

标 准 文 献

主题词标引原理与方法

主 编 胡 锦

副主编 周 洁

于良娇

中国标准情报中心信息与技术开发公司

1989年8月

标 准 文 献

主题词标引原理与方法

主 编	胡 锦
副主编	周 洁
	于良娇

中国标准情报中心信息与技术开发公司

1989年8月

编者的话

当前，标准情报工作已经进入一个新的发展阶段。标准情报检索手段的现代化，正在成为标准情报工作现代化的一个重要标志。各国标准情报中心都在逐步建立标准情报数据库。

国际标准化组织(ISO)为统一各国标准文献检索系统的主题词和著录格式，以利国际间的信息交换和资源共享，特组织BSI和AFNOR编制了ROOT主题词表和TIT主题词表，并编制了《ISO情报网手册》(ISONET MANUAL)。

中国标准情报中心根据标准情报工作的发展，以及GB 1.1—87的规定，翻译了《ISO情报网手册》，并以ROOT THE SAURUS为母本，编译出版了《标准文献主题词表》。《标准文献主题词表》的主要用途是：(1)作为新制订和新修订的国家标准、专业标准（部标准）、地方标准和企业标准标引主题词的工具书，(2)作为我国建立各级标准文献数据库和机器检索国际和国外先进标准的工具书，(3)在一定程度上可作为英汉——汉英标准化技术词典使用。

为了认真贯彻GB1.1—87的规定，做好各级标准的主题标引工作，我们编写了《标准文献主题词标引原理与方法》一书。该书对《标准文献主题词表》的结构体系和主题标引的原理与方法作了较详细的论述，并结合标准文献特点，提出了主题标引的规则，又以实例作了示范性说明。

本书不仅适用于计算机情报检索系统的文献标引与检

索，而且详细论述了手工检索工具的编制方法，也适用于手工编排目录和索引，是标准文献主题词标引的实用指南与工具。

本书共分八章，第一、二、三章由胡锦编写；第五、六、七章由周洁编写；第四八章由于良娇编写。

本书在出版过程中得到了扈明光、马随军、胡永明、李兰瑞、张楠、伊惠贤、高燕、陆绵忠等同志的大力支持，在此谨向这些同志致以谢意。

由于编者水平有限，书中恐有疏漏乃至谬误之处，希望读者在使用中提出宝贵意见。

编者

1989年8月

目 次

第一章 情报检索与主题标引	(1)
第一节 标准情报与情报检索系统.....	(1)
第二节 情报检索的定义和原理.....	(4)
第三节 检索语言与主题标引.....	(7)
第四节 文献著录与著录要求.....	(24)
第二章 标准文献主题词表的结构体系	(27)
第一节 主题词法原理.....	(28)
第二节 标准文献主题词表的结构体系.....	(35)
第三节 标准文献主题词表的语义网络系统.....	(49)
第三章 主题词的查找方法	(59)
第一节 各词表间的关系.....	(59)
第二节 词义的确定.....	(63)
第三节 使用词表的注意事项.....	(75)
第四章 文献标引工作	(78)
第一节 标引工作的意义和任务.....	(78)
第二节 文献标引的类型.....	(79)
第三节 文献的主题分析与标引的一般方法.....	(81)
第四节 标引步骤和方法.....	(88)
第五节 标引结果的审核.....	(91)
第六节 标引记录.....	(92)
第七节 标引人员的素质.....	(96)
第五章 主题标引规则	(99)
第一节 标引的一般规则.....	(99)

第二节	主题标引的专指性	(102)
第三节	通用概念词的使用	(105)
第四节	上位标引与靠词标引	(110)
第五节	标引深度的控制	(112)
第六节	过度标引的控制	(116)
第七节	标引的一致性	(120)
第八节	词表修改建议和新词的推荐	(122)
第六章	组配标引与组配规则	(127)
第一节	组配原理	(127)
第二节	组配的分类	(131)
第三节	组配规则	(146)
第七章	手工检索工具的编制	(149)
第一节	标引深度与引得深度	(149)
第二节	选定主题标题	(150)
第三节	主题目录与主题索引的编制	(155)
第八章	机检标准文档的主题标引	(169)
第一节	标引形式及著录工作单	(169)
第二节	标引词的分组与连字符的使用	(170)
第三节	适度标引和控制自由词	(173)
附录		(175)
一、	标准文献标引示例	(175)
二、	标准机读数据著录示例	(179)
三、	《ISO情报网手册》著录项目与格式	(188)
参考文献		(193)

第一章 情报检索与主题标引

随着新技术、新产业的出现，人类正面临着新的产业革命的挑战。而新技术、新产业正是大量科学技术的积累、发展和大量信息、知识的传播、应用的结果。这些新技术的发展促进了传统工业的技术改造。在此过程中，许多专家、学者发表了大量论文、专著，形成专利、标准文献，成为经济建设、科学研究、生产技术等必需的技术情报。

第一节 标准情报与情报检索系统

一、标准情报的重要性

标准情报是保证标准化事业和经济建设顺利发展的重要环节。它是科技情报的一个重要分支。标准情报工作就是对标准化活动中记载标准化科研成果和实践经验的**标准文献**，包括标准、技术法规、认证制度、专著、论文和期刊等以及其它领域中与标准化有关的情报资料，及时地、有组织地进行收集、整理、加工、存储和传递，以促进标准情报的社会交流和传播利用。标准化工作的发展在一定程度上取决于标准情报工作的好坏，取决于对标准情报的利用率和传递、吸收速度，及时了解、掌握和充分利用标准情报资料，对于社会进步，科学技术发展，提高产品质量，改进生产、促进外贸，提高经济效益，都具有十分重要的作用。

根据国际标准化组织(ISO)的有关统计,截止1980年底止,全世界共有各种标准1,000多种,75万件,连同标准化方面的会议文件、技术报告等共达120万件之多。六十年代至今,已有100多个国家和地区成立了全国性的标准化机构,其中90多个国家和地区制定了国家标准。有400至500个国际性和区域性组织参与了各类标准的编制工作,这400—500个组织中本身制定了标准或技术规程的大约有270个国际和区域性组织。标准文献的数量在不断增长。客观情况也要求我们必须做好标准情报工作。

二、建立情报检索系统的作用

用户需要的情报,有许多不同的载体形式,有印刷型情报,有记录在录音带上的声传情报,有电影或电视录象带上的图象情报,有拍摄在缩微胶卷上的缩微品情报,以及存储在计算机磁带、软盘上的机读数据情报和样品情报……等等。这些情报资料,近几年来发展异常迅速,数量庞大,类型繁杂,新陈代谢频繁,出版分散。因此要从如此众多的标准文献资料中迅速、正确地获得有用的、切合实际需要的文献资料,就必须解决情报资料的加工整理,建立情报检索系统。

情报检索是情报工作的一项重要内容。情报检索分为存储和检索两大部分。建立情报检索系统首先要解决存储问题,因为今日怎样存储,决定着日后怎样检索,但客观需要与主观可能性往往是有矛盾的,所以要建立一个先进的、经济 and 实际可行的情报检索系统是一项至关重要的,而又十分繁重的工作。

情报检索体现在标准化部门是标准情报检索,即标准

文献检索。它在标准化事业中起着重要作用。较为明显的如：

1.标准是一项引人注目，却能暗中主宰生产技术的文献。通常讲，标准化工作直接关系到生产技术和产品质量，主要是指标准的水平高低与生产技术和产品质量的关系。查找和应用国际上具有先进水平的标准文献，可以继承和借鉴国际上先进的科学技术，促进我国科学技术进步。

2.通过查找国内外标准文献，可以摸清各国有关专业的发展趋势、动向和水平，掌握世界各国的发展趋势，借以找到本国的发展途径。

3.标准是科研成果和生产实践的总结，查找国际和国外标准文献为我所用，可以提高技术水平，减少重复劳动，避免走弯路，迅速提高我国产品质量。如我国毛纺织品自1981年采用国际羊毛局标准后，已有130家工厂获得国际羊毛局的纯羊毛标志，产品在国际市场上颇有信誉，售价也提高了10~15%。

4.查找和采用国际、国外先进标准，可以缩短科研进程，节约科研时间，降低物资消耗，提高经济效益。如我国企业按照国际标准生产感光胶片，几年来节约的白银价值330万元。按照国际标准修订苧麻回潮率，每年挽回1600—2000万元损失。

据有关统计，世界上大多数科学家、学者和技术专家用于收集、整理科技情报资料 and 标准、专利文献所花的时间约占其全部科研时间的三分之一左右，如果科技情报和标准、专利情报部门工作做得好，就可以大大节省科研人员和学者的人力，节省他们的时间，并缩短科研进程。

第二节 情报检索的定义和原理

情报检索是情报工作的重要内容之一，是情报服务和情报研究的基础。建立一个科学的有效的情报检索系统是满足用户需求的基本保证。

一、情报检索的定义

情报的一词英语为information，俄语为 ИНФОРМАЦИЯ均源出于拉丁文informatio一字，原意是消息传递，还可译为信息、消息、报导、知识和见闻等。日本情报界认为“情报就是信息”，“情报存在一个发生源和吸收源，当发生源发出的信息被吸收源所理解时，就成为情报。”

情报检索是指人们对情报的存储和查找过程。查找的对象是用户需要的科技文献、标准文献或专利文献资料，故情报检索也称文献检索。

以一定方式将人类知识记录于一定载体之上所形成的物质形态资料称为文献，文献包含固定情报以及用于传递和使用情报的物质载体两个方面。换句话说，文献是固化了的知识，而情报是有固定对象，向一定方向运动着的动态性知识。

为了满足用户需要，原国家标准局筹建了中国标准情报中心，目前收藏标准文献资料有56个国家的国家标准，64个国际组织的国际标准和法规，400个各国专业协会的专业标准，以及有关标准化的期刊、专著、论文和译文等共33万件。面对如此浩瀚的标准文献资料如何建立好一个有效的检索系统以满足用户的需要，并不是一件容易的事。我国当前

正在借鉴和采用国际标准和国外先进标准，标准文献检索系统的工作效率高，直接关系到我国的采用国际标准工作，关系到我国的科学技术进步，关系到产品质量，也关系到技术监督和标准化事业的发展。

二、情报检索的原理

情报由于技术上和其它原因，不可能直接传递给分散在世界各地所有需要它的人，而且情报在传递之后的相当长一段时间里还会有新用户需要它。因此情报必须载录在载体上。亦即形成文献。这种记录下来的文献不仅能够比较容易地传递到遥远的地方，而且隔着一段时间也能为人利用，作为情报部门必须随时收集新产生的文献。为了能从大量文献资料中迅速而又准确地查找出用户需要的情报，就需予先把文献资料按照一定规则把文献特征有序地存储起来。

通俗地说，情报检索是指予先将文献按一定方式加工和存储起来，建立情报检索系统；然后根据用户的查找要求从检索系统中查找出来的过程。因此情报检索也称情报的存储和查找(Information storage and retrieval)。

情报检索工作的内容包括存储和检索两个部分。标准情报检索是指标准文献人员把搜集来的标准文献和数据，根据它们的外表特征和内容特征，按照一定的体系和规则，作：
(1)对文献内容进行主题分析，明确主题概念，从情报检索语言中选词标引；(2)按照著录规则进行标识和记录，编制检索工具和建立检索系统。文献人员又根据用户的需求，利用手工方式或机械方式按照相同规则，从检索工具中查找所需的情报线索，提供给用户；用户依照线索从文献库中提取

原始文献，从而形成检索输出。简言之，情报检索的原理乃是**根据一定的规则对文献内容进行概念分析、选词标引，确定分类号，组成检索工具**；又根据用户提问，按照相同的规则从检索工具中，查找出所需的特定情报的一种理论。

图1—1 是一个标准情报检索系统的原理图

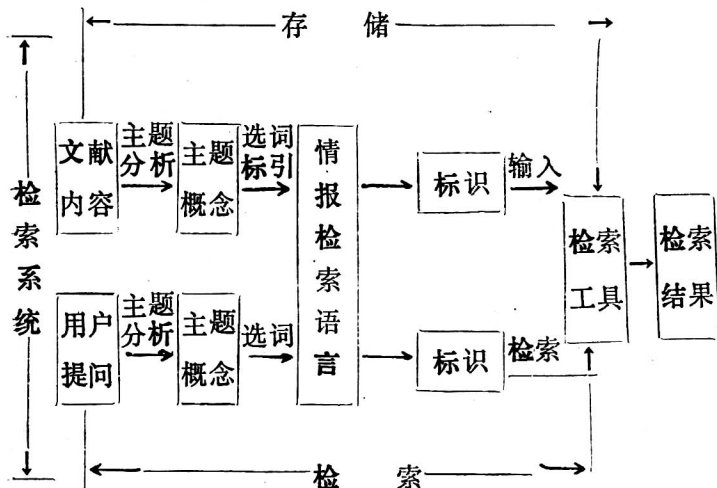


图1—1 情报检索系统原理图

情报存储和情报检索是两个有着密切联系的过程，其关系如图1—1。

当存储情报时，文献标引人员要做以下工作：

1. 首先对文献内容进行主题分析，即把它所包含的情报内容分析出来，使之形成若干能代表文献的主题概念；
2. 从文献的关键性语句中，精炼地压缩和概括出文献主题，从检索语言中，即借助于主题词表或分类表，选择一组恰当的语词（标识）把这些概念标示出来；

3.在适当的载体上,例如卡片、磁带、磁盘中存储并排序,通常称为编制检索工具,建立检索系统。

当检索情报时,情报检索人员要做以下工作:

4.根据用户的提问,首先要对检索课题进行主题分析,使之形成若干能代表情报需要的概念。

5.把这些概念借助于主题词表或分类表转换成情报检索语言的语词(标识);

6.然后从检索工具或检索系统中查找用语词标引的文献线索,从而找到所需的文献。

检索工具是特定设备与检索语词(标识)的结合。情报检索系统包括人、检索工具和文献资料三个要素。

情报检索系统分为手工检索系统和机器检索系统,两者都需要对大量的文献资料进行整理加工,按一定的规则和方法进行标识,才能有效地建立起情报检索系统。

第三节 检索语言与主题标引

情报检索语言是根据情报检索的需要而创造的人工语言。又称情报语言、情报存储与检索语言、文献语言、标引语言、检索标识系统等等。

标准文献的检索标识必须具有准确的语义概念,即应以压缩形式表达文献的内容特征,标识文献主题内容,使文献有序化,需要用特定语言形式来表示,这种特定的语言称为情报检索语言。

一、检索语言的种类

情报检索语言与情报检索效率有着密切关系。它在情报检索过程中所起的作用是极为重要的。

检索语言分为自然语言和人工语言两种。

自然语言是指人们日常生活中使用的未经规范化的语言。

人工语言是指从自然语言中优选出来的，并经规范化的名词术语或词组。

情报检索语言和人们日常生活所使用的自然语言有所区别。我们在加工文献资料时，首先遇到的是自然语言，要把自然语言所书写的文献资料的内容，用检索语言来描述，就需要有一种人工语言词汇或词典，用以把自然语言翻译成人工语言的一种词典。

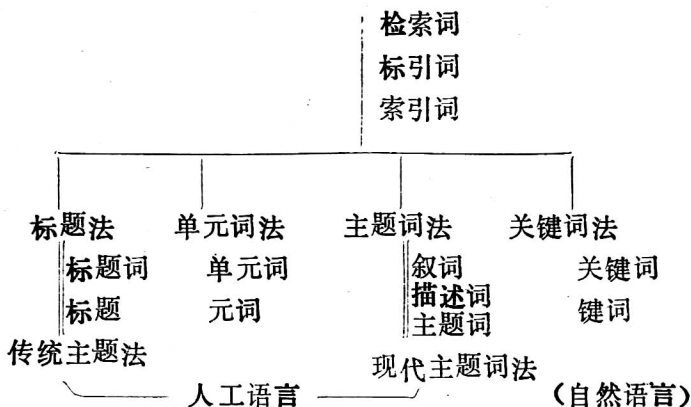
检索语言有许多类型，但基本上分为两大类，即分类语言和主题语言。此外，还有代码语言等。

从揭示和检索文献内容的方法来说，最普通、最常用的方法也是分类法和主题法两种。按文献内容的学科属性来系统揭示和检索文献的方法，称为分类法(Classification)，按文献内容的主题名称来揭示和检索文献的方法，称为主题法(Subjectindexing)。分类法和主题法是两种不断发展、不断完善的情报检索语言。

主题法语言分为标题词语言、元词语言、主题词(叙词)语言、关键词语言等四种。参见所附图1—2。

标准文献所使用的检索语言一般采用分类法语言和主题词语言两种。

图1—2 主题法语言（主题法系统）



现就分类语言、标题词语言、元词语言、主题词语言和关键词语分别扼要加以阐明。由于分类语言和主题词语言在标准文献中使用较多，故本章对分类语言和主题词语言的介绍，略多于其它几种，从第二章起论述的都是主题词语言和主题词表。

二、分类法语言

分类法语言分为体系分类语言和组配分类语言两种。

（一）体系分类语言

体系分类语言是一种直接体现知识分类的等级制概念标识系统。是对概括文献情报内容及某些外表特征的概念进行逻辑分类和系统排列而构成的。分类法的主要特点是按学科、专业集中文献，并从知识分类角度揭示各类文献内容上的区别和联系，提供从学科分类检索文献的途径。

任何概念都有内涵和外延。概念的内涵是指事物的本质

属性的总和；概念的外延是指概念的适用范围。概念按其外延是否有相同部分可以分为相容关系和不相容关系。在一般情况下，概念的外延所指的是同一类事物，但同一类事物并不是完全相同的。所谓“类”，是许多具有某种（或某些）共同属性的事物的集合，经过一次划分所形成的一系列概念称为子类或下位类（种概念），被划分的类称为母类或上位类（属概念）。由于事物往往有许多属性，它们都可以作为划分标准，所以分类是可以连续进行的，即经过一次划分所得的子类，还可利用别的属性作为划分标准再次划分。这样层层划分，层层隶属，便构成具有隶属、并列关系的秩序井然的概念等级体系。

分类语言一般以数字、拉丁字母或以拉丁字母和数字混合作为基本字符，采用字符直接连接并以圆点作为分隔符的书写法，以基本类目作为基本词汇，以类目的从属关系表达复杂概念的一种检索语言。通常以表的形式出现，因此也叫分类表。下面简单说明《国际十进分类法》和《中国标准文献分类法》两种分类语言。

1.《国际十进分类法》(Universal Decimal Classification, UDC) 是世界上比较流行的一种文献分类法，它以美国杜威十进分类法为基础发展形成的。它的特点是：(1) 类目详尽，居世界分类法首位；(2) 标记符号较灵活；(3) 比其它综合性分类法更适用于电子计算机检索；(4) 有国际文献联合会(FID)的常设机构主持增补、修订工作。但也存在一些问题，如思想体系、学科体系比较陈旧；在十个部类中科学技术仅占两个，这部分类号冗长，过于庞杂。

《国际十进分类法》的基本序列是：

- 0 总类
- 1 哲学
- 2 宗教、神学
- 3 社会科学、法律
- 4 (语言学)
- 5 数学、自然科学
- 6 应用科学、医学、工业、农业
- 7 艺术
- 8 (语言学)、文学
- 9 地理、历史、传记

(其中4 语言学已并入8 文学类, 准备将4 留作扩充科技类目用)

举“雷达的应用”为例, 用UDC进行分类:

- 6 应用科学、医学、工业、农业
- 62 工业、工业技术总类
- 621 一般机械工程、原子能、电气工程、机械工程总类
- 621.3 电气工程
- 621.39 电讯工程
- 621.396 无线电通讯工程
- 621.396.9 无线电通讯的应用、雷达
- 621.396.96 雷达
- 621.396.969 雷达的应用
- ∴ ∴
- ∴ ∴

上下位间存在着从属关系, 同位类间存在着并列关系。