

文献检索基础

高崇谦 朱孟杰 编著

书目文献出版社



文献检索基础

高崇谦 朱孟杰 编著



书目文献出版社

1983年·北京

内 容 提 要

本书编写的目的在于介绍给初学者以文献检索的基础知识，书中以浅显的语言来叙述国际上常用的文献检索方法及其一般性问题。

本书内容包括七章：概述；情报检索语言；主题标引工作；手工检索工具；机读检索工具；文献检索系统；文献检索效率。其中，特别是对情报检索语言和情报检索系统的建立作了比较详细的阐述。

本书可供文献检索初学者参考，同时也可供基层图书馆工作者、情报工作者参考。

文 献 检 索 基 础

高崇谦 朱孟杰 编著

书目文献出版社

（北京天津街七号）

秦皇岛市第二印刷厂排版

国家统计局印刷厂印刷

新华书店北京发行所发行 各地新华书店经售

开本787×1092 1/32 5 1/4 印张 127千字

1983年12月北京第1版 1983年12月北京第1次印刷

印数1—12,500册

定价：0.65元

图书分类号：G354

统一书号：7201·50

目 录

第一章 概述	1
一、文献及其类型.....	1
二、情报检索的定义.....	5
三、对文献检索的要求.....	6
四、文献检索的基本过程.....	7
第二章 情报检索语言	10
一、概述.....	10
二、分类语言.....	11
1. 分类、分类法.....	11
2. 分类表.....	13
3. 国内外几种主要分类法(表)简介.....	18
三、主题语言.....	43
1. 主题词表.....	43
2. 标题表.....	51
3. 国内外几种主要主题词表、标题表简介.....	56
四、分类语言与主题语言的比较.....	76
第三章 主题标引工作	79
一、概述.....	79
二、标引工作条件与组织.....	81
1. 标引工作的基本条件.....	82
2. 标引工作的组织与管理.....	83
3. 标引工作标准化.....	84
三、标引的一般过程.....	85
1. 主题分析.....	85
2. 检索标识的给定.....	86

四、标引规则的基本内容	87
1. 一般标引规则	87
2. 组配标引规则	90
3. 特殊标引规则	93
五、标引实例	94
第四章 手工检索工具	97
一、检索工具的概念及其要求	97
二、检索工具的种类及其特点	98
1. 期刊式检索工具	98
2. 单卷式检索工具	98
3. 附录式检索工具	99
4. 卡片式检索工具	99
5. 胶卷式检索工具	99
三、检索工具的编制方法	101
1. 编制目的与要求	101
2. 文摘的编写形式与要求	101
3. 索引的编制形式与要求	106
四、检索工具的查找方法	109
第五章 机读检索工具(市售磁带档)	112
一、概述	112
1. 磁带档的特点	113
2. 出售方式	114
3. 购买磁带档时注意事项	114
4. 调查二次文献磁带档的参考工具书	115
二、软件	116
三、磁带档再加工	116
四、标准化	118
五、磁带档记录格式	119

1. 文献目录信息及其要求	120
2. ISO标准格式	121
3. USGRA 格式	128
4. PANDEX格式	130
第六章 文献检索系统	133
一、概述	133
二、手工检索系统	134
三、穿孔卡片检索系统	135
1. 边缘穿孔卡片系统	135
2. 透光穿孔卡片系统	140
3. 机器穿孔卡片系统	142
四、电子计算机检索系统	146
1. 概述	146
2. 文献检索系统的基本流程	152
3. 数据库	155
4. 检索处理过程	158
5. 文献检索系统举例	162
五、缩微胶片检索系统	165
1. 胶卷检索装置	166
2. 缩微平片检索装置	166
3. 与计算机配套的检索装置	168
第七章 文献检索效率	169
一、概述	169
二、影响检索效率的因素	171
1. 标引质量	171
2. 主题词表的质量	174
3. 一致条件的使用	176
三、提高检索效率的措施	176
参考文献	179

第一章 概述

在科技情报工作中,情报检索(Information retrieval)占有很重要的地位。它是实现情报传递的一个重要环节,特别是随着电子计算机在情报工作中的应用,它的作用日益被人们所认识,并已逐步发展成为情报学的一门分支学科。

一、文献及其类型

文献是情报检索的主要对象,也是情报工作的基础。所谓文献,是指存贮在物理载体上按一定逻辑组织的任何知识内容的信息记录。记录文献的物理载体可以是书刊、卡片和缩微胶卷(片),也可以是录相带、录音带和计算机磁带等。由于文献记载着人类在社会实践中所积累的丰富知识和宝贵经验,汇集成无数的科学事实、数据、理论、定义和方法等,所以文献是情报自身存贮和传递过程中的一种最主要的形式。它在情报工作中的地位与作用,主要有以下几方面:

1. 文献是情报工作的基础

情报工作(一般指科技情报)是人们在生产或科研实践中记载的成果及经验经过传播为他人所利用的信息。情报工作者本身一般并不一定参加生产或科研实践。他可以是也可以不是情报的创造者,但通常是情报的加工者和传递者。文献是存贮情报最基本的形式,是情报工作的物质基础,没有

这个基础，情报工作就会变得无源之水，无本之木。

2. 文献是检索工作的基础

科学技术迅速发展，科技文献的数量与日俱增，千帙万函，难以数计；一个科研人员要浏览所有有关文献是不可能的。因此，就要靠文献检索。这种检索可以不受文献数量的限制，而且在时间上，利用累积的检索工具进行回溯性检索，可以达到特定的需求。

3. 文献是开展情报服务工作的基础

文献服务（有人也称情报服务）是情报工作的主要服务形式。其内容包括：①提供阅览；②答复咨询；③定题服务；④编译报道；⑤代译复制等。文献是提供情报的主要渠道。

按载体形式文献可分为：

（1）印刷型 包括铅印、油印、胶印等。这是一种有着悠久历史的传统形式，至今仍广为应用。

（2）缩微型 一般是指以感光材料为载体，利用光学记录的技术，使其文献缩小若干倍，如缩微胶卷、缩微胶片（平片）和缩微卡片等。

（3）磁带或磁盘 这是近年来由于计算机在情报工作上的广泛应用而产生的一种新的文献形式。它通过编码和程序，把文献转换成机器可读形式，输入计算机，存贮在磁带或磁盘上，需要时由计算机输出。因此，它必须借助于计算机才能使用。

（4）视听资料 亦称声象资料，如唱片、录音带、科技电影等。这类文献可以闻其声，见其形，给人以直接感觉。

对于科学研究，传播情报，均有其独特作用。

按内容、性质和加工程度，文献可分为：

(1) **零次文献** 指交谈或会议上交流和传递的有用情报，通常未经记载或出版。

(2) **一次文献** 亦称原始文献。所谓一次文献是指以作者本人的生产或科研工作成果为依据而创作的文献；不论撰写时是否参考或引用了他人的资料，也不论其载体或出版类型如何。

(3) **二次文献** 这是指对一次文献进行加工整理的产物，如目录、文摘、索引等。

(4) **三次文献** 这是指在一、二次文献的基础上，经过综合分析而编写的文献，如综述、述评、动态等。

按出版形式，文献可分为：

(1) **图书** 图书大部分是对科学研究成果、生产技术和经验的概括与总结。它的内容系统、全面、成熟，是传播科技知识、教育和培养科技人材的主要手段，也是科技情报工作的必要基础。图书一般分为科学图书、科普读物、生产技术图书、教科书和工具书等。

(2) **期刊** 期刊一般是指定期或不定期的连续出版物。它的刊名、出版形式固定，内容新颖，能反映当前科技水平。它是传递科技情报、交流学术思想最简便、最基本的手段。期刊的种类繁多，如学术性期刊、检索性期刊、消息性期刊、通讯性期刊等。它们是情报的主要来源。

(3) **科研报告** 科研报告是指科研工作成果的正式报告，或是对研究过程中每个阶段进展情况的实际记录。其特点是连续出版，上面刊有机构名称、统一编号，自成一册。

在内容上专深具体，往往是最新成果，比期刊论文详尽，数据也比较完整。

(4) 会议文献 会议文献一般是指国内外各种专业学术会议所发表的论文或报告。它基本上能反映出一门学科或专业的最新研究成果和国内外发展水平和趋势，是重要的科技情报来源之一。

(5) 专利文献 专利文献是指专利说明书。它是专利申请人向政府递送的说明新发明创造的书面文件。在说明书中，发明人常常论述其发明能解决什么特殊问题，使用何种方法，对旧有的产品有何改进及其用途等。因此，专利文献构成了科技情报的一个重要来源。

专利文献，还包括专利公报、专利文摘、专利索引等。

(6) 政府出版物 政府出版物一般是指由各国政府部门及其专设机构所发表出版的文件。这类文件品种多，数量大、从基础科学、应用技术，直到政治、经济、贸易等社会科学，它包括政府性文件和科技文献。政府文件，如国会记录、决议、法令、政策、统计等；科技文献，如科学研究报告、科普资料和技术政策等。

(7) 标准文献 标准文献是指对工农业产品和工程建设的质量、规格及其检验方法等方面所作的技术规定。它作为一种规章性的技术文献，有一定的法律性约束力。标准分为国际标准、区域性标准、国家标准、部颁标准、专业标准、企业标准等。

(8) 产品样本 产品样本是指对定型产品的性能、构造、用途、使用方法和操作规程、产品规格等所作的具体说明。产品样本一般包括产品目录、产品说明书、产品总览、

年鉴、手册、厂刊等。

(9) **学位论文** 学位论文是指高等院校研究生或毕业生在考取博士、硕士等学位时提出的论文。这些论文是围绕着一个专题进行的研究所取得的成果的总结，对于科研工作有一定的参考价值。

(10) **技术档案** 技术档案是指生产建设和科学技术部门在技术活动中所形成的，有一定具体工程对象的技术文件，图样、图表、照片、原始记录的原本或其复制本等。它是生产建设和科研工作中用以积累经验，吸取教训和提高质量的重要文献。技术档案一般具有保密和内部使用的特点。

(11) **其他文献** 其他文献，如报纸、新闻稿等。作为报纸刊载的科技消息，主要是科技发展展望、新的发明创造，以及新的成果等。报纸报道及时，阐明问题面广，具有通俗性，从中可获得一定重要的科技消息。

二、情报检索的定义

情报检索与世界上任何事物一样，总是从无到有，不断发展，不断完善的。在我国历史上可以说，自从有了文字就有了情报和检索，如将图书文献分为“部次甲乙”，就是使人们能“即类求书，因书究学”，以满足检索的要求。随着科学技术的迅速发展，情报检索就显得特别重要，现在它已发展成为情报工作中的重要组成部分。

情报检索，最早的含义是图书馆文献检索的同义词。可是，今天它已经跳出了这一狭窄的框框，而广泛地应用在社会活动的各个领域。由于图书馆学、目录学与电子计算机

技术相结合,开拓了传统图书馆学的新纪元,实现了计算机检索。所以情报检索又成为计算机信息处理的重要内容之一。

所谓情报检索,通常是指情报的存贮与查找的全过程。按检索对象大体上分为三类(参见图1—1):①数据检索(data retrieval),例如,喜马拉雅山有多高;②事实检索(fact retrieval),例如,长江、黄河和黑龙江,哪一河流最长;③文献检索(document retrieval),例如,关于电子计算机方面有哪些文献?数据检索和事实检索是要检索出包含在文献中的具体情报;文献检索则是要检索出包含所需情报的文献,是最典型的和最重要的情报检索。

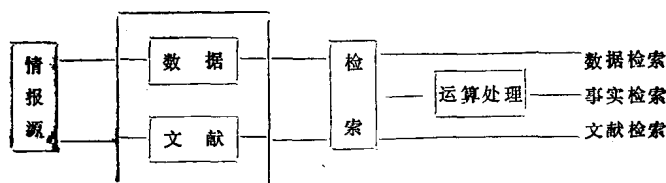


图1—1 情报检索的类型

文献检索的结果可以是与某一课题有关的文献本身,也可以是某一论文,书刊的出处和馆藏地址。如果了解和掌握了文献的查找方法,就能在最短的时间内,以最少的精力,得到前人和别人已经取得的经验和成果,为我所用。因此,文献检索是获得科技情报的主要手段。

三、对文献检索的要求

对文献检索的要求基本上有以下几方面:

(1) **灵活性** 要求能够从主题、著者、刊名、语种、出版年月等多种角度进行检索；就主题而言，要求即能按照分类号，又能按照主题词等语言进行检索。

(2) **全面性** 按照读者提问，将馆藏的相关文献尽可能都检索出来，不要遗漏，要求查全率高。

(3) **准确性** 检出的文献不仅满足提问要求，而且经读者判断也是准确的，例如，对于一个专利申请者来说，关心的是他的专利是否同已发表的专利重复。因此，要求专利检索愈准确愈好，即所谓查准率要高。

(4) **快速性** 由于文献数量愈来愈多，如何从大量的文献中迅速找出所需相关文献是一个很重要的问题。为了提高检索速度，除了合理地组织各种目录、索引等检索工具，还要采取新型检索设备，例如，电子计算机、缩微胶卷等。

四、文献检索的基本过程

文献工作，从其工作内容来看，大致包括：文献搜集、分析与标引、检索工具的组织、文献查找、复制、翻译、阅览服务等。从情报检索的定义可知，文献检索是由存贮和查找两个不可分割的过程组成。文献搜集、分析与标引和检索工具（例如，卡片目录、文摘杂志索引等）的组织是属存贮过程，也是文献查找的前提。查找可以采用人工方式，也可以采用机器方式，是文献检索的关键。为了有效地进行文献检索，必须建立一个完善的、先进的文献检索系统。现就文献存贮和查找过程简要地说明如下：

1. 文献的存贮过程

(1) 搜集文献资料 将搜集到的文献进行登记, 给出登录号。

(2) 进行文献分析和标引

①书目分析——分析文献的书目事项(包括书名、作者、出版项、稽核项等), 并按规定的格式进行著录。

②内容分析——分析文献的内容(即主题), 给出分类号和主题词, 即所谓文献标引。

(3) 给定文献存贮的地址 例如, 图书索取号、电子计算机磁带的位置、刊登论文的原始杂志的名称、卷期号、页码等, 都是文献存贮的地址。这是查找文献所在的外部标识, 相当于排架号。

(4) 编制馆藏目录、检索磁带档 将标引好的文献制成卡片, 按照一定顺序(如, 按分类号、标题字顺、书名字顺、作者字顺等)将卡片排在卡片盒中, 组成各种馆藏目录。当采用电子计算机检索时, 将标引好的文献按照规定的顺序记录在磁带上, 组成磁带档。

当上述文献存贮过程完成后, 便可将图书、期刊等放在书架上排好。

2. 查找过程

为了从存贮好的文献集合中找出满足提问的文献, 就得利用组织好的馆藏目录、磁带档等查找。具体查找步骤如下:

(1) 听取读者的提问, 弄清读者的意图。例如, 关于“教育”和“计算机”两个提问词, 要弄清需要的是计算机在教育上的应用, 还是计算机技术的教育方法。

(2) 将提问词转换为标准主题词。例如, 按照标题表或主题词表等的规定, 选用“电子计算机”而不是“电脑”。

(3) 选出的主题词(例如, 电子计算机、教育)用组配方式将读者提问内容表现出来。当采用《国际十进分类法》时, 如果提问既涉及到“电子计算机”, 又涉及到“教育”, 就用“电子计算机+教育”来组配表示。如果是关于电子计算机与教育的互相关系问题, 就可用“电子计算机: 教育”来表示。

(4) 按照上述方式组配的主题词(或分类号)查找目录或检索磁带档, 找出适合自己要求的索书号、书名、出版项等, 填写借阅单, 借出图书、期刊等。

归纳起来, 文献检索过程大体上如图1—2所示。

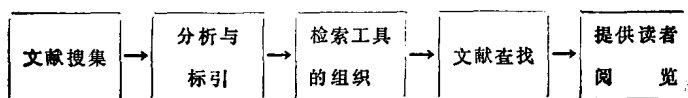


图1—2 检索一般过程

第二章 情报检索语言

一、概述

情报检索语言是根据情报检索需要而创制的一种人工语言，又称情报语言、检索语言、标引语言或索引语言，等等。它用于表达文献（或提问）主题，贯穿于情报的存贮和检索的全过程，是沟通标引（或分类）人员和检索人员双方思路的桥梁。

怎样研制和选用合适的情报检索语言是保证提高情报检索效率的重要条件之一。因此，一种理想的情报检索语言必须满足以下基本要求〔1〕*：

（1）应具有必要的词汇—语法手段，能准确地表达某一科学技术领域方面的任何文献（或提问）主题。

（2）每个词（或类名）必须概念明确，具有单义性。

（3）情报检索语言应便于将文献检索标识（主题词或分类号）同提问检索式中的项目相比较。

（4）检索语言体系应是科学、合理，便于使用。

任何书面语言，包括情报检索语言，一般由符号体系、词汇和语法组成。符号体系是指表示某种语言词汇所使用的代码系统。这种代码是由字母、阿拉伯数字或字母数字混合

* 方括号内的数字表示参考文献序号，下同。

组成的一串符号（例如，主题词表中的范畴号，分类表中的分类号等）；词汇是指文献检索中使用的主题标识（例如，主题词、标题词或类名等）。这种词汇是规范化的，以表的形式按照规定的体系或顺序汇集在一起（例如，分类表、主题词表、标题表等）。语法是指如何创造和运用上述标识来正确表达文献（或提问）主题，以实现有效检索的一套方法和规则（例如，组配标引规则、类目的复分手段和参照系统等）。

目前，国内外最常采用的情报检索语言基本上分为以下两大类：

（1）**分类语言** 这是用分类号来表达文献主题概念的语言，以分类表作为文献分类的工具。

（2）**主题语言** 这是用主题词或标题词来表达文献主题概念的语言，以主题词表或标题表作为词汇规范化的工具。下面将分别介绍分类语言和主题语言的基本内容及其特点。

二、分类语言

1. 分类、分类法〔2〕

按照塞耶兹（W. C. B. Sayers）所下的定义，所谓分类就是将事物按其属性的同异加以排列集中或分散处理的工作，图书的分类就是将图书按照某种等级体系排列在书架上或者将图书的书目排列在目录中的处理工作。如果按照美国图书馆协会格劳塞尔（Glossary）所下的定义，分类法这