

第1章 绪论



学习目标

1. 了解信息的概念和特征。
2. 了解信息源的特征。
3. 掌握信息源的类型。
4. 了解学习信息检索的目的和意义。

1.1 信息

当今,社会信息化已形成巨潮席卷整个人类社会。人类各种实践活动、各个领域都与信息和信息技术息息相关、密不可分,使得整个社会的生产方式、生活方式、活动方式、行为方式、思想观念、道德规范等方面都在发生根本变化。信息促使人类大脑在时空领域得到极大扩展,从而加速了世界经济的区域集团化、一体化的进程。

信息作为一种社会资源很早就被人们所认识,而且自古以来人类就在利用这种资源,只是利用的能力和水平很低而已。例如,我国古代发明的原始感测和显示技术的指南针、原始通信技术的烽火台,以及记录、存储信息的造纸术和印刷术等。

1.1.1 信息的定义和特征

1. 信息的定义

当今世界,“信息”已成为社会生活中使用频率最高的词汇,与之相连的“信息经济”“信息产业”也成为人们身边的现实。信息作为重要的经济、战略资源和推动社会生产力的新动力,日益受到人们的青睐,“信息时代”“信息社会”正在向我们走来,信息化成为当代社会发展的大趋势。

“信息”一词被高频率地使用,但究竟什么是信息呢?怎样认识信息的科学意义呢?我们知道世界万物都有自己的运动规律和表现特征。社会生活中人们了解到的事物间的差异都是通过一定的形式表现出来的。例如,声音、图像、文字等,这些形式所表达的内容就是信息。信息不是物质本身,而是物质的一种基本属性。

拉丁语中的“信息”一词是陈述、描述、概要的意思；英语中的“信息”是 information，有知识、资料、数据、情报、消息、报道、报告、通知等多种含义；汉语中的“信息”一词，“信”和“息”两字的意思相近，均有音信、消息的意思，但前者侧重于消息、征兆，后者强调情况、音信。

那么，信息到底是什么呢？目前，学术界还没有形成一个一致的、完整系统的定义。《辞海》中是这样定义的：信息是指客观的消息、情况、情报等，是指对消息接受者来说预先不知道的消息等。《新华字典》将信息表述为：信息是指音信、消息。信息论中是指用符号传输的消息，报道的内容是接受者预先不知道的，是事物的运动状态和关于事物运动状态的陈述。《韦伯斯特词典》将信息表述为：信息是知识和情报的通信和接受，是通过调查、研究或要求而得到的知识，是情报、消息、新闻、事实和数字等。我国科学家钱学森也对信息有过这样的阐述：信息是激活了、激化了的的知识，是为了解决一个特定的问题所需要的知识。

综上所述，信息就是消息，人们在学习、工作、日常生活中随时随地都在接受和利用信息；信息是资源，它具有使用价值和价值；信息是财富，并且是无价之财富；信息是生产力要素，更是一种不可估量的促进生产力发展的新动力。虽然学术界难以形成统一的信息定义，但从文献信息探索的角度，我们应对信息做一个基本的约定：信息是自然界和人类社会中一切事物自身的运动状态与方式，是事物的内在联系和含义的表征，信息通过一定的载体形式反映事物的本质特征及其运动变化。也就是说，信息是一种可以用来消除不定性的东西，它泛指一切事物运动的状态和运动的方式，是关于事物运动状态和运动方式的反映。正因为信息是事物运动状态和运动方式的反映，所以它才能用来消除人们认识上相应的不定性，不断提高人类认识世界和改造世界的能力。因此，人类社会的发展、科学技术的进步，都离不开信息资源的开发和利用，而且信息资源的真实状况及开发利用程度，已成为衡量一个国家经济、文化、科技，乃至综合国力的重要指标。

2. 信息的特征

信息与物质、能量并称为人类社会的三大基本资源，但是与物质和能量不同，信息具有以下特征：

(1) 信息具有客观真实性。信息既不是物质，也不是能量，信息就是信息。不论你是否认识到，信息总是存在的。信息的这一最重要的本质特征是其生命所在。

(2) 信息具有共享性。信息的共享性是它与其他资源最大的不同之处。物质和能量被专人占有和使用，进而被消耗而不复存在。信息显然不同，当某一信息被多人共享时，信息的占有者不会丧失该信息，而且，一般来说也不会因为多次使用而改变信息的自身价值。作为一种共享性的资源，信息可以通过传输和扩散的方式实现共享，在交换、传递的过程中，信息不但不会丧失，而且还可能同时获取新的信息。因此，信息还可称为一种无限再生的资源。另外，我们在实际生活中应当注意区分信息的共享性与现实共享范围，了解独占信息、有价信息和公共信息的特性，合理合法地充分享有信息。

(3) 信息具有可存储性。信息的可存储性指的是信息必须依附于各种形式的载体，信息和物质(事物)、能量具有密切的关系，但它仍需依附于另一种形式的载体。所以说，没有物质、没有能量，就没有事物的状态和方式，也就没有信息，而且信息的识别、转换、扩充、压缩、生成、分享等所有的过程都必须有物质和能量的支持。信息通过物质载体来存储，从远古的结绳记事、甲骨、泥板、纸张到磁、激光、大规模集成电路的广泛应用，大大丰富了信息的

存储形式,也极大地增加了物质载体存储的信息量。

(4) 信息具有可转换性。信息可以被加工、整理、概括、归纳使之浓缩精练和不断扩充并变换各种形态,由原始一次信息变为二次信息、三次信息等,实现信息价值的增值。运用现代信息处理系统,将原始文件进行压缩,可达到便于传输和节省存储空间的目的。信息技术的高速发展可使信息进行声音、图形、文本、数据等多种形式的相互转换,有效地促进信息的传播和交流。例如,可将声音转换为文字或将文字转换为声音、将统计表转换为拼图等。

(5) 信息具有传输性。信息不同于物质,作为一种既无大小又无重量的非实体,它的传递速度仅取决于其承载体的性能。因此,计算机技术和光纤技术的高速发展使信息的传递超出了以往载体的限制。当信息的载体是电磁波和光波时,其传输的速度将达到每秒30万千米,而运用广播、电视、Internet等大众传媒传递信息,可在瞬间将信息迅速、广泛、全方位地扩散。

(6) 信息具有极强的时效性。信息与物质、能量的价值观不同,信息有很强的时效性,过时的信息就是马后炮。例如,天气预报信息、军情与警情信息、战略预测信息等,如果不在最需要的时间内被准确地获取、传递、使用,它将分文不值。也就是说,信息的利用价值与时间成反比,及时有效地把握信息就能实现信息的最佳价值。

(7) 信息具有增值性。信息通过人脑思维或人工技术的综合、加工和处理,不断积累丰富,提高其质量和利用价值。信息交换的结果是信息的增值。

其实,信息的特征还有许多,如知识性、识别性、生成性、动态性等,这里限于篇幅,不再一一介绍。

1.1.2 信息的类型

信息类型繁多,从不同角度划分会有不同的信息类型。

(1) 自然信息和社会信息。这是依据信息存在的领域及信息的内容划分的。自然信息是指存在于人们主观意识之外,来源于自然界,在自然界中传递的信息。社会信息则来源于人类社会,是人类对于社会现象和社会活动的认识、反映,是在人际间传播的信息及思维信息。

(2) 语言信息和非语言信息。这是依据信息的符号种类及信息的性质划分的。语言信息是由作为信息沟通最基本、最重要的工具——语言来表现的信息,它由语法信息、语文信息和语用信息构成。人们总是以信息为媒介与事物打交道,通过语法信息获得语文、语用信息,经分析比较产生决策信息,根据决策信息使事物按人们预期的目标发展。非语言信息则是指表情、体态等非语言显现的信息。

(3) 动态信息和静态信息。这是依据信息的状态划分的。动态信息是指时效性极强的瞬息多变的信息,如军情、警情、新闻等类信息。静态信息则是指呈现稳定、相对静止状态的信息,如历史文献、档案资料等类信息。

(4) 一次信息、二次信息、三次信息。这是依据信息的加工处理程度划分的。一次信息是在生产实践、社会实践和科学实验中产生的新成果、新知识和经验性总结等方面的信息;二次信息是依据客观需要,按照一定的科学方法,将特定分散的一次信息进行加工整理而形成的信息;三次信息是指依据二次信息提供的线索并选用一次信息的内容进行综合分析形成的信息。

(5) 单媒体、多媒体信息和文献及电子信息。这是依据信息的媒体和信息的载体划分

的。用来表示信息的声音、图像、文字、数据,以及广播、电视等物质载体是信息的表示媒体。只有一种媒体表示的信息是单媒体信息,运用多种媒体综合、协调地表示的信息是多媒体信息。多媒体信息不同于多媒体技术和多媒体计算机系统的概念,而现代化的多媒体手段可以将多媒体信息带入一个崭新的应用领域。

(6) 政治信息、经济信息、文化信息、科技信息、学术信息、法律信息、军事信息、市场信息等,这是依据信息的社会属性划分的。

1.1.3 信息、知识、情报与数据

1. 信息与知识的关系

信息不同于知识,它是作为物质运动的反映,通过各种信息通道达到人脑成为人类意识。在人们获取、整理、利用信息之前,信息处在一种发散的状态,时刻在变化、流动、产生与消失。如果人们不断地接收信息,而且这些信息又是真实可靠的,经过人们的分析、提炼,就能准确把握客观事物的规律性,再归纳总结,这种“信息”就上升为知识、科学了。

知识是人类积累的关于自然和社会的认识和经验的总结,是人类通过收集和利用信息对自然界、人类社会及思维规律的认识与掌握,是人的大脑通过思维重新组合的系统化信息的集合。知识既包括有文字记载的认识和经验,也包括没有文字记载的认识和经验。知识是吸收了信息后凝结出来的概念,是系统化、完备化的信息,是信息中被人类总结出来的精华部分。

信息与知识往往是联系在一起的,信息可以被认为是一种可传播形式的知识,信息与知识的关系如同金钱与财富的关系。只有当信息被系统地、有目的地积累起来时,它才能转变为知识。

人类既要通过信息来认识世界、改造世界,也要根据所获得的信息去组成知识。所以说,知识是信息的一部分,信息则是“流动的”知识,它是构成知识的原料,这些原料经过人脑接受、选择、处理,就能组合成新的知识。信息相对变动,具有动态属性,而知识相对稳定,具有静态属性。

2. 信息与情报的关系

信息与情报是两个最为密切的概念,以往两者的使用没有任何区别,如英文 information 表示的就是“信息”和“情报”。随着信息科学的发展,“信息”一词才逐渐取代了“情报”一词。

情报一般是指在特定的时间、特定的状态下,为特定的人提供的有用知识,情报的本质就是知识。《辞海》里是这样解释的:“战时关于敌情之报告,曰情报。”情报往往与军事、间谍相关联,带有机密的色彩,随着社会的发展,情报已不仅仅局限于军事领域,还适用于政治、经济、军事、社会生活等诸多领域。与信息 and 知识的关系一样,信息和情报两者也是既有联系,又有区别。信息包含了情报,因为信息既包括有关军事和间谍相关的情报,又包括更多方面的“情报”“知识”,既有社会生活中人与人之间的相互沟通、思想交流、情感传递,也有人与自然的“对话”与交流,还有纯客观自然界的自然物,非自然物,动物,植物相互的信号、消息交换与传递。综上所述,信息、知识与情报的逻辑关系如图 1-1 所示。

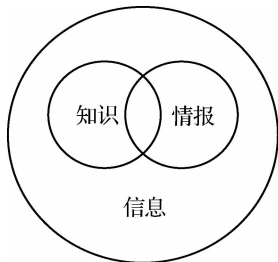


图 1-1 信息、知识与情报的逻辑关系示意图

3. 信息与数据的关系

在计算机普及的条件下,信息与数据的概念常常是混用的,实际操作中信息本身也已数据化,而数据本身也确实代表着信息。数据指的是表达事物信息的字符集合,它由特定的符号或自然数表示的数值型数据和计算机能够识别的在计算机环境下的非数值型数据构成。凡是能够用计算机进行编码和通信的数据都可以称为信息,只有被定义了的数据才是信息。信息就是加工了的数据,是数据的内容。

信息与数据通常是难以分割的,但使用中也须加以区分。比如“2008”这样一个数据,作为一般的单纯自然数并不表示什么信息,但把这一数据放在“奥林匹克运动”上,它就可能表示“北京 2008 年奥运会”这样一则信息了。如果它出现在飞机的速度表上,表示的肯定就是飞机的速度信息了。因此,数据是一种形式,是一种表示信息的形式,它所表示的内容就是信息。但信息并不仅限于数据表示,它还可以用图形、图像等形式表示。我们还可以说数据作为一种素材、一种原始资料被有序地组织和“创作”而得到的“产品”就是信息。

1.2 信 息 源

1.2.1 信息源及其特征

1. 信息源的含义

所谓信息源,就是人们为满足信息需要而借以获得信息的来源。信息源既可以是产生和持有信息的个人和机构,也可以是负荷信息的载体。人类对于自然、社会及思维的知识均来源于实践。我们从信息的内涵知道,信息本身既不是物质,也不具有能量。因此,任何信息都离不开一定的物质载体和能量,信息总是以一定类型或形式的载体作为其存在和传递的物质基础。

信息载体又可分为信息表征载体和物质载体两类。信息表征载体是指信息内容赖以存在的文字、符号、声波、光波等用以记录信息的载体,亦称第一载体。物质载体是指信息载体赖以存在和传递的纸张、胶片、磁带、光盘等物质材料,亦称第二载体或硬载体、载体的载体。只有第一载体和第二载体有机结合,才能保证信息内容得到完整地保存和有效地传递。随着信息科学、材料科学的不断发展,信息载体也会相应地发展变化。

2. 信息源的特征

(1) 信息数量多、内容庞杂。从某种程度上来说,各种信息收藏机构收藏的信息资源包括传统信息资源(纸质文献、各类光盘、缩微胶片等)及网络信息资源,这些信息资源数量多,内容纷繁复杂,涉及社会生活的各个方面。特别是网络信息资源包括成千上万种电子期刊、报纸,以及政府、学校和公司等机构和个人的详细信息,信息具有跨地区、分布广、多语种、高度共享的特点。网络信息资源既有最新的各类信息,也有毫无价值的大量冗余信息,数量巨大、内容庞杂。

(2) 信息的半衰期缩短,老化加快。近年来,随着科学技术的快速发展,各类文献资料及网上信息不断淘汰和更新。新知识、新理论、新技术、新产品层出不穷,加速了信息的新陈代谢。我们通常用文献的“半衰期”来描述文献的老化情况。

文献的半衰期是指某学科领域目前尚在使用全部文献中,较新的一版所出版的年限。国外有人统计不同学科文献的半衰期为:地理学 16.1 年、地质学 11.8 年、数学 10.5 年、植物学 10 年、化学 8.1 年、生理学 7.2 年、机械工程学 5.2 年、社会科学 5 年、化工 4.8 年、物理学 4.6 年、冶金学 3.9 年、生物医学 3 年。由于各国科技发展水平不同,相应的文献寿命也不相同。

(3) 信息类型多、范围宽、用途广。信息资源建设得以快速发展,信息类型越来越多,包括印刷型出版物、电子期刊、图书、工具书及大量的书目数据库、联机数据库、软件资源,以及个人主页、电子邮件等种种形式。信息呈现出多种表现性,仅文本就有文字、符号、表格等,另外还有声音、动画、图像等。人们通过网络可以方便地进行人文科学、自然科学的信息交流,如发送电子邮件或召开在线视频会议等。

(4) 电子信息成为一种主要的信息资源形式并表现出鲜明的特点。随着计算机技术和情报资源相结合,产生了一种新型的信息资源——电子信息,这种信息资源以数字形式存储,通过计算机或网络检索或传输,加快了信息流通和信息资源共享,以书目、索引、文摘、全文数据库、多媒体信息、电子刊物等形式为用户提供服务。

网络信息资源是一类重要的电子信息,它通常以超文本技术链接,把各类不同的相关信息都通过节点链接起来,使得检索相关信息非常方便。但是这类信息组织特殊、控制性差(尽管从局部来说某个网站、某个数据库是有控制的、相对集中的、有序和规范的),信息质量良莠不齐,所以其优势还没有完全发挥出来。

1.2.2 信息源的类型

信息源的种类繁多,但总体上可分为文献信息源和非文献信息源两大类型。两类信息源各具特点,一般来说,文献信息源信息量大,相对稳定,易于组织、管理和检索;非文献信息源相对而言,具有生动直观、选择性和针对性强、获取速度快的特点。

1. 文献信息源

文献信息源是正规的信息源,是人类社会所特有的人工信息源。凡是以文献作为载体形式的信息源均可称为文献信息源。文献在保存和传递信息方面具有重要的作用,可以说文献信息源是人们获取信息最基本、最主要的来源。文献信息源具有特殊的优势:能够促使信息得到长久的保存和广泛的传播;所载信息固定而明确,便于进行多方面的加工和利用;是确认人类活动,进行社会规范的工具。文献信息源的主要缺陷就是在传递和交流上不够灵活和生动。

2. 非文献信息源

非文献信息源包括电子信息源、实物信息源、口头信息源。

(1) 电子信息源是指含有各种信息的电子资源。电子资源按其载体形态划分,主要有两大范畴:一是以光盘、磁盘、磁带等脱机载体出版的有形电子资源,二是在网络上出版的网上电子资源。数据库光盘和多媒体光盘曾经是图书馆电子文献的主要组成部分,但随着

Internet的日益普及和飞速发展,网上资源的优势日趋显著。与光盘相比,网上资源不需要光盘塔等设备的投入和人员的日常操作维护,内容更新周期更快。因此,网上资源将成为图书馆电子资源的主体,光盘只能作为数据备份的辅助工具,这是一个不可逆转的趋势。

电子资源按发布形式不同可分为数据库、电子期刊、电子图书、电子报纸、动态性信息、软件等。其中,数据库是历史最久、影响最广的一种电子信息资源,它以先进的数据处理技术、深加工的信息,以及与网络的完美结合,成为电子信息资源的主体。数据库涉及领域广泛,按其内容可划分为以下4种:

- ① 书目数据库,包括索引、文摘等二次文献数据库,如EI、SCI等。
- ② 数值数据库,主要包含数字数据,如统计数据、科学实验数据、科学测量数据等。
- ③ 事实数据库,收录人物、机构、事务等的现象、情况、过程之类的事实性数据,如机构名录、大事记等。

- ④ 全文数据库,提供完整的原始文献数据,集检索功能与浏览原文功能于一体,是目前最便捷、最具潜力的电子资源。

电子信息源的特性如下:

- ① 易获取性。电子资源借助计算机与网络,向用户提供检索和传递资料的服务,Internet穿越时空的优势使读者在任何时候、任何地点均可通过网络获取电子资源。电子资源的提供使图书馆的服务在时间和空间上得以延伸,使信息的获取更为便捷。

- ② 共享性。与传统文献相比,电子资源使多用户可在同一时刻下共享、共用同一信息源。传统文献,如一册书或一本杂志,同一时间只能供一个读者阅读。而电子资源,如一个数据库可被多个用户同时检索,同一篇资料可供多个用户同时下载、浏览,体现了电子资源特有的实时共享性,提高了资源的利用率。

- ③ 资料更新快。在科学发展突飞猛进、知识更新日新月异的当今时代,信息的时效性显得尤为重要。在减少了印刷型文献印刷、装订、运输等烦冗环节之后,网上出版物和数据库通常以日或周为更新周期,其高效的动态更新和快捷的实时传递,在充分发挥信息的时效价值方面有着传统文献无可比拟的优势。

- ④ 形式的多样性。多媒体技术在电子资源中的利用,使文字、图形、声音、动画、三维图像等相结合,提供丰富多彩、生动逼真的信息,使读者能够闻其声、见其形,有身临其境之感。与此同时,超文本技术的运用改变了传统信息的线性组织方法,增强了人们对知识的表达能力,大大拓展了信息的获取和传播范围,使读者更容易理解和接受。

- ⑤ 知识的类聚性。受手工检索工具对文献揭示的不充分性和检索手段烦琐的限制,传统文献中大量有用知识被隐藏。而在网络环境下,依靠强大的检索软件可以使不同类型的相关信息聚集并有序化。目前网络数据库有大型化趋势,同一公司制作的多种数据库产品的检索界面趋向统一,并可多库操作,对检索的完整性提供了可靠保证。除各数据库之间的链接外,数据库还可与图书馆书目查询系统进行链接,这些链接方式可以帮助读者从书目、题录、文摘等二次文献直接查阅到全文。

(2) 实物信息源是指蕴含信息的实物载体,包括自然实物信息源和人造实物信息源两大类。自然实物信息源如植物、动物、河流、山川、土地、矿物、宇宙等,作为未经加工的自然存在的能够直接反映事物的原始状态,人们可以从中获取有关自然界的的信息,如气象的变化、矿产的蕴藏、动植物的繁衍等。人造实物信息源是经过人类有目的的加工而成的物体,

如产品、样品、样机等。它是人类活动的产物,具有内隐性。其内涵为加工工艺、化学成分、物质参数、设计指标、外观状态等。对这类信息源的利用,需要经过较为复杂的分析研究,将蕴含在实物中的信息解释出来才能够实现。实物信息源的主要优势如下:

① 真实可靠。通过实物,人们可以直观而生动地掌握其内涵。

② 信息浓度大。通过分析,人们可以从中掌握生产实物的一系列活动。

③ 易于开发利用。通过实物,人们可以充分利用其信息进行仿制、改进和创新。实物信息源的主要缺陷就是传递不方便,而且蕴含的信息必须经过分析与提炼,转换成其他形式才能被人们认识和利用。

(3) 口头信息源。口头信息以大脑为载体,是人与人之间的对话、交谈、演讲等。现实社会中的每一个人,每天都在进行信息的口头交流,不断地发送信息又不断地接收信息。口头信息源有以下优势:

① 获取信息快,交流及时。

② 具有高度的选择性和针对性。例如,从研究同一课题的同行处获得有关的口头信息,较之查找大量散见于各种文献中的信息,既方便,又对口。

③ 信息反馈及时、准确。在直接交流时,对于所涉及的问题双方可以立即询问澄清,必要时还可以互相补充和修正,避免理解上的错误。

④ 根据人们接触时的气氛、语气、手势、暗示等,可以领会语言文字不能表达出来的东西,因而有助于全面了解对方所谈的内容,对所获得的信息进行适当评价。

⑤ 可以提供其他信息不能包含的许多信息,而这些细节往往更具启发和借鉴作用。当然,口头信息源的缺陷也是显而易见的。由于人们直接通过口头交流来获取信息的机会总是有限的,机会分布也不均衡,所获信息的满意度和完整性极不稳定;由于难于对口头交流实行有效的社会监督,故其可靠性也不易检验,加之难于积累和整理,信息会随着时间的推移而失真和泯灭。所以人们在获取、传递和利用口头信息时,常常将它转换成文献信息。

1.2.3 文献及其特征

人们常说“文献信息”。其实,文献与信息密不可分。人们将信息、知识、情报记录在一定的载体上就形成文献。文献必须含有信息,信息也必须依附在一定的文献载体上;文献是信息的静态物质形式,信息则是文献的动态抽象形式。人们利用文献,实质上就是利用文献中的信息。因此,可以说文献是继承发扬科学知识的接力棒,信息与知识上升为科学记录于文献之中,又成为新的信息来源。如此反复循环,推动整个人类的文明进步。

在社会生活中,通过对文献信息资源的搜集、分析、整理和存储,向需要利用所藏文献信息资源的人们提供相关服务最常见的信息服务的机构就是图书馆。图书馆不仅为收藏知识积累了丰富的文化典籍,而且越来越成为大量信息的集散地,从而形成了多层次、全方位的馆藏文献信息资源。

1. 文献的定义

文献是伴随着人类对史料的记载而产生的,是用文字、图形、符号、声频、视频等技术手段记录人类知识的一种载体,或理解为固化在一定物质载体上的知识,也可以理解为古今一切社会史料的总称。

现在通常理解为图书、期刊等各种出版物的总和。文献是记录、积累、传播和继承知识

最有效的手段,是人类社会活动中获取情报最基本、最主要的来源,也是交流、传播情报最基本的手段。

随着时代的变迁,文献载体不断丰富,书籍日益增多,文献的概念也发生了变化,文献已不兼指书和人,而是专指著述。

随着文献交流的日趋频繁,学术界也力求对文献做一个统一的概定。其主要代表性的定义有:泛指一切文字、图形、音响资料和视听资料,是以特定的物质为载体的人类精神信息的固态品;用一定的技术手段将一定的信息内容记录到一定载体上形成的存储型传递媒介;记录信息与知识的一切人工附载物。国际标准化组织的《文献情报术语国际标准草案》中所指文献是:在存储、检索、利用或传递信息的过程中,可作为一个单元处理,依附载体而存储有信息或数据的载体。中华人民共和国国家标准《文献著录总则》(GB 3792.1—1983)中的文献则是记录有知识的一切载体。从以上文献的表述来看,笔者认为,文献是指人们对信息加以有序化、系统化,并以文字、图形、符号、声频、视频等信息符号记录在一定载体上的固态精神产品。

文献是重要的信息资源,是人类知识的物质保存形式。文献的定义由4个要素构成:

- (1) 文献内容,即被记录的信息、知识。
- (2) 信息符号,即记录信息、知识所用的符号,如文字、数字、声音、图像等。
- (3) 文献载体,即除人脑以外的用于记录信息、知识的物质载体,如最早的泥板文献中的泥板,以及竹简、甲骨、石头、绢帛、纸张、磁带、光盘、磁盘等。
- (4) 记录方式,即信息、知识被记录、存储到载体材料上的方式,如铸刻、书写、印刷、录制、摄制、刻录等。

其中,文献内容和文献载体为最基本的要素,两者相互依存构成文献。各类文献经过加工、整理、布局就构成文献资源。文献作为信息的宝库,是人类文明成果的总汇,是一种宝贵的智力资源和精神物质,是社会进步的重要标志。丰富的文献资源能够为人们提供前人或他人的已有成果,避免无效劳动,为社会创造效益。

2. 文献的特征

文献的特征分为外部形态特征和内部形态特征两大部分。文献特征是编制各种文献检索工具时的著录依据。

(1) 文献的外部形态特征。

① 文献名称:它是文献内容的概括。一般情况下,从一本书的名称就可以知道文献的主题与性质,然而有一些文献的名称不能反映其学科内容,如《根》,根据文献名称可判断是一本植物学内容的书,其实是叙述美国家族历史的书。

② 著者名称:它是指对文献负有一定责任的个人或机关团体。著者有本国和外国、团体和个人、古代和现代、主要和次要之分。

③ 版本:包括版次和版刻,指文献的排版次数,用来标明文献版本的重要变更。

④ 出版地:是指出版者所在地的地域名称,可反映地方性出版物的特点,往往也是用户选择文献的条件。

⑤ 出版者:包括负责整理、复印出版的机关团体或出版企业组织等。

⑥ 版期:它取决于版次,主要指某版次的出版时间。

⑦ 载体形态:指文献的页数、卷数、插图、折图、彩图、照片、开本、尺寸等。

⑧ 标准书号与刊号:指图书的 ISBN 号和期刊的 ISSN 号。ISBN 为国际标准书号的代号,后面有 10 位数字,分 4 段组成,每段之间以短横线间隔。第一段为国家、区域号,由国际标准组织(ISO)拨给;第二段为出版商代号,由国家、区域的有关标准化组织拨给;第三段为书号,由出版者自己给出;第四段为稽核号,为计算机核查数字。凡参加国际标准组织的国家,出版的书都可以根据国际标准书号编制法编制各国的标准书号。我国已正式加入国际标准书号系统,并确定我国第一位数字的组号为“7”(如“0”代表美国,“1”代表英国,“2”代表法国,“82”代表挪威,“90”代表荷兰)。一般单个数字代表较大的地域或国家,双位数字代表较小的地域或国家。例如,一本《现代信息检索与利用》的国际书号为 ISBN 7-5354-1988-8。

(2) 文献的内部形态特征。文献的内部形态特征是指可以揭示文献内容和实质的标识,有分类号和主题词两种。

① 分类号:文献分类就是根据确定好的逻辑系统(分类法),将文献信息所载知识单元及形式特征,利用号码标识进行分门别类的系统揭示与组织。在分类语言中,用来代表类目名称的标识符号称为分类号,分类号是现代文献分类法的重要组成部分。文献的分类号标识文献信息主题内容及形式特征,是提供分类排架和组织分类目录的依据。从文献分类号可以得知该文献的学科属性、内容实质。例如,文学用“I”字母来标识,中国小说用“I24”来标识。

② 主题词:它是一种以词语为检索标识,采用词语揭示和描述文献信息主题内容,按照主题词字顺组织文献的情报检索语言。主题词是文献的中心思想和所要表达的主要内容的高度概括。

3. 文献的类型

文献种类多种多样,是因为划分文献的标准多种多样。例如,按文献载体形态分为纸质型文献和机读型文献,而按文献记录手段又划分为印刷型文献和非印刷型文献。

(1) 印刷型、缩微型、机读型及声像型文献。

① 印刷型文献。印刷型文献以纸张为存储介质,以手写、印刷为记录手段而产生的一种文献形式。其优点是便于阅读和传播;其缺点是存储密度太低、体积大、占用空间多,不能实现自动检索,存取速度慢。

② 缩微型文献。缩微型文献以感光材料(胶片)为存储介质,以缩微摄影技术为记录手段而产生的一种文献形式。其优点是体积小、重量轻、信息存储的密度高,借助计算机可进行自动检索;其缺点是必须借助专门的设备方能阅读。

③ 计算机可读型文献。计算机可读型文献以磁性材料和光盘为存储介质,通过编码和程序设计,把文献信息变成计算机可识别的数字语言和机器语言,然后存储在计算机的磁带、磁盘或光盘上,检索时再由计算机输出的文献。其优点是信息存储的密度大,输入输出方便快捷,可实现自动检索;其缺点是硬件设备费用较高。

④ 声像型文献。声像型文献是一种非文字的文献,以磁性材料或感光材料为存储介质,借助专用机械装置直接记录声音信息和图像信息而产生的文献。其优点是直观、真切;其缺点是阅读、利用时要借助专用设备。

(2) 一次文献、二次文献及三次文献。

① 一次文献。一次文献也称为原始文献,是作者直接记录生产或科研成果的总结和理论探讨的原始文章,如科技论文、专著、研究报告、专利说明书等。确定一篇文献是否为一次文献,主要看其内容是否为原始创作,而不以其存储介质和存储方式来确定。

② 二次文献。二次文献是根据一次文献的外表特征和内容特征,经过加工整理,简化组合,使之系统化并具备检索功能的文献,包括印刷型的目录、题录、文摘和用于文献检索信息的机读数据库,统称为检索工具。二次文献是检索一次文献的工具。

③ 三次文献。三次文献是指借助二次文献广泛地选用一次文献的内容,经综合分析、比较研究而编写的文献,如专题述评、动态综述、进展报告、数据手册、百科全书、文献指南等。

4. 文献的出版类型

按文献的出版形式可将文献分为三大类型。其具体类型结构如图 1-2 所示。

(1) 图书。在各种文献中,图书的品种最多、数量最大、范围最广,包括专著、教科书、参考工具书(辞书、年鉴、手册、百科全书、名录、图谱、图录、年表、历表等)和检索工具书(目录、题录、文摘等)。图书一般给人们提供系统性、完整性、连续性的知识和信息。

(2) 期刊。期刊是一种采用统一名称,定期或不定期出版的一种连续出版物。它具有出版周期短、报道速度快、数量大、信息内容新等特点,是很重要的信息源。

(3) 报纸。这是社会信息传播的一种重要文献,也是很重要的信息源。其特点是报道及时,受众面广,具有群众性和通俗性。

(4) 会议文献。会议文献是指在各种范围的学术会议上发表的论文及与会议相关的文献。其特点是反映新技术、新成果快,专业性强,能较全面地反映某一技术领域或学科的发展水平、动态和趋势。

(5) 科技报告。科技报告是指科技人员在从事某项科研工作时撰写的阶段研究报告或研究成果的正式报告。它具有单独成册、出版日期不定、内容专深、报道迅速、内容保密、控制发行等特点。

(6) 标准文献。标准文献是指在工、农业产品和工程建设的质量、规格、检验方法等方面做出的技术规定,是从事生产、建设的一个共同技术依据和准则。它是一种具有约束力的规定性、法律性文献。

(7) 专利文献。专利文献是指一切与专利制度有关的文献,主体是发明说明书和专利说明书。专利文献有统一的格式,是一种法律文件。它具有详尽、新颖、实用、报道及时等特点。

(8) 学位论文。学位论文是指高等学校或科研单位的毕业生为取得某种学位资格而撰写的论文。一般有学士论文、硕士论文和博士论文之分。其内容较为系统,具有一定的深度

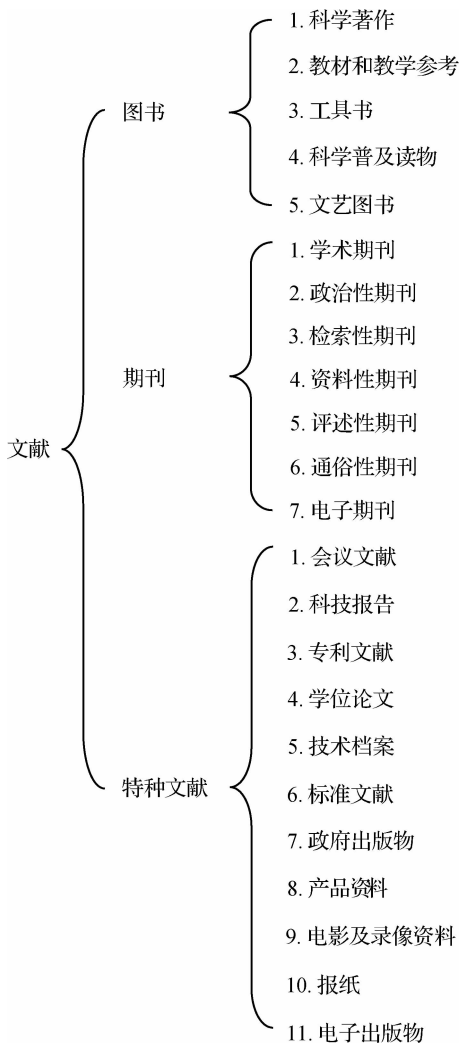


图 1-2 文献出版类型结构

和创造性。

(9) 技术档案。技术档案是指在生产建设和技术部门的技术活动中形成的,有一定工程对象的技术文件的总称,如任务书、计划书、审批文件、技术措施、设计图纸、工艺卡片等应入档的技术资料。

(10) 电子文献。电子文献是指以数字、代码方式,将图、文、声、像等信息存储在磁、光、电介质上,通过计算机或者类似功能的设备阅读使用,并可复制发行的一种文献,可分为磁盘型、光盘型、网络型文献。图书、期刊、报纸、检索工具等都可以制作成电子文献。电子文献既可以单机使用,也可以通过网络和联机的形式阅读。它具有存储量大,体积小,便于检索、保管和共享的特点。各图书馆非常重视电子文献的收藏和利用。电子文献有着十分广阔的发展前景。

1.3 科研活动和信息检索

信息技术的高速发展和广泛应用引发了全球性的信息革命,信息被看成一种战略资源,与物质、能源一起构成了现代社会的三大支柱。整个社会已形成一股强烈的信息浪潮,无论是政府、企业还是个人,都对信息给予高度重视,我国国家信息化建设的逐步开展,更使人们感受到信息时代的气息。信息时代的到来和席卷全球的国际互联网的迅猛发展,给信息利用者带来了深远的影响。如何从浩瀚的信息海洋中获取所需的信息,已成为科技人员科研工作中的首要问题。

1.3.1 科研活动概述

1. 科学研究过程

科学研究的本质是人类的一种认识和实践活动。从总体上说,人类的认识都遵循着“实践-认识-再实践-再认识……”的发展规律。科学研究同样如此,它以科学实践作为基础、出发点和检验标准,因此,科学研究的过程是“科学实践-科学认识-科学实践-科学认识……”这样一个循环往复以至于无穷的过程。科学实践作为人类以获得科学知识为目标的实践活动,是一种有意识的理性设计活动,它难以与其他按既定规程运作的物质生产实践过程一样,完全按规定的目的和计划进行,而总是在不断地“试错与改错”中取得进步。因此,“试错与改错”是科学实践最基本的特点,创造性、探索性和高风险性等是科学实践与人类其他实践活动的主要区别。

2. 科学研究的一般程序

科学研究是创造新知识和综合整理知识、推进科学技术向前发展的工作。因此,科学研究必须取得新的成果,是一种创造性的劳动。新的成果可能是新发现、新理论、新方法、新工艺和新材料。在现实的科学研究中,科学认识大多是从科学问题开始的,具体地表现为提出科学问题、分析科学问题、从思维和理论上对解决科学问题的途径进行初步构想,并根据科学实践实施的结果对原来提出的有关问题的答案进行修改、修正、重构、创造、创新等的过

程;科学实践则是根据科学问题的性质和研究的目的,围绕研究对象展开的有意识的实践活动,其过程具体地表现为通过科学观察、科学实验或科学调查等途径和方式,在实践中通过操作解决科学问题。因此,现实的科学研究就是一个不断面临新现象、新问题的动态过程。换句话说,科学研究是由科学问题开始到问题解决的过程。

从人们长期的科研实践来看,要提出一个创新的问题,需要站在前人研究成果的基础上发现问题、寻找解决问题的突破口,形成这个能力的最基本的前提能力,就是信息检索与应用能力,这个能力不是单纯的查阅文献的能力,而是一种在一定知识面扩延程度上能够高效地获得可用信息的能力,不限定专业,而是很广泛地进行想法可行性的论证,这种能力归纳起来就是应该具备良好的信息素质。在当今信息社会中,准确、快速地获取信息的能力,对科研人员是至关重要的。

1.3.2 科研活动中的信息循环和信息检索

由于信息爆炸给科研工作者知识的获取和利用带来麻烦,科技信息采集、管理和探索的现代化更具有重要意义。当下网络信息的无序、数量的巨大、类型之多,使得科技信息检索显得越来越重要。

图 1-3 所示是科学研究中信息传播循环的模式,其中环节(8)和环节(9)显示了信息检索环节在科学研究中所起的作用。科学研究的灵魂是知识创新,创新的科学研究对文献信息支撑系统提出了更高的要求——一流的科研工作需要一流的文献信息。一项科研课题无论是在立题之前还是在研究过程中,乃至在研究完成后的成果评价时,都离不开查阅有关文献资料。据统计,科研人员花费在查阅所需信息上的时间是全部时间的 30%~40%。如果没有掌握科学的信息检索方法,则花的时间会更长。

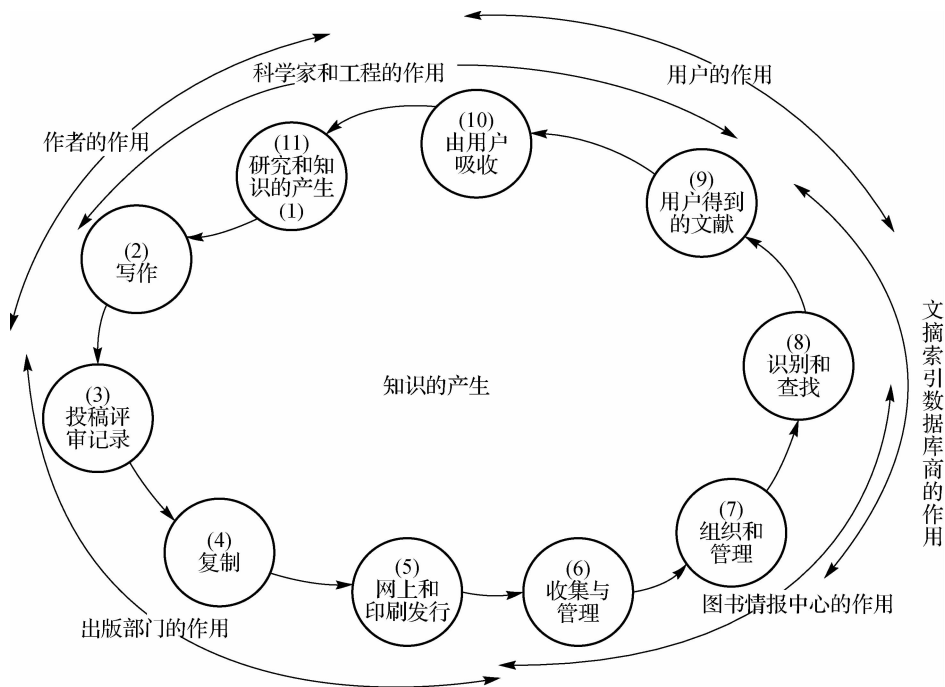


图 1-3 科学研究中信息传播循环的模式

如果在每项科学研究立项之前能够系统地检索相关的文献资料,了解当前同行在做的同样或相近的研究工作、其进展等,就能扬长避短、开拓思路,在新的起点上做出新的研究努力,取得更先进的研究成果,也可避免了人力、财力的浪费。因此,科技信息检索是科学研究中必不可少的重要环节,在科技发展中起着积极的推动作用。信息检索可以为科技成果的鉴定、评估、验收、转化、奖励等提供客观的文献依据,保证科技成果鉴定、评估、验收、转化、奖励等的科学性和可靠性。

科研所需的信息与科研活动紧密相关,包括研究生教育,获取科研信息的目的主要是减少科研活动中的“不定性”的事件描述,为科研及教学活动提供支持和帮助。科研信息有特定的用户群体,其信息需求以自然科学类为主,并具有以下特点:知识性与专业性,综合性与跨学科性,准确性、权威性与系统性,新颖性与应用性。随着网络化信息时代的发展,科研信息来源逐步向电子化、网络化方向转化。

1.3.3 信息检索的价值和地位

信息检索的重要意义体现在以下两个方面。

(1)充分利用信息资源,避免重复劳动。科学研究具有继承和创造两重性。科学研究的兩重性要求科研人员在探索未知或从事研究工作之前,应该尽可能地占有与之相关的信息,即利用信息检索的方法,充分了解国内、国外、前人和他人对拟探索或研究的问题已做过哪些工作,取得了什么成就,发展动向如何,等等。这样,才能防止重复研究,将有限的时间与精力用于创造性的研究中。

(2)为人们更新知识、实现终生学习提供途径。如上所述,在信息社会,人们需要终生学习,不断更新知识来适应社会发展的需求。美国工程教育协会曾估计,学校教育只能赋予人们所需知识的20%~25%,而75%~80%的知识是在走出学校后,在研究实践和生产实践中根据需要,不断再学习而获得的。因此,掌握信息检索的方法与技能,是形成合理知识和更新知识的重要途径。

1.4 信息社会与信息素养

随着计算机技术的快速普及,Internet热潮席卷全球,谁也无法否认人类社会正以无法估量的速度悄然迈入知识(信息)经济时代。在这个时代,经济活动对于信息(知识)活动的依赖已达到十分突出的地步,在所有可利用的资源中信息资源的利用从总体上上升到主导地位。

1.4.1 信息社会及其特征

20世纪80年代以来,由于第三次信息技术革命的推动、知识经济的勃兴,人们确信人类将逐步发展到信息社会阶段。回顾人类发展历程,我们必须清醒地看到,信息革命带给我们的是全方位的变革,正如一首歌中唱到的“这世界变化快,真让我不明白”,实际上当今时代

事事时时都处于快速变化之中。奈比斯特说：“趋势就像一匹马一样，比较容易向着它已经奔跑的方向骑去。”无疑，以知识经济为显著特征的信息社会已经成为社会发展的一种无法抗拒的趋势。

1. 信息社会的根本特征

我们今天所处的时代，有人称为“信息社会”，也有人称为“后工业社会”，其中“信息社会”的概念已为大多数人所认同。马丁曾经把信息社会定义为一个生活质量、社会变化和经济活动越来越多地依赖于信息及信息开发利用的社会。在这个社会里，人类生活的标准、工作和休闲的方式都明显地被信息和知识的进步所影响。许多研究者还归纳出衡量信息社会的若干标准：一是技术标准，信息技术必须是社会关键的能动力量；二是社会标准，信息有助于提高人民的生活质量，整个社会要有广泛而强烈的信息意识；三是经济标准，信息必须成为关键的经济因素，是资源，是服务活动，是流通的商品，也是就业和财富增值的源泉；四是政治标准，信息能够增强民主和自由，加强人们的各种参与和妥协；五是文化标准，信息具有文化价值。

如果用上述标准衡量当今社会，会得出一个无可争议的结论：我们的社会已处于信息社会发展阶段。由此可见，所谓信息社会，就是信息化了的社会，或者说信息社会最根本的特征就是社会信息化。它是对社会以材料、能源为基础的发展态势，向以信息、材料、能源并立的发展转变的一种描述和表征。社会信息化包括硬信息化和软信息化。硬信息化是指以信息基础设施建设为标志的信息技术在社会领域的全面推广；软信息化是指以信息服务发展为标志的信息资源与社会活动的深度融合。

我国政府非常重视社会信息化问题。继邓小平同志倡导“开发信息资源，服务四化建设”之后，江泽民同志又强调“四个现代化，哪一化也离不开信息化”。党和国家领导人如此重视社会信息化发展，是因为社会信息化具有强大功能：一是使社会发展空间高度拓展的激活功能；二是使社会发展因素充分组合的配置功能；三是使社会发展潜力全面发挥的诱导功能；四是使社会发展效益有效放大的倍增功能。

2. 信息化的基本特征

信息社会具有定性和定量的特征，本书主要从定性的角度来揭示信息社会的基本特征。

(1) 信息数量呈几何级数增长。据专家估计，20世纪40年代以来产生和累计的信息量已经大大超过了在此之前人类有史以来的所有信息量之和。19世纪以来，人类知识信息每50年增长1倍，20世纪中叶每10年增长1倍，20世纪70年代以后每5年增长1倍。

(2) 信息和信息技术广泛应用于社会生活的各个领域。信息产生于信息社会的各个角落，也服务于社会的各个方面。信息化手段的广泛应用和社会政治、经济、文化生活中的信息导向充分显示出信息的广泛存在和强大功能。“旧时王谢堂前燕，飞入寻常百姓家”，信息技术的普及大大缩短了信息流程，世界变成了“地球村落”。美国《时代》杂志曾破例把个人计算机选为年度风云人物，显示了现代信息技术和信息行为广泛社会化的极大影响。

(3) 人们的信息意识空前强化。人们已经充分认识到信息的重要功能和作用，全社会都在有效地开发利用各种信息资源。人们从更广泛、更快捷、更全面的途径中吸收和获取信息，主动搜寻和应用信息的意识也在不断加强，形成了一种高度的信息敏感性。人们对信息的渴望从来没有像今天这样强烈，人们的学习、工作和生活都离不开信息的参与。信息需求的自觉和自发显现，充分说明了社会信息意识——对信息功能的认同意识，对信息来源的选

择和开发意识,对信息内涵的同构和再生意识,都得到了充分的加强。

(4) 从事信息职业或半信息职业的人数超过从事非信息职业的人数。在西方,人们把担任技术、管理事务的人员称为白领阶层,把直接从事生产的人员称为蓝领阶层。在发达国家,20世纪80年代以后,白领阶层人数已超过蓝领阶层人数。美国被认为有60%以上的人主要与信息打交道,而不是与物质资源直接联系。现在人们又把专门从事数据管理的人员称为灰领阶层,他们更多面对的是计算机终端的显示屏。所谓信息职业,就是创造、处理、存储和分配信息的职业。其主要包括从事市场信息工作的劳动者(信息商品生产和销售)、提供市场信息的劳动者(信息加工和服务)和从事信息基础设施的劳动者(信息技术开发和操作)。

特别是到了信息社会,新的就业机会往往就是信息职业的就业机会。因为在信息社会,无论信息职业、半信息职业或是非信息职业,信息都成为职业因素中的重要组成,任何人都无法摆脱信息和信息技术的影响。

(5) 信息经济成为现实。按照西方经济学流行的观点,当今社会信息交流已成为经济发展的倍乘因子,即社会净产值=各部门物质生产(人力+资金)投入总和 \times 信息流量。信息已成为经济发展的重要因素,信息经济已成为客观现实,它建立在信息资源不断积累,信息传播技术手段越来越先进、便捷的基础之上。信息经济的基本特征是:第三产业的比重不断上升,非信息职业(现在已没有纯粹的非信息职业)让位于信息职业和半信息职业;理论、知识处于经济发展的中心地位;信息产业成为多数国家的“朝阳产业”。

1.4.2 信息检索与信息素养教育

掌握信息检索的基础知识和基本技能,对于提高信息素养具有非常重要的意义,具体表现在以下方面。

1. 利于创新人才的培养

当今社会正处于知识经济时代,科学技术飞速发展,新知识层出不穷。社会对人才的需求也已发生巨变。智力劳动者的社会需求日益紧缺。培养高素质、高智商的创新人才,已是目前我国高校教育迫在眉睫的重大使命。所谓创新人才,必然是发掘和捕捉最新科技知识信息的行家里手,因为无数创新成果表明:谁能优先吸收、获取和利用最新科技知识信息,谁就能优先占领科技发展前沿阵地,谁就能在竞争中立于不败之地。为此,开设本课程就是培养学生捕捉最新科技知识信息的情报意识,学习和掌握获取最新科技知识信息的技能,为科技创新锻炼过硬的文献信息检索本领。

2. 利于开发、利用文献信息(智力)资源

在当今的信息化社会里,文献信息迅猛膨胀,数量惊人,类型繁杂,新旧知识更新加快,人们处于浩如烟海的文献信息当中,迅速获取广博精准的文献信息越来越困难。学习和掌握文献信息检索的科学技术和方法,即可全面、迅速、精准地获取文献信息,为进入丰盛的文献信息的智能宝库提供了金钥匙。人们利用文献信息检索技术和方法,可以遨游文献信息大海。在这个知识经济的时代里,谁能优先掌握信息资源,谁就能在竞争中站在科学发展的前沿阵地。因此,我们必须要有被时代淘汰的紧迫感,树立强烈的信息意识,积极主动地获取所需文献信息。

3. 利于继承、借鉴和创新

继承、借鉴和创新是社会发展的客观规律。科学技术的发展、人类物质和精神文明的进步也不例外。正如牛顿所说：“如果说我看得远一点，是因为我站在巨人的肩膀上。”这就是继承、创新规律的生动描述。因此，我们从事学习、科研和生产实践，必须在前人或他人的理论、技术和科研成果的基础上，深入理解知识、学精技术、吸取研究的经验教训和发现尚待解决的问题等，只有这样才能制定出可行性创新方案，对已学知识提高认识，使技术更加专深，研究的起点更高。尤其在当今知识经济时代，继承、借鉴和创新的作用更加突出。例如，一项新的科技成果中，前人、他人的成果占90%，而创新的成果只占10%。为此，只有学习和掌握过硬的文献信息检索技能，才能在学习、科研时迅速准确地发现和了解前人和他人的研究成果。显而易见，只有强化文献信息意识和能力修养，才能有效地继承、借鉴和创新。

4. 利于节时、省力和高效

从事学习、科研都离不开查阅有关文献信息资料，它占据整个学习、研究的大量时间。美、日等国的调查结果表明，一个科研周期的各项工作环节所占时间为：计划思考占7.7%，查阅文献占50.9%，研究实验占32.1%，撰写论文占9.3%。

可见，为了提高查阅效率，必须选用最科学、最实用的文献信息检索方法，缩短时间，节省大量的精力，以便更好地学习。开设此课程的目的就是培养学生科学的检索技术，尤其是当今信息技术已引入文献信息检索领域，不仅要学习和掌握手工检索技能、技巧，而且要学习和掌握计算机检索技能、技巧，以实现省时、省力、高效。

5. 利于终生知识更新

世界进入知识经济时代以来，知识半衰期迅速缩短，新旧知识代谢加快。大学生迈入社会，走向新的岗位，如不及时更新知识，适应社会发展的新形势，将会成为新时代的知识盲人，落后于时代，必将被淘汰。有关专家估测：科技人员已具有的科技知识，12%是在大学获取的，87.5%是在工作岗位上通过学习与实践积累的。大学生在新的工作岗位上如果不更新知识，3~5年内已学的知识50%陈旧失效，8~10年再不更新知识，原有的知识将失效80%。为此，从长远出发，就要提高独立自主的学习能力、文献信息检索与利用的能力和不断更新知识的能力。

6. 有利于改善我们的生活质量

在日常生活中，小到买一本书，大到出国留学或买房买车这样的大额消费，如果掌握了信息检索的基本技能，就能及时准确地获取与生活有关的社会与市场信息，使我们的生活不仅更加丰富多彩，而且生活成本也会降低，从而为改善生活质量提供了极大的可能性。

7. 有利于缓解信息压力，有效解决信息污染

无处不在的信息虽然给我们的生活、学习、工作带来了许多便利，但不可否认，也带来了巨大的信息压力，尤其是无处不在的“垃圾”信息，往往使我们无所适从，干扰我们的思维，影响我们做出正确决策。掌握正确的信息检索技术，可以有效化解信息污染带来的风险，缓解巨大的信息量带来的精神压力。总之，学习信息检索知识是信息素养教育的重要内容，是培养高素质人才的基本要求。

第2章 信息检索的基本原理与方法



学习目标

1. 了解信息检索语言。
2. 掌握信息检索的原理和手段。
3. 了解信息检索的研究对象和范围。
4. 了解信息检索程序、策略和效果评价。

2.1 信息检索语言

2.1.1 检索语言在信息检索中的作用

检索语言是根据检索的需要而发明的一种专门的人工语言,是在文献检索领域中用来描述文献特征和表达信息检索提问的一种专用语言。

信息检索是信息传递过程中的一个重要环节。随着科学技术的飞速发展,科技文献量急剧增长,内容专深且互相渗透,而科研、生产、教学及行政部门对情报的需求十分迫切,尤其是在检索课题较专深的情况下,信息检索总是变得相当复杂,为了达到较满意的检索效果,就必须做出很大的努力。

检索语言与检索效果之间有着密切的关系,它在检索过程中起着极其重要的作用。信息检索的全过程包括信息的存储过程和检索过程。当存储信息时,文献标引人员首先要对各种文献进行主题分析,通过分析选出若干能代表文献主题的概念,并用信息检索语言把这些概念标引出来,然后纳入信息检索系统中。当检索信息时,信息检索人员也首先要对检索课题进行分析,并且通过分析明确其检索范围,选出若干能代表信息需要的概念,并把这些概念转换成信息检索语言,然后从信息检索系统中查找用该检索语言标引的文献,从而找到用户所需的信息。

由此可见,信息检索语言是信息检索系统的主要组成部分,是标引人员与检索人员之间沟通思想、取得一致意见的桥梁。

2.1.2 信息检索语言的组成、特点和功能

1. 信息检索语言的组成

实质上信息检索语言是表达、概括文献信息内容的概念及相互关系的概念标识系统。它可以是从自然语言中精选出来并加以规范化的一套分类号码,也可以是代表某类事物的某一方面特征的一套代码(如化合物的各种代码),用以对文献内容和信息需要进行主题标引、逻辑分类或特征描述。

信息检索语言是由词汇和语法组成的。在这里词汇指的是登录在分类表、词表中的全部标识,一个标识(分类号、检索词、代码)就是它的一个语词,而分类表及词表则是它的词典;语法是指如何创造和运用那些标识(单个标识或几个标识的组合)来正确表达文献内容和信息需要,以有效地实现信息检索的一整套规则。

2. 信息检索语言的特点

检索语言不同于自然语言,它所表述的概念只有一种解释,不允许一词多义、多词一义,而使概念的表述模棱两可。检索语言的这种单一性保证了表述概念的唯一性和标引与检索的一致性,从而使信息检索人员全面、准确、快速地检索到含有所需信息的文献。

3. 信息检索语言的功能

通过检索语言标引文献的主题概念,不仅能简明地提示文献包含的信息内容及其外表特征,而且还能将同一主题概念的文献集中在一起,使文献的存储集中化、系统化、组织化,便于进行有规律的检索。

2.1.3 检索语言的分类

目前世界上使用的检索语言有一两千种,但各种信息检索语言的基本原理是一致的,只是它们在表达各种概念及其相互关系和解决对它们提出的共同要求时所采用的方法不同,因而形成了不同的类型和语种。

(1) 信息检索语言按其构成原理可分为分类检索语言、主题检索语言和代码检索语言三大类。

(2) 信息检索语言按其包括的学科或专业范围可分为综合性语言和专业性语言。

(3) 信息检索语言按其适用范围可分为一种检索工具或检索系统专用的语言,某一图书、信息机构专用的语言,各类图书信息机构通用的语言。此外,信息检索语言可分为手工检索工具适用的语言和机器检索系统适用的语言。

(4) 信息检索语言按语种可划分为单语种语言和多语种语言。

(5) 信息检索语言按其标识的组合使用方法还可划分为先组式语言和后组式语言。

上述5种划分方法中,按构成原理划分的方法是最基本、最常用的方法,以下详细说明用这种方法划分的各种信息检索语言。

(1) 分类检索语言是用分类号来表达各种概念,将各种概念按学科性质进行分类和系统排列。分类检索语言包括体系分类语言(体系分类法)和组配分类语言(组配分类法)。体

系分类法主要应用概念划分与概括的方法。组配分类法主要应用概念分析与综合的方法。实际上,无论是前者还是后者,都采用概念划分与概括的方法建立等级体系结构,又采用概念分析与综合的方法进行组配。

(2) 主题检索语言是用语词来表达各种概念,将各种概念按字顺排列。主题检索语言包括标题词型检索语言(标题法)、单元词型检索语言(单元词法)、叙词型检索语言(叙词法)和关键词型检索语言(关键词法)等。它们可统称为主题法系统。

(3) 代码检索语言一般只就事物的某一方面特征,用某种代码系统来加以标引和排列。下面仅就几种常用语言进行介绍。

1. 体系分类语言

(1) 体系分类语言的定义。体系分类语言是一种直接体现知识分类的等级结构的标识系统。它根据一定的观点,以科学分类为基础,以文献内容的科学性质为对象,运用概念划分与概括的方法,按照知识门类的逻辑次序,从一般到具体,从简单到复杂,进行层层划分,每划分一次就产生许多类目,逐级划分就产生许多不同级别的类目。

所有不同级别的类目,层层隶属,形成一个严格有序直线性的知识门类的等级制体系。每个类目都用分类号作为标记,每个分类号都是表达特定知识概念的词汇,这些词汇即是体系分类语言。

(2) 体系分类语言的特点。

① 体系分类法具有按学科或专业集中地、系统地提示文献信息内容的功能,有利于从学科或专业角度进行全面的检索。

② 体系分类法按结构逐级划分,具有等级结构,便于扩检和缩检。

③ 体系分类法用分类号作为检索标识,不能直接表达概念;检索时,首先要将检索的主题内容转换为分类号,然后进行检索。若不熟悉分类法,转换时就可能发生误差,但用分类号检索,可不受文种的限制。

④ 分类法不宜经常修订,列举在分类表中的类目不能随时改变,因而分类法不能及时反映新的科学技术。遇到检索主题属于新兴学科或主题狭窄的课题时,有可能找不到相应的类目和分类号。

⑤ 分类法是按直线或序列编制类目,遇到跨学科的课题时,与课题相关的文献只能标引在一门学科的类目之下,因此检索时可能漏检。

⑥ 体系分类语言是一种先组式检索语言,因此,不能按检索课题的需要进行组配检索。体系分类语言除了以上特点以外,还有一个独特的优越功能,就是组织图书资料的排架,这是其他检索语言不可比拟的。

2. 标题词语言

(1) 标题词语言的定义。标题词语言是规范化了的自然语言。它以经过标准化处理的名词术语作为标识,来直接表达文献所论及或涉及的事物之间的相互关系(这种关系是借助于参照系统来间接显示的)。

(2) 标题词语言的结构及类型。

① 标题词语言结构。标题词语言的三大基本构成要素是主标题、副标题和说明语。

- 主标题。这是表达文献核心内容的主题概念,一般由经过控制的自然语言中的词、词组和短语来充当。
- 副标题。它是从一特定方面对主标题进行说明、限定,并经过规范化的词、词组进行限定,说明主标题所表达的文献主题的某个部分、某一应用方面、研究对象和方法、性质、场所以及文献类型。
- 说明语。它是用来详细说明和限定主标题、副标题的内容与方面的词、词组或短语。它由未经控制的自然语言表达,但不纳入词表正文,而是排在主、副标题之下。

② 标题词语言类型。标题词语言是按事物来集中相关文献的。因而一个标题下的副标题就可能涉及相当于分类表中多个类目的范畴。一般都将某些标题下的内容进行细分。细分方法多种多样,相应地形成了不同类型的标题形式,主要包括以下几种:

- 单级标题。仅由一个名词术语构成的标题,它可以是一个单词,称单词标题或元词标题;也可以是一个词组,称词组标题或复词标题,但都只有一级,故称单级标题。这种标题系统的优点是比较简单,缺点是专指性差,在一个专题下往往集中资料太多,不便于检索。
- 带说明语的单级标题。例如,“起重机,用于建筑施工的”“期刊,建筑工程的”等,这实际相当于一个复词标题。由于用说明语表达复杂概念比较自由,既能表达事物的特征,也能表达事物的方面,因而可以达到较高的专指度。但其缺点也是显而易见的,那就是比较冗长,排列次序不够明确。
- 多级标题。即在标题下再加标题。汉语中常用破折号将两级标题分开,破折号后的标题称为子标题,用以表示该标题所表达事物的某一“方面”。子标题之下还可以有子标题,称为次子标题;次子标题下还可以有子标题,称为再次子标题。不过,实际应用中超过三级的多级标题形式很少见,绝大多数是采用二级或三级标题,各级子标题可统称为副标题,第一级标题则称为主标题。
- 倒置标题。将复词标题中起限定作用的词倒置于被限定词之后,例如:

钢结构桥梁	桥梁,钢结构
木结构桥梁	桥梁,木结构
水泥结构桥梁	桥梁,水泥结构
石结构桥梁	桥梁,石结构

由此可见,倒置排列的结果可以将某一方面的资料集中在一起,这是其主要优点,但它不能单独构成标题系统。

- 带限定词的标题。例如,“建筑学(辞典)”“古建筑(维修)”“苹果(水果)”等。限定词一般都置于括号中,用于做形式复分,或区分两个字面相同但无法用其他方式区分的标题。显而易见,这种标题形式也不能单独构成标题系统。
- 混合标题。同时采用上述几种标题形式中的若干种构成。混合标题系统的优点是能提高检索效率,查找比较方便,但其结构复杂,编制难度较大。

(3) 标题词的书写格式。标题词若是一个词组,则有以下3种书写格式:

① 正序式。按词组的正常书写顺序书写。

② 倒序式。把词组的正常书写顺序颠倒过来,并加逗号表示。利用倒序式可将某些同族事物集中显示其种属关系,例如,“COMPUTERS, ANALOG”(模拟计算机)和“COMPUTERS, DIGITAL”(数字计算机)。

③ 并列式。用“与”的形式把两种概念组合在一起,如“METAL AND ALLOYS”。

一般来说,标题词间的相互关系,即它们所表达的各种概念之间的相互关系,可分为等同关系、等级关系、相关关系 3 种类型。我们可以用“见”和“见自”这一对参照来显示等同关系,用“参见”和“参见自”这一对参照来显示等级关系和相关关系。具体说明如下:

- 等同关系一般有标题词的不同写法、同义词、学名与俗称、新称与旧称、全称与简称、不同译名、其他等同词等。在每组等同关系的几个词(词组)中,都只能选用其中一个作为“标题词”,其余的则作为“非标题词”保留在标题系统中,并用“见”和“见自”这一对参照来标引它们与标题词之间的关系,例如:

电脑 见 电子计算机

这样一来,进行检索时,就不一定要判断清楚哪一个词是被采用的,哪一个词是不被采用的。可以从所想到的任何一个词或任何一种标题形式入手检索。如果入手检索的词或标题形式正好是不被采用的,则“见”参照就会把检索人引到正式被采用的标题词中。

- 等级关系一般包括属种关系、学科与其分支的关系、整体与部分的关系等。

等级关系实际上是广义词(泛指词)与狭义词(专指词)之间的关系。广义词是上位词,即上级标题词;狭义词是下位词,即下级标题词。在标题法中,一般只做上级标题词“参见”下级标题词的参照,而不做下级标题词“参见”上级标题词的参照。例如:

电机 参见 电动机 发电机

海洋学 参见 海洋地质学

消化系统 参见 肠胃

- 相关关系一般有交错关系,矛盾关系,因果关系,某种原理、方法、材料、设备与其某个应用方面的关系等。由于具有相关关系的两个标题词下的资料是可以互相参考、互相补充的,所以应用相互“参见”,例如:

图书馆学 参见 情报学

情报学 参见 图书馆学

(4) 标题词表。标题词表是把标题词按一定方式组织与展示的词汇表。它由编制说明、主表、副表 3 部分组成。

下面以美国“工程索引”使用的《工程标题词表》(简称 SHE)为例加以说明。《工程标题词表》是《工程索引》(EI)文摘条目编制和检索的依据(1993 年前),EI 的月刊和年刊都是按此标题表编排的。此表由美国工程情报公司编制。

SHE 由主标题词、副标题词、“见”或“参见”项及说明语 4 部分组成。词表的后面有 5 个附录,它们是:附录 1,副标题索引;附录 2,此部分附录中包括计算机辅助分类代码,主题指南和机构索引;附录 3,缩略语表;附录 4,希腊字母及发音;附录 5,美英不同拼写对照。

SHE 的正文是按主标题词的字母顺序排列的。主标题词全文大写黑体印刷,其下含数量不等的若干副标题词,副标题词的第一个字母用大写。

主、副标题词共计 14 000 多个。其具体结构举例如图 2-1 所示。

BUILDINGS ^①	402,(404) ^②
(Use for general subject ,applying code 404 if for civil defense or military purposes . Use geographical location as subheading . If significant, use heading for specific type of building . For specific types represented by main headings, see BUILDINGS/REFERENCE LIST ^③)	
Acoustics	See ARCHITECTURAL ACOUSTICS
Air Supported Structures	See ^④ INFLATABLESTRUCTURES
Aluminum Applications ^⑤	415
Arctic	
Brick	
Ceilings	
Climate Control	(Beginning 01/83)
Cion Systems	(See also ^⑥ Signal Systems)
Concrete	718

说明:①标题词;②分类代码;③说明语;④“见”项;⑤副标题词;⑥“参见”项。

图 2-1 主、副标题结构示意图

SHE 是用计算机编制的,可提供计算机检索磁带,词表中标题词右侧的 3 位数字,便是该标题相对应的计算机分类代码。

“见”(see)项可指引读者选用更切题的标题词,“参见”(see also)项用以指引读者进一步选用含义相关的其他标题词,以扩大检索范围。

说明语位于主标题词或副标题词下,置于括号内,用斜体字排印,用于说明该标题词的定义、适用范围以及使用方法等。

在 SHE 的副标题索引中,每个副标题词的右边分别标有一个符号,它们是关于该副标题的简略范围注释,具体符号及其含义如下:

- S 表示它可以作为普通副标题使用,而且是直接从主标题转过来的。
- # 表示它只能做副标题使用。
- @ 表示它与主标题词的形式不同。
- + 表示它是属于地理位置、材料或船只的名称。

SHE 是查找 EI 的主题索引和相同的辅助索引的工具,通过 SHE,可以准确了解 EI 采用的标题。

(5) 标题词语言的特点。

- ① 标题词语言标引文献,是以事物为中心,从有关该事物出发便于进行特性检索。
- ② 标题词在词表中按字顺排列,无字级性。
- ③ 标题词语言只能通过参照系统进行扩检和缩检,因此,查全率较低。

3. 关键词语言

(1) 关键词语言的定义。关键词语言是为适应主题目录、主题索引编制自动化的需要而产生的一种主题语言型检索语言,是文献量迅猛增长、信息传播速度加快的必然产物。

关键词是指在文献的标题、摘要或正文中出现的,对表达文献内容具有实质意义,能作

为检索入口的,起关键性描述作用的词汇。

关键词语言,就是针对文献中的关键词选定或抽出,不做或极少做词汇控制,按字顺排列,从而提供检索途径的一种检索语言。

(2) 关键词语言的特点。关键词语言可实现检索工具编制过程的计算机化,因而降低对编目人员知识和业务水平的要求,降低标引难度,节约人力。利用计算机自动标引文献,编制关键词索引非常迅速,从而大大缩短检索刊物的报道时差。由于这些优点,关键词语言在国外被普遍应用。关键词语言直接使用自然语言,因此用它作为信息检索语言,质量较差。

4. 单元词语言

单元词又称为元词,它是从文献中抽取出来并经过控制处理的,能表达文献主题的最小、最基本的词汇单位。它可以是一个单纯词,也可以是一个合成词。这些词具有一个共同的特点:它们在概念上不能再进一步分解,如再分解,就再也不能表达原来所代表的特定概念,从而失去检索本意。

单元词语言就是经过词汇控制的,能表达文献主题的元词集合。

元词的灵活组配性使它处理文献的能力很强,但在检索中也产生了相当的误检,因此,这种语言现在很少用,它被叙词语言所代替。

5. 叙词语言

(1) 叙词语言的定义。叙词是经过词汇控制的,在标引中用来显示文献主题,在检索中用来构成提问式的一种检索词汇。

叙词这一术语是由美国人穆尔斯在 20 世纪 40 年代末提出的。叙词语言是以受控的自然语言为语词基础,以字顺和分类系统为词汇显示的基本手段,以语词的概念组配为重要特征的一种标引和检索文献的理论方法。

叙词语言采用科技术语作为标识文献主题概念的标目。叙词语言可谓是博采各种信息检索语言之长,吸取了多种信息检索语言的原理和方法。叙词语言是一种采用规范化的单词或词组,由标引人员或读者自行组配,来表达文献(或课题)主题概念的一种后组式索引语言,或者叫做后组式的检索语言。

(2) 叙词的组配原则。叙词所用的词类主要是名词和动名词。叙词的词序都采用自然序列,尽量采用单词作为叙词,利用单词的组配来表达复合概念,使叙词语言的优越性能(多途径检索和自行扩检、缩检)得以发挥。直接采用专指性的词组是受到适当控制的,只是在用单词组配会造成错误和混乱时才采用词组。语词的组配方式有概念组配和字面组配两种。叙词的组配是概念的组配,应当严守概念组配的原则,只有使之符合概念逻辑,才具有检索意义。

概念组配有以下 3 种:

① 概念相交。时钟和收音机是同级词,谁也不属于谁,组配后得时钟收音机。两个或多个同级词组配后,由于不同概念相交,就会形成一个新概念,这个新概念是组配前各概念的下位概念。若用 A 和 B 表示两个不同概念,则概念相交可用图 2-2(a)表示,其中阴影部分表示组配后形成的新概念。

② 概念并列。两个或多个同级词组配后,由于不同概念并列,也会形成一个新概念,例

如,长篇小说和历史小说这两个概念并列后,得到长篇历史小说。和概念相交不同,概念并列后的新概念是组配前各概念的上位概念,如图 2-2(b)所示。

③ 概念限定。刀具和热处理可以组配成刀具热处理。刀具和热处理不是同级词,刀具是事物,热处理是该事物的某一方面,称方面词,它对事物起限定作用。因此将表示事物的叙词和它的方面词组配后,由于方面词概念的限定作用,就会形成一个新概念,它能说明事物的某一方面。概念限定也可用图 2-2(a)表示。

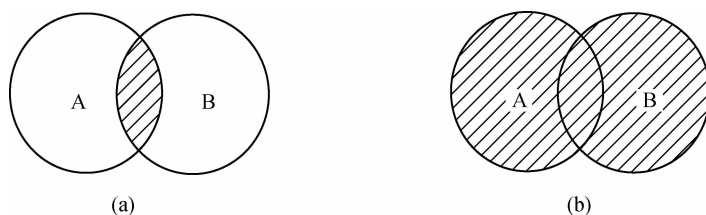


图 2-2 概念组配示意图

在各方面概念组配中,常见的是第一、第三种,特别是概念限定。因此,叙词经过概念组配后,往往能够表达更加复杂的概念,从而提高检索标识的专指性[把标识限制在图 2-2(a)中的阴影部分],以满足特性检索的需要,提高查准率。

(3) 叙词表。叙词语言所对应的叙词表是在吸收了各种词表优点的基础上发展起来的具有现代意义的词表。按其选词的学科范围,可以分为专业性叙词表和综合性叙词表。我国的《汉语主题词表》就是一部综合性叙词表。《冶金专业叙词手册》《机械工程主题表》等属于专业性词表。

叙词表一般都由一个主表和几个辅表组成。主表是叙词表的主体,包括的内容最全,对每个叙词的著录也比较完备,所以可独立存在。辅表则是为方便主表使用而编制的各种辅助索引,不能单独使用。

① 主表。主表即叙词字顺表,包括全部的叙词和非叙词;叙词条目下列有与该词有关的所有上位词、下位词、被代替词和相关词,从而可以详细了解到这类词的组成情况,同时也可以得到相关文献的“参见”项,以便扩大检索,提高查全率。

② 辅表。辅表包括叙词等级索引、叙词分类索引、叙词轮排索引。叙词等级索引按概念的等级关系排列,即将有上下从属关系的词按等级集中在一起,这样的一组词称为词族,最上位的一级词称为族首词,全部款目按族首词的字顺排列。叙词分类索引是按学科类别划分叙词,然后按类汇集,同类下的叙词按字顺排列,因此从学科角度可以查到该类的全部叙词,也可以从已知的类号转换成相应的叙词,以便从主题途径检索文献。叙词轮排索引就是将叙词中的复合词,按其包含的单词做轮番排列。一个叙词词组由几个单词构成,便可轮排几次,在几处都能查到它。

以下是由《工程索引》EI 的《工程叙词表》中摘录下来的主表的著录格式及几种受控主题词表中的参照符号对照。

《工程叙词表》的主表是由 9 400 个叙词和 9 000 多个非叙词按词的字顺排列而成,在叙词款目下列出一系列参照符号,在非叙词款目下列出 EI 的叙词。其著录格式如下。

叙词款目:Acoustic imaging

SN: Production of read-time images of the internal structure of opaque objects

DT: Ptrdates 1975

UF: Acoustic lenses

BT: Imaging technigues

NT: Ultrasonic imaging

RT: Acoustic holography

Acoustics

Diagnosis

Lenses

非叙词款目: Acoustic Lenses

Use: Acoustic imaging

其中一些符号的含义见表 2-1。

表 2-1 《工程叙词表》中符号的含义

符 号	含 义	说 明
Use	用	非正式叙词转为正式叙词
UF	代	只用于正式叙词,指出正式叙词所代替的非正式叙词
NT	分	指出所列叙词下分得更专指的词,这些词均为正式叙词
BT	属	指出所列叙词的上位概念的词
SN	注释	对所列叙词的词义进行说明
DT	日期	所列叙词起用时间
RT	参见	指出与所列叙词相关的叙词

(4) 叙词语言的特点。叙词语言是在体系分类语言、组面分类语言、标题词语言、单元词语言和关键词语言的基础上,扬长避短而发展起来的一种检索语言,它具有很多优异的性能。

① 叙词语言是后组式检索语言。概念组配是叙词语言的核心。叙词经过自由的、广泛的组配后,能形成概念网罗性高的检索标识,表达主题所含的复杂概念,从而能进行特性检索,提高查准率;能形成若干检索标识,索引剪度高,从而可以进行多途径检索。

② 叙词表的等级结构,便于增删、修订叙词。科学技术有了新的进展,增添相应的叙词后,就能标引和检索这些新科技的文献。

③ 叙词语言直接用叙词作为检索标识,不必转换,具有语言的直接性。

④ 叙词语言有一套较完整的参照系统,能显示叙词之间的相互关系,可供扩检或缩检用。

⑤ 利用叙词的词族索引和范畴索引,还可以按学科分类进行族性检索。

总之,与现有其他检索语言比较,叙词语言是一种多功能的较完善的检索语言。科技人员都熟悉自己的专业,从叙词表中找到切题的叙词后,只要组配得当,就能大大提高检索效率。

2.2 信息检索原理和手段

2.2.1 信息检索原理与类型

1. 信息检索原理

人类的信息检索行为总是从特定的信息需求开始,并在特定环境和信息检索系统中完成,这里所说的环境包括产生需求的环境、信息检索系统的运行环境和其他制约因素。特定的检索系统包括完成检索过程所需的一定设施和工具,它可以是图书馆、信息中心或信息经济人,也可以是某种工具书(如文摘索引、目录、资料集、手册、词典等)或机读信息源(如各种机读数据库)。

广义的信息检索包括信息的存储和检索两个过程。信息的存储就是将搜索到的一次信息,经过著录其特征(如题名、著者、主题词、分类号等)而形成款目,并将这些款目组织起来成为二次信息的过程。信息的检索是针对已存储好的二次信息库进行的,是存储的逆过程。存储是为了检索,而为了快速而有效的检索,就必须存储。没有存储,检索就无从谈起。这是存储与检索相辅相成、相互依存的辩证关系。

然而,由于职业、知识水平、个人素质,甚至习惯等因素的差异,信息存储人员(标引者)与信息检索用户(检索者)对同一信息的分析、理解也会存在不同。例如,《计算机在生物化学中的应用》一文,标引者可能将其归为“生物化学”类,而检索者则可能在“计算机”类查找该文。这样,标引者与检索者之间发生了标引错位,存储的信息就无法检索到。

怎样才能保证信息能存储且能取出呢?那就是存储与检索所依据的规则必须一致,也就是说,标引者与检索者必须遵守相同的标引规则。

信息存储与检索共同遵循的规则称为信息检索语言。只要标引者和检索者用同一种检索语言来标引要存入的信息特征和要查找的检索提问,使它们变成一致的标识形式,信息的存储过程与检索过程就具备了相符性。相应地,存入的文献也就可以通过信息检索工具(系统)检索出来。如果检索失败了,就要分析检索提问是否确切地描述了待查课题的主题概念,在利用检索语言标引时是否出了差错,从而导致检索提问标识错误。只有检索提问标识和信息特征标识一致时,相关的文献才能被检索出来。

信息检索正是以信息的存储与检索之间的相符性为基础的,如图 2-3 所示。如果两个过程不能相符,信息检索就失去了基础。检索不到所需的信息,存储也就失去了意义。

2. 信息检索类型

掌握了信息检索的原理,就有必要了解信息检索的类型。按照不同的标准,信息检索可划分为不同的类型,其特点也各自不同。

(1) 依检索内容划分。

① 文献信息检索。凡是利用目录、文摘或索引等二次信息查找某一课题、某一著者、某

一地域、某一结构、某一事物的有关信息以及这些信息的出处和收藏单位等,都属于文献信息检索范畴。其检索的结果是文献信息。例如,“设计人行天桥的参考文献有哪些?”便属于该类命题。

② 数据信息检索。凡是利用参考工具书、数据库等检索工具检索包含在文献中的某一数据、参数、公式或化学分子式等,都统称为数据信息检索。其检索结果为数据信息。例如,“某一新型载货汽车的载重量是多少?百公里油耗是多少?”就属于数据检索。

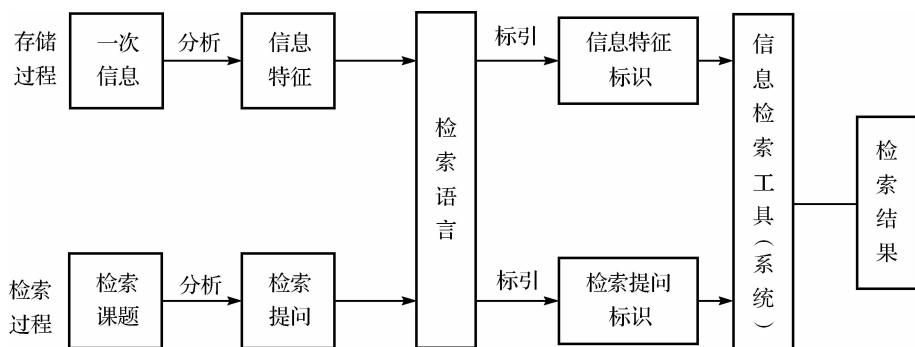


图 2-3 信息检索基本原理示意图

③ 事实信息检索。凡是利用百科全书等检索工具从存储事实的信息系统中查找出特定事实的过程,都称为事实信息检索。其检索结果是基本事实。例如,“世界上最长的斜拉桥是哪座?”“该桥位于什么地方?”“何时建成?”等。

文献信息检索是一种相关性的检索,检索的结果是文献线索,还必须进一步查找才能检索到有关的一次信息;数据与事实信息检索是一种确定性检索,检索的结果是可供用户直接利用的信息。一般情况下,文献信息检索通过二次信息来实现,而数据与事实信息检索则通过 3 次信息来完成。

(2) 按照信息存储和检索的方式和技术划分(详见 2.2.2)。

① 手工检索。手工检索是指人们通过手工的方式来存储和检索信息。

② 计算机检索。计算机检索是指人们利用数据库、计算机软件技术、计算机网络以及通信系统进行的信息存储和检索,其检索过程是在人机协同作用下完成的。计算机从其存储的大量数据中自动分拣出与用户提问相匹配的信息,用户则是整个检索方案的设计者和操纵者。其检索的本质没有发生变化,发生变化的只是信息的载体形式、检索手段、存储方式和匹配方法。

2.2.2 信息检索手段

根据信息检索的基本原理,实现信息检索的基本方式可分为传统信息检索和现代信息检索。传统信息检索简称手检;现代信息检索简称机检。

1. 传统信息检索

传统信息检索是检索人员利用手工检索工具手翻、眼看、大脑思维判别,索取原始文献的一种方式,其优点是:

① 检索条件简单,成本低。

- ② 在检索过程中可以随时获取反馈信息,及时调整检索策略。
- ③ 可对不同的检索工具同时进行对比,从而提高检索质量。
- ④ 可以参阅检索工具中的附图。

其缺点是:

- ① 速度慢、效率低,检出的文献款目必须抄录。
- ② 手工检索工具提供的检索点有限,很难进行多元检索。
- ③ 难以找到涉及几个概念组合的多主题的文献。

手检的方法有以下两种:

(1) 直接检索。这是常用的检索方法,即利用一次文献进行检索。人们常常在需要信息资料时,直接到图书馆查阅各种图书、期刊及其他资料,从中找到所需的信息。

(2) 间接检索。间接检索是利用各种检索工具(目录、题录、文摘)查找信息的方法。这是文献检索最常用的方法。

2. 现代信息检索

现代信息检索是检索人员利用计算机检索系统查找文献的一种检索方式。计算机检索系统包括数据库技术、计算机技术和通信技术。机检可以克服手检的缺点,但机检对设备条件的要求比较高,所需的投资比较大。计算机检索已从单机检索、联机检索发展到今天的网络检索,并正向智能化的方向发展。

(1) 单机检索。单机检索是计算机检索的最初形式,随着计算机存储介质的发展变化,也在不断发生变化。目前主要有计算机磁盘检索、计算机磁带检索和计算机光盘检索3种形式。

(2) 网络检索。网络检索是近十几年发展起来的,它目前主要有图书馆局域网的检索、联机检索及 Internet 信息检索3种形式。

3. 传统信息检索与现代信息检索的比较

现代信息检索是在传统信息检索的基础上发展起来的,因此两者的检索原理相同,但在网上的信息检索技术与以往的计算机检索技术又有很大的差别,主要有以下几方面的异同点:

(1) 检索语言方面。检索语言就是在检索信息时使用的专门的语言词汇,手工检索与计算机检索使用的检索语言基本相同,不同的是手工检索使用规范语言较多,而计算机检索则可以使用较多的自然语言,Internet 检索一般使用自然语言。

(2) 概念的组配方面。手工检索无检索运算符,检索时各种概念之间用人脑进行组配,用计算机检索时,则是利用各种逻辑运算符进行概念的组配,网上检索概念组配方式还不成熟,目前一般只用布尔逻辑运算符进行组配。其中计算机检索的概念组配最严谨,手工检索的概念组配最灵活。

(3) 检索入口方面。检索入口就是手工检索工具中的检索途径,是根据信息的某种特征编制的索引进行的检索,如题名、著者等。手工检索的检索入口较少,而计算机检索和网络检索的检索入口较多而且非常灵活,计算机检索过程中一个信息的所有特征都能作为检索入口,如年代、语种、文献类型、各种代号等,检索时还可以多途径同时进行检索。

(4) 检索结果方面。手工检索的检索结果查准率最高,误检率最低,而查全率较低。计

计算机检索的结果查准率要低于手工检索,误检率要高于手工检索,查全率则较高。Internet检索结果误检率和查全率也较高。

2.3 信息检索研究范围与对象

信息检索作为一门学科,有它自己的研究范围和对象,也有自己的理论、方法和技术。从总体上看,信息检索的研究对象是比较明确具体的,研究范围广泛而边界有些模糊,理论和方法已经逐渐形成体系。

2.3.1 信息检索的研究范围

信息检索的研究范围包括一切与信息存储与检索有关的系统、过程、理论和方法。一切可供存储和检索利用的信息类型,如文献、数据、事实、知识、声音、图形等;各种信息检索系统及其运行过程,如信息采集、标引、组织、存储、处理、匹配、传送等各种过程中使用的方法;以及在信息检索实践和研究的基础上形成的各种理论和假设,均包括在这个范围内。信息类型侧重于文献,其次是数据和事实。

2.3.2 信息检索的研究对象

信息检索的研究对象主要有以下几方面:

(1) 信息检索理论。它主要包括检索语言与标引理论、信息检索的数学模型、知识表示理论、相关性理论,以及有关的哲学问题。

(2) 信息检索系统。它是实现信息检索的物质基础,是现实的研究对象,主要研究其结构、功能、演变,其设计开发技术、管理维护技术和评价技术,还研究它与其他信息系统乃至整个外部世界的关系。其中,数据库是信息检索系统的核心部分之一。数据库的建造和维护是非常重要的信息技术。数据库的质量直接影响信息检索系统的功能和效率。

(3) 计算机信息检索。它涉及许多计算机设备、软件技术、存储技术革新、检索技术、系统管理和经营知识、市场营销技术等,是一门综合性很强的技术。

(4) 检索策略与方法。它是用户从信息检索系统中获取有关信息所必需的。好的检索策略是检索成功的必要前提,计算机的应用为检索策略和方法的发展提供了有力的支持。近30年来,涌现了许多新颖而有效的检索技术和方法,如布尔检索、位置检索、截词检索、加权检索、聚类检索等,人们利用、研究、评价和完善现有的各种检索策略和方法,研究开发新的更有效的策略和方法。

(5) 用户研究与培训。用户是信息检索系统的生命,无论是系统的研制开发、管理维护、功能和服务的扩展,还是系统评价,都离不开用户研究工作。用户培训是用户研究的继续,是与用户建立紧密联系和发展新用户的一种非常有效的措施。此外,还有自动标引、自动分类和自动摘录,以及相关设备等研究领域。

2.4 信息检索方法与程序

信息检索是根据课题的要求,利用各种有关的检索工具,按照一定的检索程序和方法,采用各种检索策略查找信息的过程。

2.4.1 信息检索方法

常用的检索方法如下:

1. 追溯法

追溯法是指在没有检索工具的情况下,人们利用书、刊之后的引用或参考文献查找资料的方法。

2. 常用法

常用法这是指利用各种类型检索工具,通过技巧查找资料的方法。常用法又可分为顺查和倒查两种。

顺查法:即由远而近的查找方法。如已知某研究课题的起始年代,需要了解它的全面发展情况时,可从最初研究的年代开始,由远及近地进行查找。

倒查法:也称为叫逆查法。和顺查法相反,倒查法是由近及远的查找方法,这种方法常用于正在研究课题的过程中。

3. 循环法

循环法又称为分段法,这是前两种方法的结合,即首先利用检索工具查找一部分资料以后,为了了解其前因后果,再利用这部分资料的引用文继续追查的方法,利用这种方法能够保证所查的资料较全面、系统。

2.4.2 信息检索程序

查找信息的过程如下:

1. 分析课题

信息检索之际,首先要对课题进行充分的了解,掌握课题有关的基本知识,详细摸清课题所需资料范围及要求,以及弄清课题有关的名词、术语等。

2. 选择检索方法

根据检索人员已掌握的资料情况,确定采用何种检索方法。如果是新立的课题,一般采用顺查法和追溯法结合查找,正在进行中的课题一般用倒查法。

3. 确定检索途径

在检索过程中,根据所掌握的有关课题的资料,确定检索途径。一般情况下,所研究的课题较专深、单一时,常用主题途径;课题所涉及的内容较复杂时,常用分类法。

4. 查找信息线索

在确定检索途径之后,根据检索标识,通过有关索引进行信息检索,查找信息线索。

5. 索取原始信息

通过检索获得信息线索后,确认有一定的参考价值,需要详细查阅信息的全文时,即可根据文摘指出的信息出处,向有关部门索取原始信息。

2.5 信息检索策略

信息检索总是根据一定的课题进行,但并不是有关该课题内容的全部信息都需要。要想在数量庞大、类型复杂、内容专深、文种多样的信息海洋中迅速、准确地查找到满足要求的信息,必须制定查找信息的策略。

下面仅介绍手工检索策略的制定方法。

1. 先国内后国外

首先利用国内检索工具或用中文编辑的报道外文的检索工具入手检索。这样,检索者从自己熟悉的文字可以很容易地掌握课题的概况,并可以从中选择出切题的关键词、主题词,为进一步查找外文检索工具提供方便。

2. 检索时多选几个同义词、近义词

信息语言复杂多变,作者使用的语词不尽相同,为了避免漏检,检索时尽量多选几个同义词、近义词作为检索词。

3. 巧妙地利用上、下位词的关系

当要求查全率高而不计查准率时,可利用上、下位词进行检索。

4. 尽量避免从字面出发选择检索词

在检索时,尽量从内容上选词,以避免有关问题因选词不当被漏检。

5. 变换检索词的词序

某些课题,在检索中反复查找而没有结论时,应考虑倒置词序的可能。检索词序的准确与否是检索成败的关键。

6. 利用参见系统,扩大检索范围

检索工具中的参见(see also)项,把内容有关的主题联为一体,以便族性检索。例如,CA中每一大类末尾均列有与该类有关的全部相关大类和具体相关文摘。绝大多数检索工具中的主题词表、主题索引或文摘部分中均没有参见项,检索时分别查找各有关的标题,以获取某一课题的全部信息。

7. 各种索引配合使用

要想把各种索引配合使用,首先熟悉各种索引的结构、特点、款目内容及其功能。例如,

CA中的化学物质索引与分子式索引、登记号索引、环系索引、杂原子索引关系十分密切。如果从化合物登记号索引入手检索,只能获得化合物的名称与分子式而查不出文摘号,必须转查化学物质索引才能获得信息线索。又例如,对商品名称与俗称,必须首先查索引指南。类似的情况在检索时常常遇到,只有熟练掌握各种检索工具的结构体例、编排特征、各索引款目结构及其相互关系,检索起来才能得心应手,取得满意的效果。

2.6 信息检索结果的评价方法

利用检索工具或检索系统查找信息时,检索结果并不是在任何情况下都能如愿以偿的。其主要原因是:从目前的检索系统看,它只能处理信息标识,而信息的标识只能表示信息的中心主题,不表示信息的次要主题。即系统只能回答信息提问而不一定能满足信息需求。从检索者来看,一般只能大概地表述自己的信息提问,而不能确切地衡量信息与提问的切题程度。再加上,检索课题的性质、范围、深度不断变化,查找信息的方法尚未形成固定模式,补充词表的出版不及时,人们不可避免地受到知识面的局限等,致使信息标引与检索失误在所难免。

在目前情况下,对检索结果的分析评价方法有以下几种。

1. 查全率

查全率是对所需信息被检出程度的量度。

2. 查准率

查准率是衡量检索系统拒绝非相关信息的能力。

查全率和查准率均可以用公式来表示:

$$\text{查全率}(R) = \frac{a}{a+c} \times 100\%$$

$$\text{查准率}(P) = \frac{a}{a+b} \times 100\%$$

式中, a 表示被检出的相关信息, b 表示被检出的非相关信息, c 表示未检出的相关信息。

除此之外, $c/(a+c) \times 100\%$ 为漏检率,是查全率的补数; $b/(a+b) \times 100\%$ 为误检率,是查准率的补数。

一般来说,查准率与查全率之间存在反变关系。当某一课题的查全率与查准率处于某一比例关系时,继续提高查全率,查准率会降低。

3. 相关性

相关性是用户判断信息与信息需求之间关系的标准。信息系统回答的不是用户提问本身,而是检索式表达的信息提问。因此,检出的是与信息提问相关的信息,但不一定是对用户切题的信息。

第3章 信息检索工具



学习目标

1. 了解检索工具的概念与类型。
2. 了解检索工具的排检方法。
3. 了解检索工具的相关知识。
4. 掌握检索工具的使用方法、检索程序和途径。

3.1 检索工具的含义

检索工具是用来报道、存储、查找信息、汇集各种信息,按特定的方法加以描述和编排,以供查考的工具。检索工具与普通文献的主要区别在于:它是一种工具书刊,是专供人们查找特定信息的,虽然它也具备可读性,但它不是供人们进行系统阅读的。

3.2 检索工具的类型

1. 按检索工具的出版形式划分

(1) 期刊式检索工具。期刊式文献信息检索工具有着一般科技期刊出版物的特点,有固定统一的刊名,以年、卷、期为单位,为各种定期或不定期出版的检索期刊。它的优点是迅速、及时、系统、完整,便于回溯检索,是进行近期或早期文献信息检索的较为主要的工具。例如,以书本形式出版的期刊《全国报刊索引》就属于这种类型。

(2) 书本式检索工具。书本式检索工具具有以下优点:可以一次编成,印刷分发,便于长久使用,实用价值高;体积小易于携带,使用不受时间空间限制,便于馆际互借。

书本式检索工具的不足之处是:出版周期长,不能随时反映最新信息;查阅不方便,书本式检索工具落后于实际发展;陈旧过时的信息无法从书本式检索工具中剔除。

(3) 附录式检索工具。附录式文献信息检索工具一般不单独出版,系指附在书刊后面的参考文献和引用文献目录。这种检索工具比较分散,不引以关注,利用它们往往查不全所需文献线索。但是,其专业性强,参考价值大。

(4) 卡片式检索工具。卡片式检索工具具有以下优点:可以随时编辑、随时排序、随时使用,具有流水作业的性质。

其不足之处是:体积庞大,占用较大空间,排序不易,检索点少。这种工具现在基本被淘汰了。

(5) 缩微式检索工具。缩微式检索工具的优点是体积小、编印速度快、发行广泛;缺点是不能随时增减,阅读需要专用阅读器,使用不方便。

(6) 机读式检索工具。机读式检索工具的优点是:一次输入多次检索,检索速度快,质量高,编目自动化,利于集中统一编目,而且可以利用各种现代化通信设备形成检索网络,实现联机和网络检索。

2. 按检索工具的基本作用划分

(1) 目录型。一切可供检索使用的目录均可称为目录型检索工具。它主要报道信息的外表特征,著录具体出版事宜及其收藏信息。常见的目录按其收录范围可以分为国家目录、联合目录、馆藏目录;按报道的范围分为综合性目录和专题性目录。

(2) 参考型。这是分析和著录大量具体而常用的科学数据库与事实,以备查用的各种常用工具书的总称。其中包括字典、词典、年鉴、百科全书、指南、手册、名录等。

(3) 综合型。该类工具是收录多种文献(如图书、期刊、科技报告、会议文献等)和多种学科的文献信息检索工具(如《化学文摘》《工程索引》等)。

(4) 搜索引擎。搜索引擎是将网络信息分门别类地组织起来,通过搜索网址的方式来实现信息检索的工具。它是一种网络检索工具,检索到的可以是一般的信息线索、原始信息全文、一般的文本信息和多媒体信息。

3. 按检索工具的内容与功用划分

(1) 目录。目录是著录一批相关的出版物,并按一定方式编排的一种揭示与报道信息的工具,包括国家书目、在版目录、期刊目录、收藏目录、联合目录、推荐书目、专题书目。

(2) 索引。索引是将出版物中只有检索意义的信息特征(篇名、语词、人名、地名等),按一定的顺序编排组织起来,以供检索之用的工具。其分为篇目索引和内容索引(语词索引、主题索引、人名索引、地名索引、关键词索引)。

(3) 文摘。文摘是索引的延伸。

(4) 辞典。辞典是字典、词典的总称。

① 字典是解释字的形、音、义及其用法的工具。

② 词典是说明语词的定义、意义和用法的工具。

(5) 百科全书。百科全书是记载人类一切门类的或某一门类的知识,以辞典的形式编纂的系统而完备的检索工具。

(6) 年鉴。年鉴是汇集一年内的重要时事和统计等信息,按年度连续出版的检索工具。

(7) 手册。手册是汇集某一领域中最经常使用的信息以备查检的工具。

(8) 名录。名录亦称机构名录、机构指南、便览、大全,是一种系统编排的、有关个人和组织机构的目录。

(9) 表谱。表谱是以编年或表格形式记载事物发展的检索工具。其主要有年表、历表和其他表谱,主要用于查检不同纪年方法的年代对照,不同历法的年月日对照以及历史大事、人物生卒年、地理沿革等信息。

(10) 图谱。图谱主要指地图和图录。

① 地图是根据一定的数学法则和光学技术,将地球表面的自然和社会现象,以各种不同的图式符号在平面纸上缩小成概括的图形,包括普通地图、历史地图、专题地图。

② 图录是以图像表示事物形象的工具,包括历史图录、人物图录、艺术图录。

(11) 类书。类书是采集各种古籍中有关事物的资料,将它分类编排,以供查阅的工具。类书是中国古代的百科全书。类书主要可用来检索中国古代的诗赋文章、丽词骈语、史实典故、名物制度等百科资料,如《太平御览》《册府元龟》《太平广记》《文苑英华》《永乐大典》《古今图书集成》《佩文韵府》《骈字类编》。

(12) 政书。政书中国古代专门记载典章制度的书,它记载典章制度的基本内容及其沿革变化和其他史实。如《通典》《通志》《文献通考》《续通典》《续通志》《续文献通考》《清朝通典》《清朝通志》《清朝文献通考》《清朝续文献通考》,合称为“十通”。

4. 按检索工具的收录范围划分

按收录范围,检索工具可以分为综合性检索工具、专业性检索工具和单一性检索工具。

(1) 综合性检索工具。综合性检索工具收录的学科范围较广,涉及的信息种类和语种也比较多。例如,美国的《工程索引》。

(2) 专业性检索工具。专业性检索工具收录的学科范围比较窄,仅限于报道某学科或专业的信息,但收录该学科信息的类型比较全。例如,美国的《化学文摘》。

(3) 单一性检索工具。单一性检索工具收录的文献类型比较单一,文献的著录格式标准、规范。例如,《中国专利公报》就是典型的单一性检索工具,它仅仅收录单一的文献类型——专利文献,著录格式简单规范,易于查找。

3.3 检索工具的排检方法

3.3.1 按字顺编排

这里只介绍中文字顺排检法。中文字顺排检法主要有以下几种。

1. 音序法

(1) 汉语拼音排检法。其优点:原理简单、查检迅速;科学性强,符合国际上检索工具的

编排规则,是一种理想的中文排检法。其缺点:如果不知道一个字正确读音时,就难以查到。

(2) 注音字母排检法。注音字母是由汉字偏旁改造而成的40个字母,其中声母24个、韵母16个。它以北京语音为标准音,1958年前后出版的字典(词)典大多数都以注音字母注明其读音,如《新华字典》1956年、1959年版和《同音字典》等。

(3) 声部韵部排检法。声部韵部排检法也称“声韵法”,是我国古代按音韵排列汉字的一种方法。按韵部编排的字典称为“韵书”。我国在不同的历史时期有不同的韵部。宋代《广韵》分206个韵部,为《集韵》所沿用。自宋以来流行的平水韵(《平水新刊礼部韵略》)将其删并为106韵,从而成为文人作诗用韵的规范。一些主要工具书也按平水韵编排,如清代编的《佩文韵府》《经籍纂诂》以及现代编的《辞通》等都按词目的末字分韵编排。《九史同姓名略》则按条目首字分韵编排。

利用韵部编排的工具书来查字词,可通过新印本所附的索引先查出该字的韵部后,再按韵部去查。如新印本《佩文韵府》和《辞通》都编有首字四角号码和笔画索引。亦可先利用有关的字典查出该字的韵部后,间接来查。

2. 形序法

(1) 笔画法。优点是:原理简单,易于掌握,且适合汉字的特点。局限性是:同一笔画容纳的汉字较多,不便很快查到要查检的汉字;由于汉字结构的多样化,简体与繁体、规范书写与手写体的差别很大,每个字的笔画不易计算准确。应用:不仅可以用来编排检索工具的正文,也可以用来编制检索工具的索引,它还是一种常见的补充编排法,音序法、部首法、笔形法等往往用它对条目进一步编排。

(2) 部首法。优点是:部首法符合汉字的结构和人们的查找习惯,不认识的字念不出来,不能用音序查检时,可以根据字形用部首法查出来;部首法能够容纳数量庞大的汉字,是编制大型语言工具的必要排检法。局限:部首的位置不固定,判断一个字的部首必须遵循固定的原则和规定,比较难掌握;不同检索工具中的部首数量往往不同。

3. 号码法

(1) 四角号码法。其基本规则包括笔形规则和取角规则。笔形规则即确定用什么号码来代替何种笔形的规则;取角规则是关于汉字四角取笔的一些特殊问题的具体规定。优点是:四角号码法是一种单一的排检法,不必与其他方法相结合,只要熟记10种笔形所代表的号码,见字知号,查检迅速。缺点是:四角号码法的局限在于不易准确掌握。

(2) 中国字度撮(guǐ xiè)法。此法类似于四角号码法,但不十分科学,规则更复杂,掌握很难,已被淘汰。哈佛燕京学社运用它编了64种引得(索引),多为古籍整理类的线索性工具书,但也附上了笔画法、四角号码法和汉语拼音法。

4. 起笔笔形代码法

起笔笔形代码法按汉字起笔笔形进行字顺排列的方法和规则。即将汉字起笔笔形分为横、竖、点、撇、折5种,对5种笔形进行编号,将一个汉字的前5个笔形转换成笔形编码,然后按编码排序。

(1) 应用。起笔笔形法一般不用来编排检索工具的正文,多数检索工具仅用它作为其他编排方法的一种补充手段,是其他编排方法的附属方法。

(2) 优点。起笔笔形法强调起笔,位置固定。

(3) 缺点。每一类所统属的条目多,检索不便。

3.3.2 按内容性质编排

1. 分类编排

分类排检法是按信息内容的学科属性和逻辑次序编排检索工具的一种方法。分类排检法大致可分为按分类法编排和按自编的分类体系编排两类。

使用分类排检法便于按类检索信息,并且能够较全面地获得同类的相关信息,有利于检索。

【例 3-1】 Yahoo 的类目体系:Yahoo 是以网络资源为对象的指南型分类检索工具,它设置了 14 个基本大类,按字顺排列为艺术与人文、商业与经济、计算机与网络、教育、娱乐、政府、健康与医药、新闻与媒体、休闲与运动、参考资料、地区、科学、社会科学、社会与文化。

优点:从类目设置的角度看,基本放弃了文献分类法以学科为中心建立类目体系的传统,采用以主题为中心并与学科结合的设类方式,可以使其在具有直接性的同时增加包容性;类目体系的展开,采用了多重划分、多元展开的类目结构,收录有关该类的所有信息资源,使用户可以同时从不同角度出发进行查找;采用链接的方式充分揭示横向联系。

缺点:类目的展开缺乏统一的标准,随意性大,无规律可循;同位类中包括不同等级的类目;按照类名字顺排序不能反映并列类目之间的联系,缺乏对知识门类系统显示的能力。

【例 3-2】 Sohu 分类目录的类目体系:娱乐休闲、工商经济、公司企业、文学、体育健身、卫生健康、生活服务、社会文化、社会科学、国家地区、电脑网络、教育培训、艺术、新闻媒介、科学技术、旅游交通、政法军事、个人主页。

2. 主题编排

主题排检法是以主题词来揭示、标引和排列信息的一种方法。用主题法编排检索工具,其过程分两步:一是用能够描述表达信息内容的主题词作为条目的标识,二是将所有条目的标识按字顺组织起来。

主题排检法能够把不同学科、不同知识体系中的同一主题的信息集中在一起,有利于人们按主题检索,但使用主题编排的检索工具必须正确选定主题词。

主题编排与分类编排的区别如下:

(1) 主题法直接用主题词来描述、标引和组织信息,强调直观性;分类法则按知识体系分类归并信息,强调系统性。

(2) 在主题排检法中,主题词既揭示信息的内容,又是用来标引和检索的词汇;而在分类排检中,类目只是对信息条目学科属性的概括,而不是标引、检索的词汇。

(3) 主题排检法需要借助于字顺排检法来编排组织主题词的顺序;而分类排检法本身不需要借助于其他方法,它按信息内容逐级分类。

3.4 常用的中文检索工具

3.4.1 书刊、论文和成果检索工具

1. 中文检索工具概况

中文检索工具专业面广,无语言障碍,多为题录、简介、文摘的结合形式,著录格式规范(采用国家标准《检索期刊条目著录规则》),正文多按分类编排,主题索引采用《汉语主题词表》或专业主题词表,原文容易获取。但许多检索工具索引不齐全,给使用带来不便。

2. 书刊、论文、成果检索工具选介

(1)《全国新书目》和《全国总书目》。由中国版本图书馆编辑出版,1949年创办。《全国新书目》为月刊,反映我国近期图书出版情况。《全国总书目》是《全国新书目》的累积本,一般一年出一册。按规定,各出版社出版新书必须向中国版本图书馆呈缴样本,该书目根据该馆馆藏的新书样本编辑出版,所以,它较全面地反映了我国图书出版的基本情况,是检索国内图书(包括公开发行,限国内发行的初版和改版图书)的主要工具。

著录格式:

书名/著者. —版本. —出版地:出版者,出版时间. —页码;尺寸. —ISBN:价格

著录格式举例如下:

①TU^②建筑科学^③室内设计简明教程/^④庞瑾,赵彬编著. —^⑤北京:^⑥电子工业出版社,^⑦2002.10^⑧. —331页;^⑨26cm^⑩. —ISBN7-5053-8061-3:^⑪¥38.00

注释说明:①分类号,②类目名称,③书名,④著者,⑤出版地,⑥出版单位,⑦出版时间,⑧总页数,⑨书的载体形态,⑩国际标准书号,⑪价格。

《全国总书目》有分类目录、专题目录和附录3部分。专题目录包括少年读物目录等,附录包括报纸杂志目录、出版社一览表和书名索引。

(2)《外文新书通报》。为馆藏外文图书目录,目前有北京图书馆编辑出版的,还有中国科学院图书馆等单位编辑出版的,是检索外文图书的重要工具,反映了收藏单位在某一时期收藏的外文图书。

(3)《全国西文连续出版物联合目录》(1978—1984年)。由北京图书馆报刊联合目录编辑组编,书目文献出版社出版,收录了全国692家图书情报单位入藏的120个国家和地区的56种文字的18860种连续出版物,较全面客观地反映了1978—1984年西文连续出版物在我国的入藏情况。

(4)《全国报刊索引》。《全国报刊索引》由上海图书馆编辑出版,以题录形式按篇报道报刊上的文章,目前年报道量为40万条以上,是我国重要的综合性检索工具。

从1980年起,《全国报刊索引》分为哲社版与科技版同时出版,均为月刊。哲学社会科学版收录了国内公开和内部发行的中文期刊2000余种、中文报纸160多种。自然科学技术版收录了国内公开和内部发行的中文期刊3000余种。两种版本每期均附有本期引用期刊一览表。

近几年在题录正文后附有作者索引,按作者姓名汉语拼音音序排列,在作者姓名后提供题录号。

(5)《中国学术期刊文摘》。1994年创刊,月刊,由中国科学技术协会学会学术部、国家自然科学基金委员会计划局主办。该刊报道范围涉及自然科学、农业科学、医药科学、工程与技术科学、人文与社会科学等领域,每期附有关键词索引和作者索引。关键词索引先按类排,每类下再按汉语拼音顺序排。作者索引按作者姓名的汉语拼音顺序排列。

(6)《中国化学化工文摘》。1983年创刊,1993年由《中国化工文摘》改为现刊名,1997年由月刊改为双月刊,由中国化工信息中心编辑出版。该刊是检索全国化学化工科技信息、了解和掌握中国化学化工发展现状的主要工具。

文摘正文按分类编排,每年编有年度主题索引、作者索引、分类索引。

(7)《中国生物学文摘》。1987年创刊,月刊,由中国科学院文献情报中心、中国科学院上海文献情报中心和中国科学院生物文献情报网主办。本刊报道的内容有普通生物学、细胞学、遗传学、生理学、生物化学、生物物理学、分子生物学、生态学、古生物学、病毒学、微生物学、免疫学、植物学、动物学、昆虫学、人类学、生物工程学、药理学以及生物交叉学科与相关科学技术领域。

文摘正文按分类编排,并有主题索引。

(8)《科学技术研究成果公报》。1981年创刊,月刊,由中华人民共和国国家科学技术委员会主办、国家科委科技成果管理办公室编辑出版。该刊按类(主要为自然科学)编排,第12期后附年度分类索引,是检索我国科研成果的权威性工具。

(9)《中国学术会议文献通报》。1981年创刊,月刊,由中国科学技术信息研究所、中国农业大学主办。该刊汇集了全国各学会与协会、各部委及其所属单位召开的全国性学术会议和国际性学术会议的文献信息,按类(主要是自然科学)编排,第12期为年度主题索引,是我国目前唯一报道学术会议文献的检索刊物。

(10)《中国学位论文通报》。1985年创刊,双月刊,由中国科学技术信息研究所编辑出版。该刊专门报道国内学位论文信息,收录了我国高等院校、科研单位招收的硕士、博士研究生的毕业论文,内容包括数理科学、化学、天文学、地球科学、农、医、卫生、工程技术等各个领域。

3. 检索方法

查找文献的方法分为以下3种。

(1)直接法。直接法是指直接利用检索工具(系统)检索文献信息的方法,这是文献检索中最常用的一种方法。它又分为顺查法、倒查法和抽查法。

① 顺查法。顺查法是指按照时间的顺序,由远及近地利用检索系统进行文献信息检索的方法。这种方法能收集到某一课题的系统文献,它适用于较大课题的文献检索。例如,已知某课题的起始年代,现在需要了解其发展的全过程,就可以用顺查法从最初的年代开始,逐渐向近期查找。

② 倒查法。倒查法是由近及远,从新到旧,逆着时间的顺序利用检索工具进行文献检索的方法。此法的重点放在近期文献上。使用这种方法可以最快地获得最新资料。

③ 抽查法。抽查法是指针对项目的特点,选择有关该项目的文献信息最可能出现或最多出现的时间段,利用检索工具进行重点检索的方法。

(2) 追溯法。追溯法是指不利用一般的检索工具,而是利用已经掌握的文献末尾所列的参考文献,进行逐一地追溯查找“引文”的一种最简便的扩大情报来源的方法。它还可以从查到的“引文”中再追溯查找“引文”,像滚雪球一样,依据文献间的引用关系,获得越来越多的内容相关文献。

(3) 综合法。综合法又称为循环法,它是把直接法和追溯法加以综合运用的方法。综合法既要利用检索工具进行常规检索,又要利用文献后所附参考文献进行追溯检索,分期分段地交替使用这两种方法。即先利用检索工具(系统)检到一批文献,再以这些文献末尾的参考目录为线索进行查找,如此循环进行,直到满足要求为止。综合法兼有常用法和追溯法的优点,可以查得较为全面而准确的文献,是实际中采用较多的方法。对于查新工作中的文献检索,可以根据查新项目的性质和检索要求将上述检索方法融汇在一起,灵活处理。

4. 文献检索程序

文献检索工作是一项实践性和经验性很强的工作,对于不同的项目,可能采取不同的检索方法和程序。检索程序与检索的具体要求有密切关系,大致可分为以下几个步骤。

(1) 分析待查项目,明确主题概念。首先应分析待查项目的内容实质、所涉及的学科范围及其相互关系,明确要查证的文献内容、性质等,根据要查证的要点抽提出主题概念,明确哪些是主要概念,哪些是次要概念,并初步定出逻辑组配。

(2) 选择检索工具,确定检索策略。选择恰当的检索工具,是成功实施检索的关键。选择检索工具一定要根据待查项目的内容、性质来确定,选择的检索工具要注意其所报道的学科专业范围、所包括的语种及其所收录的文献类型等。在选择中,要以专业性检索工具为主,再通过综合性检索工具相配合。如果一种检索工具同时具有机读数据库和刊物两种形式,应以检索数据库为主,这样不仅可以提高检索效率,而且还能提高查准率和查全率。为了避免检索工具在编辑出版过程中的滞后性,还应该在必要时补充查找若干主要相关期刊的现刊,以防止漏检。

(3) 确定检索途径和检索标识。一般的检索工具都根据文献的内容特征和外部特征提供多种检索途径,除主要利用主题途径外,还应充分利用分类途径、著者途径等多方位进行补充检索,以避免单一途径不足所造成的漏检。

(4) 查找文献线索,索取原文。应用检索工具实施检索后,获得的检索结果即为文献线索,对文献线索进行整理,分析其相关程度,根据需要,可利用文献线索中提供的文献出处,索取原文。

5. 文献检索途径

检索途径也称检索入口,文献的特征是存储文献的依据,也是检索文献的依据,因此,文

献特征便构成了检索途径。一般来讲,检索途径可以分为以下4种:分类途径、主题途径、著者途径和其他途径。前3种检索途径如图3-1所示。

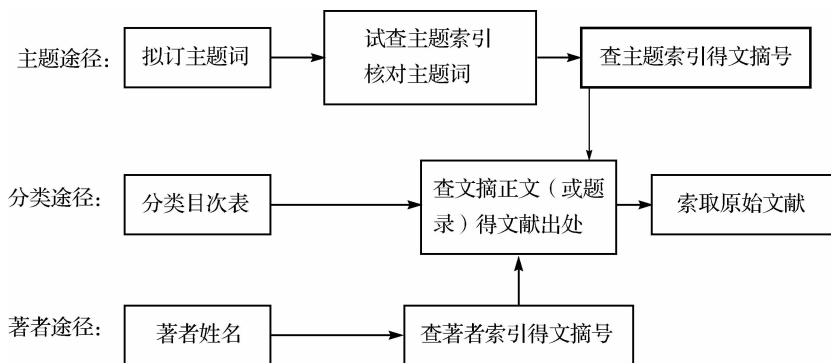


图 3-1 检索途径示意图

(1) 分类途径。按学科分类体系来检索文献。这一途径是以知识体系为中心分类排检的,因此,比较能体现学科系统性,反映学科与事物的隶属、派生与平行的关系,便于从学科所属范围查找文献资料,并且可以起到触类旁通的作用。从分类途径检索文献资料,主要是利用分类目录和分类索引。

检索关键:获取课题的分类号。

分类依据:《中国图书馆分类法》(《中图法》)、《中国图书资料分类法》(《资料法》)、专业分类法,等等。

课题的一般分类方法如下:

① 单概念课题。单概念课题指课题的主题概念只有一个。了解分类表的体系结构,掌握大类的分布;在相关大类中,由大而小逐步查找最接近的类号;注意分类表中有助于确定类号的类目注释和类目指引。

② 多概念课题。多概念课题指课题的主题概念有两个或以上。对涉及同一研究对象的几个方面或几个并列研究对象,有主次者,取其重点或主要研究对象;若无主次,则几个方面都要检索;研究理论、方法、工艺、材料、设备、产品等在某方面应用或对某方面影响的课题,应在所应用或受影响的类目查找;研究理论、方法等等多方面应用或对多方面造成影响的,在该理论、方法等本身所属类目查找。

【例 3-3】 利用分类途径检索关于网络安全技术研究的文献。

第一步:分析课题,选择检索工具。该课题是关于网络安全技术方面的研究,网络安全包括网络的系统安全和网络的信息安全,而保护网络的信息安全是最终目的。

可选择综合性检索工具《全国新书目》《全国报刊索引(自然科学技术版)》《中国学术期刊文摘》《中国学术会议文献通报》《科学技术研究成果公报》《中国学位论文通报》及专业性检索工具《电子科技文摘》等进行查找。

第二步:确定检索途径。以上检索工具可提供多种途径检索,根据课题的要求,选用分类途径检索。

例如,查《全国报刊索引》(自然科学技术版)的分类目录,有“TP 自动化技术、计算机技术”,进一步确定分类号与类名为“TP393 计算机网络”。

第三步:采用倒查法。逐年查找《全国报刊索引》(自然科学技术版),以2003年第4期为例,检索结果(选一例)为:

030414228 若干网络安全解决方案的比较与讨论/祁明(华南理工大学计算机学院,510640);李坚超;康敏//计算机应用研究.—2002,19(12).—70-72,99

第四步:根据文献线索,获取原始文献。

第五步:根据文献线索,可在《计算机应用研究》2002年第19卷第12期第70页至第72页和第99页上查得原始论文。

(2) 主题途径。通过反映文献资料内容的主题词来检索文献。由于主题法能集中反映一个主题的各方面文献资料,因而便于读者对某一问题、某一事物和对象做全面系统的专题性研究。通过主题目录或索引,即可查到同一主题的各方面文献资料。

【例 3-4】 利用主题途径检索关于网络安全技术研究的论文。

第一步:分析课题,自拟主题词。分析课题,自拟主题词“网络安全”“系统安全”“信息安全”等。

第二步:选择检索工具,确定检索途径。可选择《中国学术期刊文摘》《电子科技文摘》《中国学术会议文献通报》检索工具,根据课题的要求,选用主题途径检索。

第三步:查找主题索引,获得相关文摘号。逐年查找《电子科技文摘》,如2003年的年度主题索引,可得“网络安全-安全措施 10486”。

第四步:根据文摘号,查阅文摘。

0310486 主动型安全防护措施-陷阱网络的研究与设计沟通[刊]/刘宝旭//计算机工程.—2002,28(12).9-11,16……

第五步:根据文献出处,获取原文。根据文献出处,可到《计算机工程》2002年第28卷第12期第9页至第11页和第16页查得原始文献。

(3) 著者途径。从著者途径检索,可以集中某一著者关于某一学科或专业范围内的同类文献,但不够全面。

(4) 其他途径。其他途径包括利用检索工具的各种专用索引来检索的途径。专用索引的种类很多,常见的有各种号码索引(如专利号、入藏号、报告号等),专用符号代码索引(如元素符号、分子式、结构式等),专用名词术语索引(如地名、机构名、商品名、生物属名等)。

3.4.2 标准文献检索工具

1. 概况

1978年5月国家标准总局的成立和1979年7月《中华人民共和国标准管理条例》的颁布,标志着我国标准化工作进入了一个新的发展时期。1979年以来,我国已成立了200个专业标准技术委员会,327个分标准化技术委员会。1978年9月又以中国标准化协会(CAS)名义加入了国际标准化组织(ISO),并参加了其中103个技术委员会。

我国标准的分类是采用字母、数字混合分类法。字母标志大类,数字代表小类。我国标准号结构形式为标准代号+标准编号+发布年份,如GB/T 208—2014。

2. 具体检索工具

(1) 《中国标准文献分类法》。其由国家技术监督局编制,1989年由中国标准出版社出

版。我国标准文献是依据该分类法进行的。

(2)《标准文献主题词表与分类法对照索引》。其由中国技术监督情报研究所、上海技术监督情报研究所编辑,中国标准出版社于1999年11月出版。《标准文献主题词表与分类法对照索引》是标准文献检索语言的综合性对照索引,它将《标准文献主题词表》《国际标准分类法》和《中国标准文献分类法》进行分析对照,形成了标准文献主题检索语言和分类检索语言统一的对照索引。利用对照索引可以实现3种检索语言的相互对照转换。

(3)《中华人民共和国国家标准目录及信息总汇》。其有多种版本,收录多种类型的标准,以《中国标准文献分类法》顺序编排,全文都有中英文对照,并附有标准顺序号索引。

(4)《国家标准、国外标准、作废代替对照目录》。现有最新版本。该目录收集了1991年年底前,国家标准的全部作废、代替内容,包括专业分类、现行国家标准号、首次发布与历次代替日期、被代替的标准号等。

(5)《中国国家标准汇编》。这套汇编是查阅国家标准原件的重要工具书,在已知标准号的情况下,可以直接查到标准原文。

(6)《中华人民共和国国家标准目录》。全书按行业标准名称代号的汉语拼音排列顺序依次编排,各行业标准目录按分类和标准序号编排。该目录是检索我国行业标准的重要检索工具。

3.4.3 专利文献检索工具

1. 概况

专利是由国家专利局根据发明人的申请,经过一定的审查程序,认为其发明符合专利法规定的条件,而授予发明人在一定时期内独占该项发明的权利。

专利文献是指一切与专利有关的文献的统称,包括专利说明书和用以检索专利的工具书和专利公报、专利文摘、专利的各种索引、专利分类表、专利目录等各种专利出版物。专利说明书是专利的主体。

2. 具体检索工具

(1)《中国专利公报》。《中国专利公报》是中国专利局的官方出版物,专门公布和公告与专利申请、审查、授权有关的事项和决定。《中国专利公报》是查找中国专利文献,检索中国最新专利信息和中国专利局业务活动的主要工具书。

《中国专利公报》分为《发明专利公报》《实用新型专利公报》和《外观设计专利公报》3种。3种公报均为周刊。

《发明专利公报》的主体是报道申请公开、申请审定(1993年1月1日前)、专利权授予和专利事项变更的内容及索引。申请公开部分,著录每一件专利申请的IPC分类号、申请号、公开号、申请日、优先权、申请人、发明人、发明名称、摘要及附图等内容,即说明书扉页上的内容。款目按IPC号字母数字顺序编排。申请审定和专利授权部分无文摘,其他著录项目除个别变动外,与申请公开部分相同。这3部分分别编制了IPC索引申请号索引和申请人(专利权人)索引,以及公开号(公告号)/申请号对照表,提供从国际专利分类、申请人(专利权人)和公开号(公告号)检索中国专利的途径。

此外,《发明专利公报》的专利事务部分,还通报实质审查,申请的驳回与撤回、变更,专

利权的继承和转让,强制许可,专利权的无效宣告和终止等事项。《实用新型专利公报》和《外观设计专利公报》只有申请公告(1993年1月1日前)、专利权授予和专利事务等部分以及相应的索引。编排体例与《发明专利公报》相似。

(2)《中国专利索引》。《中国专利索引》是《中国专利公报》中索引的年度累积本,分为分类年度索引和申请人、专利权人年度索引两个分册。各分册都包括发明专利、实用新型专利和外观设计专利3部分。

分类年度索引的款目按IPC顺序排列。检索者根据检索课题所属的国际专利分类号,可由索引查出有关专利的公开号(公告号)、申请人(专利权人)、发明名称及《中国专利公报》刊登的卷期号。

申请人、专利权人年度索引按申请人或专利权人姓名或译名的汉语拼音字母顺序排列。检索者根据申请人或专利权人的姓名或译名,可由索引检索出其专利申请的公开号(公告号)、IPC号、发明名称及《中国专利公报》刊登卷期号。

根据年度索引的检索结果,可以在《中国专利公报》中找到文摘或者向专利说明书收藏单位索取专利说明书。

3.4.4 《全国报刊索引》和《全国报刊索引数据库》

1. 《全国报刊索引》

取材:国内公开发行的全国性、专业性、省(市)自治区报刊及部分地方报刊,也有部分内部报刊,约8000种。

刊期:月刊。

结构:分类编排+著者索引。

价值:及时反映了全国最新的科研成果和学术动态。

2. 《全国报刊索引数据库》

《全国报刊索引数据库》是《全国报刊索引》的电子版,由上海文达信息公司出版。

上海图书馆网址为 <http://www.libnet.sh.cn>。

布尔检索:8个字段(分类、题名、著者、著者单位、出处、年份、人名、主题)。

索引词检索:系统提供索引词列表。

3.4.5 中国人民大学书报资料中心《复印报刊资料》数据库

中国人民大学书报资料中心成立于1958年,是国内最早搜集、整理、存储、提供社会科学和人文科学信息的学术机构,其产品被誉为“中华学术的窗口”“中外文化交流的桥梁”。其有8个分册:经典、哲学、社会科学总论类,政治类,经济类,文化教育类,语言文艺类,史地类,科技类,著者索引。

在国内公开出版的多种报刊上选材,按专题分册影印出版。

全文数据库的检索:原文出处、原刊地名、原刊期号、分类号、分类名、复印期号、标题词、作者、关键词、任意词、复合检索等。

网址为 <http://book.zlzx.org>。

3.4.6 《中文科技期刊数据库》

《中文科技期刊数据库》于1989年研制成功,以软盘的形式在全国发行,1992年推出光盘版,1997年推出网络版。

网址为 <http://lib.cqvip.com/index.aspx>。

3.4.7 万方数据资源系统

万方数据股份有限公司是中国科技信息研究所组建的国家级综合性科技信息中心。万方数据资源系统又称“数据库之王”“科技信息集散地”。万方数据资源系统有以下3个子系统。

1. 科技信息系统

科技信息系统是中国唯一完整的科技信息群。它汇集了中国学位论文文摘、会议论文文摘、科技成果、专利技术、标准法规、各类科技文献、科技机构、科技名人等近百个数据库。其上千万元的海量信息资源,为广大科研单位、公共图书馆、科技工作者、高校师生提供最丰富最权威的科技信息。

2. 数字化期刊群

“数字化期刊群”是国家“九五”重点科技攻关项目,由国家科技部组织实施,原中国科技信息研究所万方数据网络中心编辑制作,现中国科技信息研究所绝对控股的万方数据股份有限公司开发运作。

“数字化期刊群”是万方数据资源系统中的重要信息服务栏目之一,是中国核心期刊(遴选)数据库的重要基地,是以因特网为载体的集成化、网络化、数字化的全文期刊文献服务系统。《数字化期刊全文数据库》以中国数字化期刊群为基础,整合了中国科技论文、引文数据库及其他相关数据库中的期刊条目部分内容,基本包括了我国文献计量单位中自然科学类统计源期刊和社会科学类核心源期刊。它不仅是我国首家网上期刊的出版联盟,而且是核心期刊测评和论文统计分析的数据源基础。

3. 商务信息子系统

凭借数十年的商务信息采集经验,万方面向企业用户推出工商资讯、经贸信息、咨询服务、商贸活动等项服务内容,其主要产品为《中国企业·产品库》。

《中国企业·产品库》始建于1988年,多年的发展使企业产品库规模渐大,内容渐全,服务更好,在企业产品信息服务界有一定的知名度;产品库现已建立由全国200多个地(市)信息中心组成的信息收集、处理与维护系统,并在上海、广东、辽宁等26个省(市)建立了分库站点,实现了信息的互联、互通与互换。

网址为 <http://www.ceie.com.cn/>。

3.4.8 《中国科学引文数据库》和《中国社会科学引文数据库》

《中国科学引文数据库》由中国科学院文献情报中心出品,网址为 <http://www.las.ac.cn>,可检索核心期刊的发文情况、科技文献被核心期刊引用的情况以及全部作者及作者所在机构的发文情况。

《中国社会科学引文数据库》由南京大学中国社会科学研究评价中心出品,网址为 <http://cssci.nju.edu.cn>。

“中文社会科学引文索引”(CSSCI)是国家、教育部重点课题攻关项目。作为我国社会人文科学主要文献信息查询与评价的重要工具,CSSCI 提供来源文献、被引文献、优化检索等多种信息检索。

3.5 外文检索工具

3.5.1 SCI(美国《科学引文索引》)

SCI(Science Citation Index)是由美国科学信息研究所(Institute for Scientific Information, ISI)于1961年创办出版的引文数据库。SCI是一部国际性的检索刊物,包括数、理、化、农、林、医、生命科学、天文、地理、环境、材料、工程技术等自然科学,主要侧重基础科学。所选用的刊物来源于94个类、40多个国家、50多种文字,这些国家主要有美国、英国、荷兰、德国、俄罗斯、法国、日本、加拿大等,也收录了一定数量的中国刊物。

1. 引文索引

(1) 概念:把引文作为检索词,标引出所有引用过该引文的信息,即构成引文索引。

(2) 特点:引文索引将引文作为检索词,简便易查;利用引文索引检索信息,检索范围不断扩大;引文索引具有引文统计分析的功能。

2. SCI 印刷版

(1) 结构。

Citation Index(A、B、C分册)

Source Index(D分册)

Permuterm Subject Index(E、F分册)

(2) 检索途径。包括机构途径、作者途径、关键词途径。

① 机构途径。

目的:了解某个机构发表的文献被SCI收录的情况。

检索步骤:

第一步:已知机构名称,但不清楚该机构所在地,先查机构索引的机构部分,查到该机构所在国家和城市。

第二步:在机构索引的地理部分按字顺查该机构所在国家(美国为州名),再查城市名及省名,其下列出各级机构被SCI收录文献的作者和出处。

第三步:按作者姓名字顺查来源索引,了解来源文献的详细信息。

【例 3-5】 试查中山大学中山眼科中心发表的论文被SCI收录的情况。

步骤:

第一步:查团体索引的地理部分,先按字顺查 PEOPLES R CHINA,然后查 GUANG-DONG 或 GUANGZHOU,再分别查一级机构 SUN YAT SEN UNIV MED SCI 二级机构 ZHONGSHAN OPHTHALM CTR,其下列出了眼科中心被 SCI 收录文献的作者和刊名。

第二步:按作者姓名字顺转查来源索引。

② 作者途径。

目的:查某个作者有哪些文献被 SCI 收录。

检索步骤:按作者姓名字顺查来源索引。

③ 关键词途径。

目的:检索某个研究领域或研究课题被 SCI 收录的相关文献。

检索步骤:

第一步:根据课题内容确定关键词,按字顺查轮排主题索引,由主标题和副标题可查到一系列相关文献的作者。

第二步:按作者姓名字顺查来源索引,可查到与该课题有关的所有来源文献的详细信息。

3. SCI 网络版

SCI 的网络版,即 SCIE(Science Citation Index Expanded)由美国汤姆森科技信息集团(Thomson Scientific)开发,是 Web of Science(简称 WOS)检索平台的一个子库。现版本的 Web of Science 包括三大引文库[Science Citation Index-Expanded (SCI-Expanded)——1991 年至今、Social Sciences Citation Index(SSCI)——1998 年至今和 Arts & Humanities Citation Index(A&HCI)——1998 年至今]、两个国际会议录[Conference Proceedings Citation Index-Science(CPCI-S)——1996 年至今和 Conference Proceedings Citation Index-Social Science & Humanities (CPCI-SSH)——1996 年至今]和两个化学数据库[Current Chemical Reactions(CCR-EXPANDED)——1985 年至今(包括 Institut National de la Propriete Industrielle 化学结构数据,可回溯至 1840 年)、Index Chemicus(IC)——1993 年至今]。

3.5.2 EI——The Engineering Index(美国《工程索引》)

1. 报道范围

EI 的报道范围很广,涉及工程技术领域的各个方面。它报道的信息主要来源于各国的科技期刊和会议文献,图书、科技报告、学位论文,政府出版物也少量收录,一般不报道专利信息。从出版物的文种看主要是美、英、德三国的英文出版物。目前,EI 收录了 50 个国家、15 种文字出版的 3 500 种科技期刊和 1 000 多种世界范围的会议录、科技报告、图书、论文集等。

2. 出版形式

EI 的出版形式包括印刷版(月刊、年刊)、缩微胶卷、机读磁带、光盘、网络版。

(1) 印刷版的结构及检索途径。

① 文摘正文。从 1993 年起,文摘正文均按《工程索引叙词表》(EI Thesaurus)的字顺编排。该词表采用单级叙词,文摘条文按其内容收录在相关的叙词下。

② 辅助索引(主题索引、著者索引)。印刷版检索途径有主题途径、著者途径、著者单位途径。

(2) 《工程索引》网络版。EI Compendex Web 自 1998 年 11 月起,美国工程信息公司(Engineering Information Inc.)与清华大学合作,在清华图书馆设立镜像服务器,提供基于 Web 方式的 EI Village 信息服务。

3.5.3 SA——Science Abstracts(英国《科学文摘》)

1. SA 印刷版

(1) 构成。SA 印刷版有文摘本、半年度累积索引、多年度累积索引及 INSPEC 叙词表。

(2) 结构。文摘是 SA 印刷版的主体,每辑都按类编排。索引有主题指南、主题索引、著者索引、会议索引等。

(3) 检索途径。检索途径包括分类途径、主题途径和作者途径。

2. SA 数字版

(1) 联机版。联机版早在 1973 年就投入使用,其数据库在 DIALOG、ORBIT、ESA/IRS 等系统中均有相应的文档。

(2) 光盘版。SA 光盘版(INSPEC 数据库)每年出版一张光盘,每张光盘覆盖的文献时间跨度大约为 5 年,每季度更新一次。

(3) 网络版。20 世纪 90 年代后出版的,称为 INSPEC 数据库。目前,基于 Web of Knowledge 检索平台,可以提供 INSPEC 数据库的检索。其网址为 <http://ieeexplore.ieee.org/Xplore/home.jsp>。INSPEC Web 高级检索界面如图 3-2 所示。

图 3-2 INSPEC Web 高级检索界面

3.5.4 CA——Chemical Abstract(美国《化学文摘》)

1. CA 概况

CA 是著名的世界性化学化工文摘刊物,它摘录的内容除化学化工领域外,还涉及生物学、医学、农业、食品、冶金、轻工、物理等领域的信息,这些信息来源于 150 多个国家出版的科技期刊,以及专利信息、会议信息、研究报告、学位论文、图书、视听资料、寄存手稿等,年信息报道量为 50 多万篇,自称囊括了全世界化学信息的 98%,也是举世公认的收录化学信息最全的权威工具。

CA 于 1907 年创刊,由美国化学学会(ACS)和化学文摘社(CAS)编辑出版。目前它在形式上是周刊,每年两卷,每卷 26 期。从内容上看,CA 是双周刊,它按内容分为单期、双期交替出版,单期出版生物化学和有机化学类,双期出版高分子化学、应用化学与化学工程、物理工程、无机化学与分析化学。

2. CA 文摘结构

(1) 正文。CA 分五大部分,80 个大类,文摘部分按 80 大类编排。

(2) 索引。期索引(关键词索引、著者索引、专利索引)、卷索引。

(3) CA 的索引系统。

CA 索引本:普通主题索引(General Subject Index,GS)、化学物质索引(Chemical Substance Index,CS)、分子式索引(Formula Index,FI)、环系索引、母体化合物手册、登记号索引、来源索引(CASSI)、索引指南、著者索引(Author Index,AI)、专利索引(Patent Index,PI)。主体索引:KI、GS、CS、FI。补充索引:AI、PI。辅助索引:环系索引、母体化合物手册、登记号索引。工具索引:索引指南、来源索引。

(4) CA 的特点。

① 收录范围广,信息量大。期刊收录多达 9 000 余种,另外还包括来自 47 个国家和 3 个国际性专利组织的专利说明书、评论、技术报告、专题论文、会议录、讨论会文集等,涉及世界 200 多个国家和地区 60 多种文字文献。到目前为止,CA 已收文献量占全世界化工化学总文献量的 98%。

② 索引完备,检索途径多。

③ 报道迅速。自 1975 年第 83 卷起,CA 的全部文摘和索引采用计算机编排,报道时差从 11 个月缩短到 3 个月,美国国内的期刊及多数英文书刊在 CA 中当月就能报道。

网络版 SciFinder 更使用户可以查询到当天的最新记录。CA 的联机数据库可为读者提供机检手段进行检索,大大提高了检索效率。网址为 <http://www.acs.org/content/acs/en.html> 或 <http://www.cas.org/products/scifinder>。

3.5.5 OCLC(联机计算机图书馆中心)

OCLC 是世界上最大的提供网络文献信息服务和研究的机构之一。该中心创立于 1967 年,是俄亥俄州 54 所大学图书馆共同开发的计算机图书馆系统。从 1977 年起 OCLC 的成员扩大到了俄亥俄州以外的图书馆。网址为 [http://www.oclc.org/zhen-asiapacific/home.html? redirect=true](http://www.oclc.org/zhen-asiapacific/home.html?redirect=true)。

First Search 联机信息服务是 OCLC 从 1991 年开始推出的一个新产品。1999 年 8 月, OCLC 完成了新版的 First Search——New First Search, 它是一个综合的以 Web 为基础的联机参考服务系统。利用 New First Search 可以检索到 75 个数据库, 其中有 30 多个数据库可检索到全文, 总计有 7 500 多种期刊的联机全文和 3 000 多种期刊的联机电子映像, 共达 900 多万篇全文文献。

免费检索的 12 个数据库列表如下:

(1) Article1st——12 500 多种期刊的文章索引。Article1st 数据库包括 12 500 多种学术期刊的文章引文, 主题覆盖了工商、人文学、医学、科学、技术、社会学和大众文化等。虽然大多数期刊是英文资料, 但也收录了部分其他语言的期刊。该库覆盖了 1990 年到现在的资料, 每天更新。

(2) ECO——联机电子出版物(只能查到书目信息)。ECO 是一个全部带有联机全文文章的期刊数据库。它的主题范畴广泛, 目前记录来自 2 000 多种期刊, 期刊的数量还正在逐步增加。数据库中的文章都以页映像的格式(PDF、RealPage 或 HTML)显示, 在页映像中包括了文章的全部原始内容和图像。

(3) ERIC——教育方面的期刊文章和报告。ERIC 是由教育资源信息中心生产的已出版的和未出版的教育方面的资料来源的一个指南。它囊括了数千个教育专题, 提供了最完备的教育书刊的书目信息, 覆盖了从 1966 年到现在的资料, 每月更新记录。

(4) GPO——美国政府出版物。GPO 包含 52 万多条记录, 报道了与美国政府相关的各方面的文件。这些文件的类型有国会报告、国会意见听证会、国会辩论、国会档案、法院资料以及由美国具体实施部门(如国防部、劳动部、总统办公室等)出版发行的文件。它覆盖了从 1976 年 7 月以来的资料, 每月更新记录。

(5) MEDLINE——医学期刊的文章摘要。MEDLINE 覆盖了所有医学领域, 包括临床医学、实验医学、牙科学、护理、保健服务管理、营养学以及其他学科。它索引了国际上出版的 3 500 多种期刊, 覆盖了从 1965 年到现在的资料, 每月更新记录。

(6) NetFirst——OCLC 的 Internet 资源的数据库。NetFirst 数据库的引文(附有摘要和主题词)描述了 Internet 的信息资源, 使用户能够借“指、点”的动作, 即刻从万维网的 FirstSearch 访问其他站点。

(7) PapersFirst——国际学术会议论文索引。该数据库包括在世界各地学术会议上发表的论文, 它覆盖了自 1993 年 10 月以来在“大英图书馆资料提供中心”的会议录收集的在每一个大会、专题讨论会、博览会、讲习班和其他会议上发表的论文, 每两周更新一次。

(8) Proceedings——国际学术会议录索引。Proceedings 是 PapersFirst 的相关库, 它包括在世界各地举行的学术会议上发表的论文的目录表。该库提供了一条检索“大英图书馆资料提供中心”的会议录的途径。

(9) UnionLists——OCLC 的期刊联合列表库。该数据库包括数千种期刊的馆藏情况, 有 740 多万条记录, 每一条记录列出了 OCLC 的成员馆收藏的一种期刊每期的情况, 每半年更新一次。

(10) WilsonSelect——H. W. Wilson 公司的全文库。该数据库是一个包括联机全文、索引和摘要的记录的集合, 这些全文文章选自 H. W. Wilson 公司的普通科学文摘、人文学科文摘、读者指南文摘和 Wilson 商业文摘。它覆盖了 800 多种期刊的从 1994 年到现在的

资料,每周更新一次。

(11) WorldAlmanac——世界年鉴。该数据库在 1868 年第一次出版,它是适用于包括学生、图书馆的读者、图书馆的参考咨询人员和学者等几乎每个人的一个十分重要的参考工具。它涉及的范畴包括艺术和娱乐、新闻人物、计算机、科学和技术、经济学、体育运动、环境、税收、周年纪念日、美国的城市和州、国防、人口统计、世界上的国家等,每年更新一次。

(12) WorldCat——世界范围图书馆的图书和其他资料。该库是 OCLC 的一个联机的联合目录数据库。它目前包括 4 100 多万条记录,这些记录来自 370 多种语言的文献,覆盖了从公元 1000 年到现在的资料,基本上反映了世界范围内的图书馆所拥有的图书和其他资料。它的主题范畴广泛,并以每年 200 万条记录的速度增长。该库每天更新。

3.5.6 Elsevier Science 学术期刊

Elsevier Science 是世界上公认的高品位学术出版公司,也是全球最大的出版商,已有 100 多年的历史。该公司除了出版图书外,还是当今世界最大的学术期刊出版商,内容涉及生命科学、物理、医学、工程技术及社会科学,其中许多为核心期刊。近年来,Elsevier 公司合并了一些出版社,如 Academic Press、Cell Press 的学术期刊全文数据库也已加入 ScienceDirect,目前国际网站的服务器上安装有 1 500 多种期刊。

3.5.7 Springer Link 全文期刊

德国施普林格(Springer-Verlag)是世界上著名的科技出版集团,通过 Springer Link 系统提供其学术期刊及电子图书的在线服务。2002 年 7 月开始,Springer 公司和 EBSCO/Metapress 公司在国内开通了 Springer Link 服务。

在 2004 年 Kluwer Academic Publishers(KAP)并入 Springer 后,目前 Springer Link 中已包含 1 200 多种全文学术期刊。(特别提示:凡是在期刊列表前有符号,即表示可阅读全文。)

Springer Link 中的期刊及图书划分为 12 个学科:建筑学、设计和艺术,行为科学,生物医学和生命科学,商业和经济,化学和材料科学,计算机科学,地球和环境科学,工程学,人文、社科和法律,数学和统计学,医学,物理和天文学。