



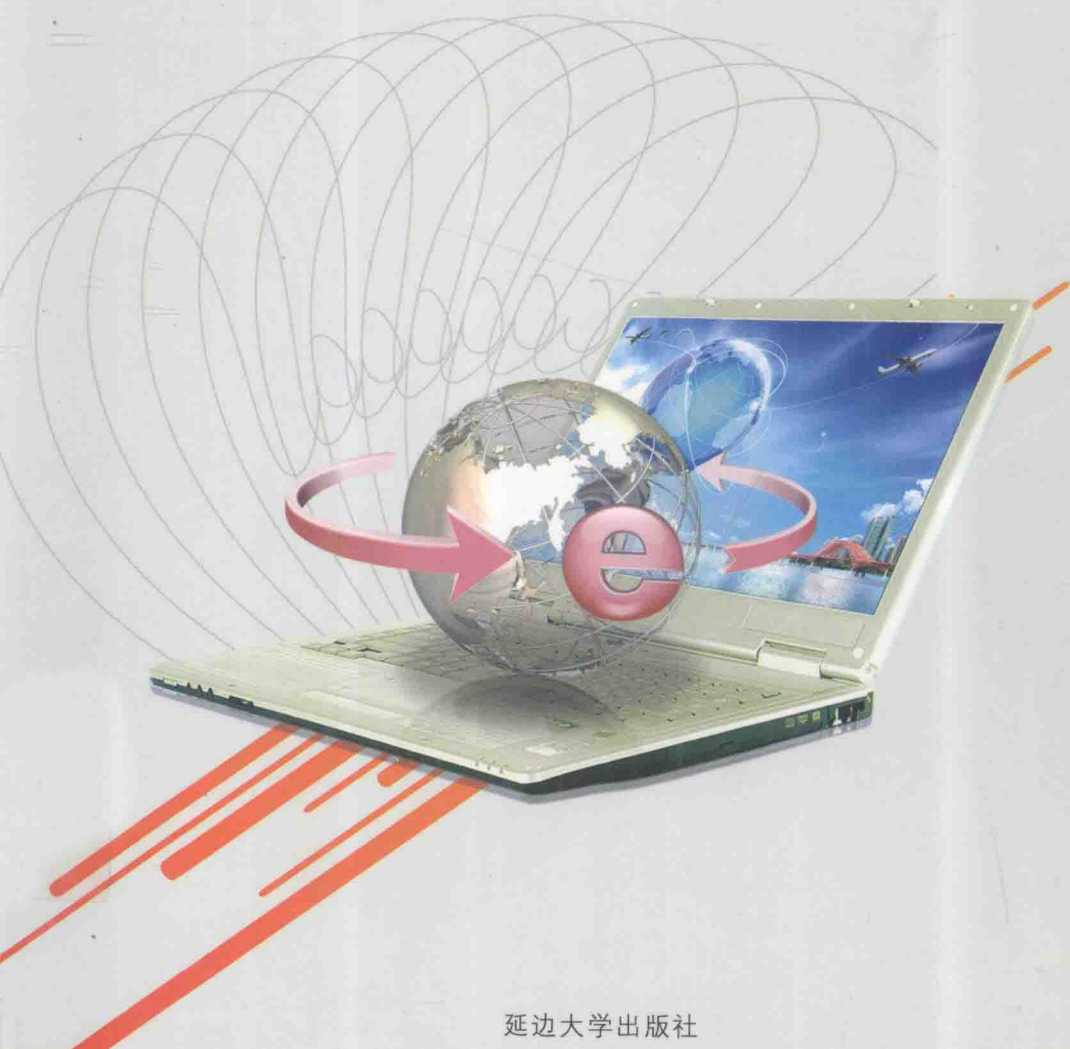
DAXUESHENG

# 大学生

XINXILIYONG YU JIANSUO

## 信息利用与检索

方力◎主编 刘乃熙 张桂荣◎副主编



延边大学出版社

# 大学生信息利用与检索

主 编 方 力  
副主编 刘乃熙  
张桂荣

延边大学出版社

**图书在版编目(CIP)数据**

大学生信息利用与检索/方力主编. —延吉:延边大学出版社,2009.5

ISBN 978-7-5634-2347-7

I. 大… II. 方… III. 大学生—情报检索—基本知识  
IV. G252.7  
中国版本图书馆 CIP 数据核字(2009)第 085159 号

书 名:大学生信息利用与检索

---

主 编:方 力

副 主 编:刘乃熙 张桂荣

责任编辑:贾 锐

装帧设计:揽胜视觉

出版发行:延边大学出版社

社 址:吉林省延吉市公园路 977 号

邮 编:133002

网 址:<http://www.ydcbs.com>

E-mail:ydcbs@ydcbs.com

电 话:0433-2732435

传 真:0433-2732434

0433-2133001(发行部电话)

传 真:0433-2733266

印 刷:北京龙跃印务有限公司

开 本:16 开 规 格:710×1000

印 张:24 字 数:380 千字

印 数:2000

印 次:2009 年 9 月第 1 版

2009 年 9 月第 1 次印刷

ISBN 978-7-5634-2347-7

---

定 价:43.80 元

## 前 言

高等学校肩负着向学生传授基本知识、培养学生自学能力和独立研究能力的重任。信息检索课程是培养学生能力的一个重要环节,是培养现代创新人才的重要内容之一,通过这门课的学习,可以培养学生强烈的信息意识,提高学生的信息能力,丰富学生的信息知识,在文献检索过程中了解文献的基础知识,掌握获取与利用文献信息的方法与技能,为今后更新知识、跟上现代科学技术发展的步伐打下坚实的基础。

目前,在各类高等学校中,文献检索已经普遍作为研究生、本科生、专科生的必修课或选修课。

当今社会是信息社会,虽然学生对计算机、网络很熟悉,但是真正地掌握信息检索方法的人还很少,学校每年在数据库方面投入资金很多,但是学生利用率却很低,德国柏林图书馆门前有这样一段话:“这里是知识的宝库,你若掌握了它的钥匙,这里的全部知识都是你的”。这个钥匙就是文献检索的方法,文献检索课将把这个钥匙交给学生,使他会在任何一个图书馆、任何一个网络终端前,随心所欲地漫游世界的知识海洋,获取全世界的最新信息。我校学生在十年前就开设了信息检索课,但是一直以来就没有适合我校校情的教材,学生上课大多是记笔记的形式,学生累,教师也累,根据《关于高校文献检索课基本要求》,通过多年的教学实践,我们在使用了几年讲义的基础上编写了这本《大学生信息检索》来填补这个空白,该书是从我校综合院校实际出发,针对我校各类学

生的专业特色和图书馆信息收藏实际情况专门为学生编写的教材。

本书首先介绍了信息检索的基础理论知识,然后分别逐章讲述了图书、期刊、会议文献、学位论文、专利文献等信息及其检索方法,介绍了国内外几种权威参考数据库及网络搜索引擎的使用和检索方法,最后就建立个人资料库及有关学术论文的撰写与发表做了介绍。

方力负责编写了第七章至第十一章;并审校了全部章节。刘乃熙负责编写第一章、第二章、第三章、张桂荣负责编写了第四章、第五章、第六章。

本书在编写过程中参考了许多教材、论文和网页,这里向所有参考文献的作者表示诚挚的谢意。由于编者的水平所限,加之编纂时间仓促,缺点和疏漏在所难免,恳请专家和学者指正。

## 目 录

<b>第一章 信息资源及其检索概述</b> .....	(1)
第一节 信息资源 .....	(1)
第二节 信息检索基础理论 .....	(8)
第三节、关于《信息检索与利用》课 .....	(27)
<b>第二章 汉语字、词、句检索</b> .....	(33)
第一节 现代汉语字、词检索 .....	(33)
第二节 古代汉语字、词检索 .....	(38)
第三节 汉语成语、典故检索 .....	(48)
第四节 汉语方言、谚语、俗语、歇后语检索 .....	(52)
第五节 诗词、文句出处检索 .....	(55)
<b>第三章 图书信息检索</b> .....	(61)
第一节 概述 .....	(61)
第二节 网络环境下书目信息的检索 .....	(64)
第三节 电子图书及其检索 .....	(65)
<b>第四章 期刊信息检索</b> .....	(68)
第一节 期刊概述 .....	(68)
第二节 期刊(出版、收藏)信息检索 .....	(71)
第三节 电子期刊及其检索 .....	(73)
第四节 网上免费电子期刊简介 .....	(93)
<b>第五章 会议文献及其检索</b> .....	(94)
第一节 学术会议与会议文献 .....	(94)
第二节 几种重要的中文数据库 .....	(95)
第三节 美国《科学技术会议录索引》 .....	(105)
<b>第六章 学位论文的检索</b> .....	(115)
第一节 概述 .....	(115)

第二节	学位论文的检索 .....	(115)
<b>第七章</b>	<b>专利文献信息检索 .....</b>	<b>(130)</b>
第一节	专利与专利文献 .....	(130)
<b>第八章</b>	<b>科技查新 .....</b>	<b>(144)</b>
第一节	查新的作用及查新的程序 .....	(144)
第二节	国内科技查新主要利用的信息检索系统 .....	(152)
第三节	国际查新联机检索系统 .....	(155)
<b>第九章</b>	<b>参考数据库及网络搜索引擎 .....</b>	<b>(171)</b>
第一节	常用中文数据库及检索 .....	(171)
四、维普资讯中文科技期刊数据库 .....		(194)
第二节	常用外文全文数据库及检索 .....	(225)
第三节	常用外文文摘数据库及检索 .....	(253)
第四节	原文传递服务 .....	(297)
第五节	网络搜索引擎 .....	(298)
<b>第十章</b>	<b>建立个人资料库 .....</b>	<b>(307)</b>
第一节	个人资源管理的必要性 .....	(307)
第二节	信息资料的收集 .....	(308)
第三节	信息资料收集的工作步骤 .....	(313)
第四节	如何提高个人的信息收集能力 .....	(314)
第五节	信息资料的选择与整理 .....	(316)
第六节	信息资料整理应该注意的两个问题 .....	(319)
第七节	个人信息资料库的组织方法 .....	(320)
第八节	超文本方式的电子信息资料库构建 .....	(321)
<b>第十一章</b>	<b>论文的撰写与投稿 .....</b>	<b>(325)</b>
第一节	论文概述 .....	(325)
第二节	学年论文 .....	(328)
第三节	学位论文的写作 .....	(329)
第四节	学术论文的撰写 .....	(330)

# 第一章 信息资源及其检索概述

## 第一节 信息资源

### 一、信息资源的概念

人类社会与信息紧密相连。人类不仅需要对生于斯的世界有所了解,而且不同人群之间需要协作与沟通,因此人类不仅时刻需要获得自然界信息,也需加强相互之间的信息交流。近年来,赛伯空间(cyberspace,又称为网络空间)和虚拟现实(virtual reality)的出现已使“秀才不出门,便知天下事”的神话成为现实,信息的高速流动使地球变成一个小小村落。信息越来越重要地影响着人们的生活。

#### 1. 什么是信息?

信息对应的英文单词是 information,它来源于拉丁字 *informare*,意思是解释、陈述,是消息的意思,英文常译为情报、资料、消息、报道、知识的意思。“信息”一词最早出自南唐诗人李中的《暮春怀故人》中的“梦断美人沉信息,目穿长路倚楼台”。因此,很多人长期以来就把信息看作是消息的同义语,简单的把信息定义为能够带来新内容、新知识的消息。由于人们接触信息的角度不同,所以不同的人对信息给出了不同的定义和描述,据不完全统计,有文可考的“信息”定义已不下一百个。

在我国国家标准《情报与文献工作词汇基本术语》中,对信息的概念定义为:信息是物质存在的一种方式、形态或运动状态,也是事物的一种普遍属性,一般指数据、消息中所包含的意义,可以使消息中所描述事物的不确定性减少。这一解释基本涵盖了对信息的属性(客观存在性)、作用(消除不确定性)及形式(数据、消息等事实)的定义。

#### 2. 什么是信息资源?

信息资源(information resources)最早是由国外学者提出的,时间大约在 20 世纪 60 年代末 70 年代初。信息资源的定义与信息的定义一样,目前仍是众说纷纭。尽管在认识上存在一定的分歧,但国内外研究信息资源管理的大多数专家都认为,应该从狭义和广义两种角度来认识和理解信息资源的涵义。

(1)狭义信息资源:指人类社会活动中经过加工处理的、有序化的并大量累积后的有用信息的集合。

(2)广义信息资源:信息和它的生产者以及信息技术的集合。也就是说,广义的信息资源一般由 3 部分构成:一是人类社会活动中经过加工处理有序化并大量累积后的有用信息的集合;二是为某种目的而生产有用信息的信息生产者的结合;三是加工、处理和传递有用信息的信息技术的集合。



## 二、信息资源的分类

按照不同的标准可以将信息资源划分为不同的种类,下面分别介绍。

### 1. 按信息资源所依附的载体划分

可以分为体载信息资源、实物信息资源、文献信息资源、网络信息资源 4 种类型。

(1)体载信息资源:以人体为载体并能为他人识别的信息资源。按其表述方式又可分为口语信息资源(谈话、授课、讲演、讨论、唱歌等)和体语信息资源(表情、手势、姿态、舞蹈等)。

(2)实物信息资源:以实物为载体的信息资源。依据实物的人工与天然特性又可将事物信息资源分为以自然物质为载体的天然实物信息资源和以人工实物为载体的人工实物信息资源(产品、样品、模型、雕塑等)。

(3)文献信息资源:指以文字、图形、符号、声频、视频等方式记录在各种载体上的知识和信息资源,是人们目前利用的最多的信息资源。

(4)网络信息资源:以网络为纽带链接起来的信息资源和以网络为主要交流、传递、存储的手段与形式的信息资源。网络信息资源通过网络将原本相互独立、分布于世界各地的数据库、信息中心、文献中心等联结在一起,形成一个内容与结构全新的信息整体。

本课程学习的就是文献信息资源以及网络信息资源的检索。第二章到第七章讲述文献信息资源的检索,关于网络信息资源的检索将在第八章、第九章讲述。

### 2. 按信息资源传递的范围划分

可以分为公开信息资源、半公开信息资源和非公开信息资源 3 种。

(1)公开信息资源:又称共享信息资源或白色信息资源,指公开发行、流通和传递的信息资源,其蕴涵的信息人人均可以使用。一般来说,公开信息资源的数量最大,而且能够作为信息商品进入流通领域。

(2)半公开信息资源:又称灰色信息资源,指非公开发行、流通和传递,从常规途径难以获取的信息(内部消息)。这类信息出版量小,发行渠道复杂,流通传递范围有限,不易收集。例如,内部的刊物、技术报告、会议资料等。

(3)非公开信息资源:又称黑色信息资源,指人们未破译或未被辨识的,也指处于保密状态的信息。例如,考古发现的古老文字,未解密的政府文件,内部档案,个人日记,私人信件等。

## 三、信息资源的特性

### 1. 载体依附性

(1)信息不能独立存在,需要依附于一定的载体。

(2)同一个信息可以依附于不同的载体。如:交通信息既可通过信号灯显示,也可以警察的手势来传递;写教案即可以手写,也可以用电脑来写。

(3)载体形式多样,有印刷型、缩微型、机器型、声像型、网络型。

(4)载体的依附性具有可存储、可传递、可转换特点。

### 2. 价值性

(1)信息不能直接供给人们物质需要,但能满足人们精神生活的需要。

(2)信息可以促进物质、能量的生产和使用。

(3)信息可以增值。

(4)信息只有被人们利用才能体现出其价值。

### 3. 时效性

(1)信息只有满足“价值”条件时,才具有时效性。

(2)信息时效性不受时间、空间限制。

(3)信息在某一特定时刻的状态,会随着时间的推移而变化。如:交通信息、股市信息、天气预报、会议通知、求职报名、市场动态等。

### 4. 共享性

(1)信息可以被一次性或多次或同时利用。

(2)Internet 实现全球信息资源的共享。

(3)信息共享不会丢失、改变。

### 5. 传递性

(1)按流向的不同划分,有单向传递、双向传递。

(2)按信息传递时信息量的不同划分,有集中传递和持续传递。

(3)按信息传递范围或环境关系的不同划分,有内部传递和外部传递。

### 6. 可处理性

(1)可以压缩、存储、有序化、转换形态。

(2)信息处理的类型:分类、比较、综合、研究、编制。

## 四、文献信息资源

文献信息资源是信息检索的主要对象之一,也是信息资源建设和利用的主要类型之一,因此文献信息检索是本门课讲述内容的重点。

### 1. 文献的概念

广义上凡是人类的知识信息以文字、图形、代码、符号、声频、视频等形式,用一定的技术手段记载在物质载体上,由此而形成的每条记录,都统称为文献。

《中华人民共和国国家标准? 文献著录总则》对文献的定义:文献是记录有知识的一切载体。从这一精炼的定义中,我们可以提炼出构成文献的三要素:记录、知识、载体。

### 2. 文献的类型与特点

(1)按文献的载体形态(物质形式)划分为:刻写型文献、印刷型文献、缩微型文献、声像型文献和电子型文献。

①刻写型文献:是指在印刷术尚未发明之前的古代文献和当今尚未正式付印的手写记录,以及正式付印前的草稿。如古代的甲骨文、金石文、绵帛文、竹木文以及现今的手稿、日记、信件、原始档案、碑刻等。

②印刷型文献(printed form):指以纸张为存储介质,以印刷(油印、铅印、胶印等)为记录手段,将信息固化在纸张上而产生的一种传统的文献形式。

优点:直接阅读,携带方便;缺点:存储密度低,需要占用大量空间,不便于整理和保存。

③缩微型文献(micro form):指以感光材料为存储介质,以光学摄影技术为记录手段而生产出来的一种文献形式,可分为缩微胶片、缩微平片、缩微胶卷、缩微卡片等。

优点:体积小,存储密度高,易于保存和传递;缺点是阅读不方便,需要专门的阅读设备

和环境。

④声像型文献(audio-visual form):又称直感型或视听型文献,指通过专用设备,使用声、光、磁、电等技术将信息以声音、图像、影视和动画等形式表现出来的一种文献。

优点:存储密度大、内容直观形象、表现力强,它在帮助人们观察罕见的自然现象和探索物质的微观结构方面,能起到印刷型文献不能具备的独特作用;缺点:需要专用设备。

⑤电子型文献(electronic form):又称机读型文献,指以磁性材料为存储介质,以穿孔、打字或光学字符识别装置为记录手段,通过计算机对电子格式的信息进行存取和处理而产生的文献。

优点:存储容量大,存取速度快,原记录可以修改、删除或更新等;缺点:必须借助计算机阅读,费用大。

(2)按文献的加工级次(程度)划分为:一次文献、二次文献和三次文献。

①一次文献(primary literature):是著者以自己的经验、研究或研制成果为依据而创作的原始文献,经公开发表或交流后,称为一次文献。包括期刊论文、图书专著、研究报告、科技报告、专利说明书、学位论文等。

特点:一次文献是重要的情报源,具有新颖性、创造性和系统性等特点,但是数量庞大、分布分散、不便于直接查找利用。

②二次文献(secondary literature):指通过科学的方法,将分散无序的一次文献经过筛选、分析、整理,按其内容特征(如主题、分类)和外部特征(如著者、篇名)进行加工提炼、浓缩简化,编制而成的有系统的文献,包括索引、目录、文摘等。

特点:一是具有简明性和系统性;二是能提供文献线索,是查找文献的工具;三是仅对一次文献进行加工,不改变一次文献的内容。

③三次文献(tertiary literature):指在利用二次文献的基础上,将大量的一次文献进行综合分析、归纳、整理,并对原始文献进行重新组织、加工、提炼而编写出来的文献,以便于人们了解某学科或专题的过去、现在和对未来进行预测。包括专题述评、动态综述、手册、年鉴、百科全书、字典等。

特点:一是具有资料性和实用性;二是对一次文献进行浓缩、提炼、重新组织,内容有较大变化。

各次文献之间的关系:上述的三种文献是一个从分散的原始文献到系统化、密集化的过程。一般说,一次文献是形成二次文献和三次文献的基础,是文献的基本形式,是检索利用的对象;二次文献是检索一次文献的工具,故又称检索工具;三次文献是一次文献内容的高度浓缩,是对众多文献分析、综合归纳、整理而形成的,它往往从新的角度和深度揭示相关的一次文献,是科技信息的重要来源。

(3)按文献信息的编辑出版形式划分为:图书、期刊、会议文献、学位论文、专利文献、科技报告、标准文献、政府出版物、产品样本和说明书、技术档案等10种主要的文献类型。

①图书:这是一种论章成册的出版物,是对已有研究成果、生产技术、实践经验或某一知识的论述或概括。图书按其内容性质和作用可分为:普及读物、教科书、丛书、专著、论文集、参考工具书,如书目、索引、字典、辞典、手册、年鉴、指南和百科全书等。

特点:内容成熟、系统、全面和可靠,是传播知识、教育和培养人才的主要工具。但是由

于其出版周期较长,因此传递信息的新颖性、快捷性不如期刊及会议文献。

②期刊又称杂志,这是一种有固定名称,有一定的出版规律,每期刊载不同著者、译者、编者所编写的论文,按一定编号顺序出版的连续出版物(serials)。期刊按其报道内容性质,可分为学术性、技术性期刊,如 Acta(学报)、Journals(杂志)、Bulletin(通报)、Transactions(汇刊)、Precedings(会刊)、Review(评论)、Progress; Advances in ... (进展)等;快报型刊物,如 Communication(通讯)、letters(通讯);消息性刊物,如 news(新闻)、等;资料性刊物,如 data(数据)、event(事件)等。

特点:出版周期短,报道速度快,信息量大,内容丰富新颖,传播面广,时效性强,能及时反映学科新观点、科研新成果、社会新动态,是科研人员了解科研动态、前沿信息和进展情况的首选文献。据统计,科技人员所获取信息的 65% 以上来源于期刊,它是十分重要和主要的信息源和检索对象。

期刊的识别依据主要有:期刊名称,期刊出版的年、卷、期,国际标准刊号(ISSN)等。

期刊常用的名称:

i. 学术性、技术性刊物: Acta(学报)、Journals(杂志)、Bulletin(通报)、Transactions(汇刊)、Precedings(会刊)、Review(评论)、Progress; Advances in ... (进展)

ii. 快报性: Communication(通讯)、letters(通讯)

iii. 消息性: news(新闻)、news letters 等

iv. 资料性: data、event

③会议文献:会议文献是指在各种学术会议上发表的论文和报告。世界上每年举行上万次的学术会议来交流科研成果,讨论重大的学术问题。会议文献按其出版时间为会前文献、会间文献和会后文献三大类。会前文献主要有会议论文预印本(Preprints)和论文摘要(Advance Abstract),前者是论文的全文,后者是文摘,它们有的只发给与会者,有的公开发行。会间文献主要是指会议期间的开幕词、讲演稿、闭幕词、讨论记录、会议决议等文献,一般均在会议举办时分发给与会者,此类文献大都为行政事务性和情况报道性的资料,学术价值不大。会后文献主要是指会议结束后正式发表的会议论文,是会议文献中的主要部分。如会议录(Proceedings)、会议论文集(Symposium)、学术讨论论文集(Colloquium)等。会议文献常以图书或者期刊等形式出版。

特点:会议文献论题集中,针对性强,学术性较强,内容新颖专深,信息传递速度快,一些重要的研究成果或新的发现,通常首先通过会议文献向社会发布,因此往往反映了当前的某学科最新成就、最高水平和发展趋势。所以,会议文献成为了解各国科技发展水平和动向的重要文献源,因而受到科技界的高度重视。

会议文献识别依据有:会议名称、时间、地点及会议举办单位等信息。

④学位论文:学位论文是指高等院校或科研机构的毕业生为取得各级学位而撰写的学术论文,它按级别可分为学士论文、硕士论文和博士论文。

特点:学位论文一般偏重理论,所论及的内容一般比较专深。特别是硕士、博士学位论文在某些方面往往提出了独到的见解和观点,对研究工作有较大的参考价值。因此,学位论文也是构成文献资源的重要部分之一。但是由于学位论文一般不公开出版发行,除极少数以科技报告、期刊论文形式发表,一般在本单位内得到收藏、流通。我国在中国科技信息

研究所等单位收藏有部分学位论文的复制品。不过现在大家可以通过中国知网 CNKI 和万方数据资源系统来获取一些学位论文全文。

学位论文的识别:往往注明授予学位头衔,授予学位的单位名称,授予学位的地点、时间以及责任者(不仅著录作者姓名,还有指导教师的姓名)。

⑤专利文献:专利文献是指一切与专利制度有关的在专利申请和授权各阶段产生的文献,包括专利说明书、专利公报、专利分类表、专利检索工具以及专利的法律文件,其中主体是专利说明书。据专利的技术水平和应用情况,其类型有:invention(发明)utility patent(实用新型)design patent(外观设计)。

特点:内容丰富新颖、实用、可靠、详尽,且统一编号。然而由于同一个发明,往往在多个国家申请并拥有专利权,从而产生了大量的相同专利及其不同语种的专利文献。这种状况可以有效地促进专利信息的交流,增加不同语种的用户利用专利信息的机会,但同时因为大量重复,也增加了用户信息识别和去掉重复的负担。

关于专利信息的检索将在第七章详细讲解。

⑥科技报告:科技报告是科技人员从事某一专题研究所取得的成果和进展的实际记录。科技报告每份单独成册,有专门编号,用以识别报告类型及其主持机构。最著名的是美国出版的四大报告,即美国军事国防部门的 AD 报告、美国商务出版局的 PB 报告、美国能源部的 DOE 报告和美国宇航部门的 NASA 报告。

特点:反映新技术、新学科较快,内容比较专深、新颖,数据比较可靠,保密性较强,有相当一部分科技报告不公开发行。

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及共同规范。是从事研究、生产、开发及工程建设的一种共同技术依据。按批准机构和适用等级可分为国际标准、区域标准、国家标准、行业标准和企业标准等;按其内容可分为基础标准(包括术语、符号、单位、定义等)、产品特性标准(包括特性、尺寸、形状、成分、质量等)以及方法标准(包括生产方法、作业方法、试验及检测方法等)。

特点:标准文献是一种经权威机构批准的规章性文献,具有一定的法律约束力;有一定的有效时间,需要随着技术发展而不断修订,补充或废除,新陈代谢比较频繁。

⑧政府出版物:政府出版物是指各国政府部门及其所属机构所发表各类文件,主要包括行政性文件和科技文件两大类。查找政府出版物的常用工具有:《中华人民共和国国务院公报》、《中华人民共和国全国人民代表大会常务委员会公报》美国的《政府报告通报及索引》。

特点:政府出版物全面系统地反映了一个国家(地区、部门等)在政治、经济、军事、文化、教育等方面的思想、方针、政策及措施等,对于了解、掌握上述情况,做出相应的决策、采取适当的行为都具有指导性意义和重大的参考价值。

⑨产品样本和说明书:这是一种对定型产品的性能、规格、构造、用途及其使用方法等所作的说明。

特点:图文并茂、形象直观,出版及时,更新迅速,是进行技术革新、技术设备改造、试制新产品的重要技术参考文献。

⑩技术档案:是指在生产建设中和科技部门的技术活动中形成的,涉及一定的工程对

象的技术文件的总称。其内容包括任务书、审批文件、研究计划、技术指标、技术措施、调查材料、设计计算、工艺记录等。它是科研和生产建设中积累经验、提高质量的重要依据,具有重要的信息价值。

特点:明显的保密性和内部控制使用。

此外,科技文献还包括报纸、新闻稿件、科技译文、手稿、地图等。

以上所述的这 10 种主要的文献类型中,图书、期刊、会议文献、科技报告、专利文献、学位论文、技术标准均有其相应的二次文献,即检索工具,所以查找起来比较方便,而政府出版物、产品样本和说明书、技术档案、报纸则多数没有相应的二次文献,所以查找极为不便。

### 3. 现代文献信息资源的发展趋势与特点:

现代科学技术的迅猛发展,使记录科技知识信息的文献呈现出一些新的变化和特点:

①数量多、增长快:科学技术的进步和信息产业的飞速发展,特别是互联网的兴起和普及,极大地推动了社会信息数量的增加和信息的流动。据统计,全世界每年出版近 70 万种期刊,60 余万种新书,登记 40 多万项专利,发表科技文献 500 多万篇,编写 25 万份学术报告、学位论文。1999 年到 2002 年 3 年间,世界范围内信息生产量以平均每年 30% 左右的速度递增。

②形式多、文种多:现代文献的生产突破了传统的纸张印刷方式,声、光、电、磁等现代技术和化学塑胶新材料的广泛应用,使现代文献载体形式发生了重大变化,缩微、声像以及电子文献相继问世。除形式多样化外,世界各国文献使用的语言也在不断增多。过去,世界上科技文献大多数只用“英、德、法”几种文字出版,而现在,有 70~80 种之多,比较集中的文种就有 7~8 种,语言障碍已经成为文献收集、整理和利用的严重问题。

③类型多数量多、内容庞杂:从某种程度上来说,各种情报机构收藏的信息资源包括传统信息资源(包括纸质文献、各类光盘等)及网络信息资源,这些信息资源数量多、内容纷繁复杂,涉及各行各业以及生活的各个方面。特别是网络信息资源包括成千上万种电子期刊和报纸以及政府、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。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。

④信息类型多、范围宽、用途广。近年来,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,单就文本就有文字、符号、表格等,另外还有声音、动画、图形、图象等。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,如发电子邮件或召开在线视频会议等。

⑤文献分布分散:现代科技发展呈分化和综合两大趋势,分化使学科越分越细、分支越来越多,而综合表现为各学科的交叉渗透,出现许多边缘学科。这使学科之间严格的界限逐渐消失,各学科之间的相互联系逐渐加强。由于这一原因,使得文献的分布呈现出既集中又分散的不均匀现象:一方面,一种期刊登载多个学科的文献;另一方面,一个学科的文献往往分散在许多不直接相关的期刊上。

⑥文献老化迅速:现代科技发展的日新月异,使科研成果从发明到推广应用的周期缩短,知识的有效期缩短,不成熟的观点被成熟的观点所代替,错误、片面的数据被客观的事实

校正,不完善的方法被较先进的方法更替,这自然而然形成了科技文献的新旧更替。我们把这种现象称为文献信息的“老化”。1958年,美国科学家引用放射元素衰变过程中的半衰期描绘文献的老化速度。1960年美国伯顿、凯布勒合作,通过对科技文献使用的引文进行载文统计,他们对半衰期的定义为:某学科或某专业现在仍在利用的所有文献的一半是在多长时间内发表的。它是一个时间长度,指正在被利用的文献中的二分之一使用效用所经历的时间。

#### 4. 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及更新。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。通常用文献的“半衰期”来描述文献老化情况。所谓文献的半衰期是指某学科领域目前尚在使用的全部文献中,较新的一半所出版的年限。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 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。

5. 文献内容交叉重复:文献信息内容交叉重复的原因主要有以下几个方面:(1)各国政府部门、学术机构、研究单位的科研选题互相重复,使得反映成果的文献信息内容也出现重复;(2)为获取更多的利润或提高自己的声誉,出版商竞相出版内容相同的热门书刊和新兴学科、尖端学科书刊资料;(3)同一内容的文献以不同形式分别出版,如印刷型文献再以缩微型、声像型或者电子型文献出版;(4)各种原因的再版和改版文献数量增多。冗余文献虽然能扩大人们获得和接触这些文献的几率,但也使识别和使用这些文献变得非常困难。有材料表明,各类文献中,有用的信息内容仅占 25%,这就要求我们在收集文献时,必须加以认真区分和筛选。

### 五、网络信息资源

网络信息资源已经成为一种主要的信息资源形式并表现出鲜明的特点。网络信息资源是一种重要的电子信息。

电子信息是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,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。

,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。所以其优势尚无法完全发挥出来。

## 第二节 信息检索基础理论

### 一、信息检索的概念

广义的信息检索(information retrieval)是指将信息按一定的方式组织和存储起来,并根据用户的需要找出有关信息的过程和技术,又叫信息存储与检索。狭义的信息检索则是指

从已存储的信息集中查找出所需要信息的过程。

通俗地说,信息检索就是信息用户为处理解决各种问题而查找、识别、获取相关的事实、数据、知识的活动及过程。随着信息社会的到来,信息检索的重要地位日益突出,因此我们有必要学习和研究信息检索相关知识,如信息检索原理、信息检索语言、信息检索工具和系统以及信息检索技术和方法等。信息检索的全过程包括信息存储和信息检索两个过程(见下图)。存储过程就是将原始信息进行分析从而提炼出主题概念,通过信息检索语言对主题词进行标引,从而形成信息特征标识,再将众多信息特征标识进行排序,最终形成信息检索系统。检索过程则是对用户的信息需求加以分析从而提炼出检索主题,同样通过信息检索语言将主题概念转化为检索提问标识,然后在信息检索系统中循序匹配,当检索提问标识与信息特征标识一致或被它包含时,则为命中结果。

## 二、信息检索的原理

### 1、信息检索原理

信息检索的全过程包括信息存储和信息检索两个过程(见下图)。存储过程就是将原始信息进行分析从而提炼出主题概念,通过信息检索语言对主题词进行标引,从而形成信息特征标识,再将众多信息特征标识进行排序,最终形成信息检索系统。检索过程则是对用户的信息需求加以分析从而提炼出检索主题,同样通过信息检索语言将主题概念转化为检索提问标识,然后在信息检索系统中循序匹配,当检索提问标识与信息特征标识一致或被它包含时,则为命中结果。

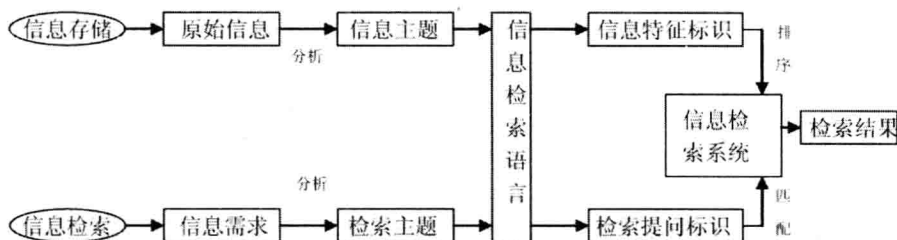


图 1—1 信息检索原理图

一方面是组织有序的文献信息集合,即信息组织,另一方面是用户的信息需求。信息组织是对大量杂乱无章、分散的信息资源进行搜集、整理,并根据信息资源的外部特征和内容特性,按学科、主题、分类、地域、时序等,进行筛选、整理、分析、标引,使其归纳到一个统一的系统中,成为一个有序的检索系统,或称检索工具(目录、题录、索引、文献)。所以,信息组织是存储的前提,是检索的基础,只有实现信息的规范、科学的组织,才能实现信息的有效检索,才能便于利用。

检索就是从用户特定的信息需求出发,对特定的信息采用一定的方法、技术手段,根据一定的线索与规则从中找出相关的信息。

匹配有其匹配标准,这里涉及到两者一致性(存储过程和检索过程必须采用相同的检索标识进行信息标引)、相关度等问题,按一定的标准筛选符合要求(信息特征标识与检索提问标识相一致,或者信息特征标识包含了检索提问标识)的信息。



信息检索的过程往往需要一个评价反馈途径,多次比较匹配,以获得最终的检索结果,存储的过程主要对信息源进行标引,将其外表和内容的特征(如文献的标题、作者、来源和主题等)用特定的检索语言转为一定的标识(如主题词、分类号和类目名称等),再将这些标识按一定的顺序编排后输入检索系统,从而为检索提供“有规可循”的途径。

为了保证文献信息能存得进,取得出,就必须使文献存储所依据的规则与信息检索所依据的规则尽量做到一致。也就是说,为了检索过程的顺序进行达到较高的检索效率,除了在存储和检索过程的各个环节必须依据一定的方法和规则外,还必须有统一的检索语言和名称规范作为存储人员和检索人员的共同依据。

## 2、计算机信息检索原理

计算机检索的核心是数据库,它是在计算机的存储设备上合理存放的相互关联的数据集合。数据库通常由一组相关的文档组成。文档的基本组成单位是记录,记录是作为一个单位来处理的信息单元的有关数据的集合,是对某一信息实体属性进行描述的结果。字段是纪录的下级数据单位,用来描述实体的某一属性。一个字段与一个著录项目相对应,每个字段都可以是一条记录的检索途径。有些字段因内容较多,还可以进一步划分为若干子字段,如主题字段含有若干个主题词,出版字段含有出版地、出版者、出版年等,通常把这样的字段划分成若干子字段。在文献数据库中,记录中含有题名、著者、出版物名称、出版年份、主题词、文摘等字段。

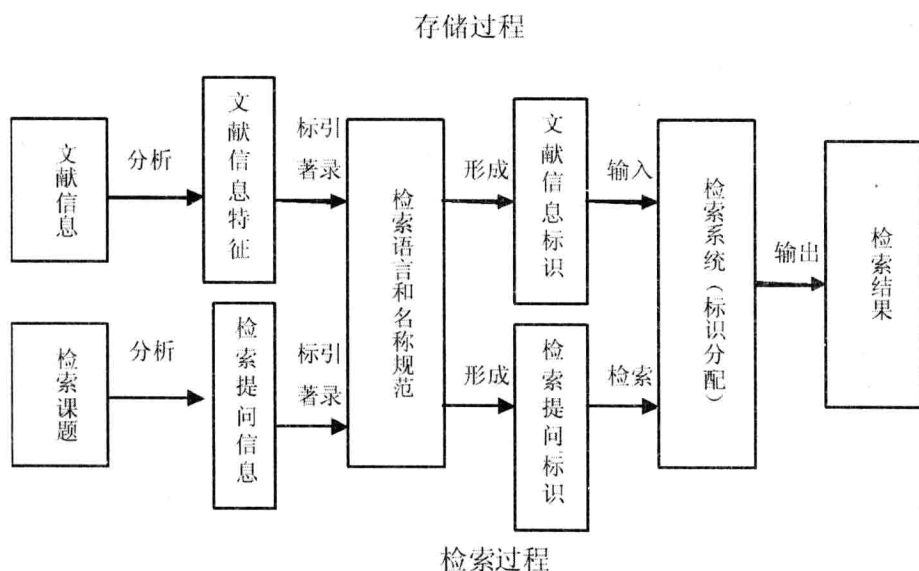


图 1-2 计算机信息检索原理图

计算机检索的过程主要步骤是:用户已知所需文献的特征,直接输入计算机查询,如果用户不知道文献的线索,就要先对检索课题进行主题分析和概念分析,参照系统语言(受控主题表、分类表)确定检索词,或直接用自然语言作检索词,并且明确各检索词间的关系,从而编写出计算机能识别的检索表达式,输入到计算机中。计算机执行检索程序,自动把检索词或检索表达式与数据库中文档的信息特征标识进行比较、匹配、运算,凡信息特征标识