSPEC（科学文摘）、COMPENDEX（工程索引）、METADEX（金属文摘）和 ISMEC（机械文献题录）等四种国外文献磁带。据不完全统计截至 1982 年初，我国已经引进 25 种国外文献磁带。表 1·1 是载至 1980 年 9 月我国引进的国外文献磁带及收藏单位的情况。

表1·1 我国收藏国外文献磁带的情况 (数字统计到1980年9月止)

磁 带 名 称	收 藏 单 位	年 份	磁带盘数	估计文献量 (万)	备 注
世界专利索引 WPI	上海科技情报所	64—78 78—	16	100	累计带
科学文摘 INSPEC	机械工业部情报所	74—80	168	112	
机械文摘题录 ISMEC	机械工业部情报所	73—77 78—80	60 38	7.5 4.5	
工程索引 COMPENDEX	机械工业部情报所	74—80	84	60	
金属文献 METADEX	机械工业部情报所	74—78 78—80	60 24	15 6	题录文摘
化学文摘 CAS	化工部情报所	79—80	104	45	题 录
聚合物科学与技术 POST	化工部情报所	75—80	156		
生态与环境 EE	化工部情报所	79—80	52		
生物化学活性 CBA	化工部情报所	79—80	52		
能源 ENERGY	化工部情报所	75—80	156		
化工札记 CIN	化工部情报所	80	26	6	
可检物理情报通报 SPIN	中国科学院图书馆	76—79上 79—80上	7 18	8 4	累计带
会议文献索引 CPI	中国科学院图书馆	79—	10	20	累计带
美国政府报告通报 GRA	北京文献服务处	64—79 80	96 26	80 5	累计带
生物学文摘 BA (索引) (文摘)	南 京 大 学	77—78 77—78	4 8		
生物学研究索引	南 京 大 学	77—78	8	30	
石油文摘 TULSA	石油部情报所	65—69 70—80	10 33	} 22	
炼油文献和专利索引 API (文献) (专利)	石油部情报所	64—70 71—80 64—70 71—80	7 20 7 10		
地质文摘 Gec Ref	地质部情报所	78—80	33	13	
污染文摘 PA	科学院环化所	79—80	2	1.2	
美国国会图书馆目录 MARC	北京图 书 馆	78—79	104	44	
英国农业文摘 CAB	农科院情报所				

§ 1.3 提问逻辑与表示

1.3.1 引言

前节已经述及,情报的检索是根据用户的提问要求,从二次文献的资料档中检索出符合提问的文献。然而用户的检索要求是用提问逻辑式表达给计算机的。提问逻辑式能否正确地表达检索要求,将直接影响检索效果。提问逻辑式是由检索词组配而成。

所谓检索词是指表示文献任一属性词汇表的元素,它可以是作者、分类号、标题或文摘中的词、杂志名称、出版日期等。也可以是主题词表中的主题词。检索词之间用分隔符(如分号、逗号、空格等)分开。所以每个检索词都是由分隔符之外的字符串组成。在文献的描述中,各字段内的字段值是由各种检索词表的子集组成。检索词是描述文献的“语言”。要进行检索,相应也必须用描述文献的“语言”来描述用户提问,只有这样才能进行匹配。也就是说,文献的描述语言同用户的提问语言、存储用语与检索用语必须是一致的。

由于用户的检索需求是复杂的,单靠一个个孤立的检索词是不能反映出用户的提问逻辑,必须采取某些方法来表示构成一个提问的检索词之间的逻辑关系。又由于语言文字本身的特点和各种变化,也需要有一定的方法加以表示,以满足检索要求。由此就产生了用户的提问逻辑和具体表示方法。这些表示应该满足两方面的要求:一是明瞭,可以输入到计算机中进行检索运算;二是要能全面描述和表达用户的提问。

1.3.2 提问逻辑与表示

一、用布尔代数表示提问逻辑

一个提问逻辑可以用布尔代数的逻辑运算符将检索词组配成提问逻辑式。

布尔代数的逻辑运算符有 AND (与)、OR (或)、NOT (非) 分别表示逻辑乘、逻辑和与逻辑差运算。

AND (与)——这个算子是用来连接不同的检索概念的,例如:

COMPUTER AND LIBRARY

凡是在描述中包含 Computer (计算机) 和 library (图书馆) 的文献都将被检索。

OR (或)——这个算子用来连接同义的概念。例如:

COMPUTER OR AUTOMATION OR MECHANIZATION

凡是含有 Computer (计算机)、automation (自动化)、Mechanization (机械化) 三个词中任何一词的文献都将被检索。

NOT (非)——这个算子用来从检索中排除概念。例如

COMPUTER AND CATALOGUE NOT MARC

凡是含有 Computer (计算机) 和 CATALOGUE (目录) 而不含有 MARC (机器可读目录) 的文献将被检索。

利用这三个算子,还可以列出更复杂的提问检索式。例如

(COMPUTER OR AUTOMATION OR MECHANIZATION) AND LIBRARY

