第 3 章　信息检索与利用

教学目标

(1) 了解信息与信息检索的基本知识。

(2) 掌握常用信息检索的使用方式。

(3) 熟悉常用的信息检索资源。

在人类社会的演变和发展过程中，信息活动从来就没有间断过。尤其是 20 世纪以来，随着科学技术的空前进步和信息化过程的深入实施，信息与物质和能量一起构成了现代社会文明的三大支柱。为提高大学生的全面素质以适应信息社会的需要，国家教育部将信息素养教育作为培养人才的重要内容，而信息检索与利用则是实施信息素养教育的必修内容，教学目标是培养大学生的信息意识、信息检索能力、信息吸收能力和信息利用能力，最终使大学生具有信息利用和知识创新能力。

3.1　信息检索系统

获取信息的方式多种多样，简单的可以从日常生活和人际交往中获取信息，复杂的可以从计算机网络系统中检索得到。本节主要介绍信息检索原理，文献分类法，信息检索方法、途径和步骤，文献检索工具系统，数字图书馆。

3.1.1　信息检索原理

1. 信息检索原理

(1) 信息检索概念

信息检索是指为满足用户信息需求而建立的、存储经过加工的信息集合，拥有存储、检索、传送等装备，以便提供存储和检索服务的实体。实质是将用户的检索需求与信息集合中的信息进行比较，如果匹配，则表示信息被找到，否则没有找到。实际上，信息检索就是根据社会需要和信息交流目的而建立的一种有序化的信息资源集合。

信息检索可以分为两种：狭义信息检索与广义信息检索。狭义信息检索就是查找信息，广义信息检索则还包括合理的信息存储和有效的信息检索。其中，信息存储是将搜集到的一次信息，经过著录特征(如著者、分类号、题名、主题词等)后而形成款目，并将这些款目组织起来成为二次信息；信息检索是在已存储好的二次信息数据库进行指定查找。

换言之，合理存储是为了有效检索，有效检索要求存储组织和结构必须合理，二者是相互依存的辩证关系，如图 3-1 所示。

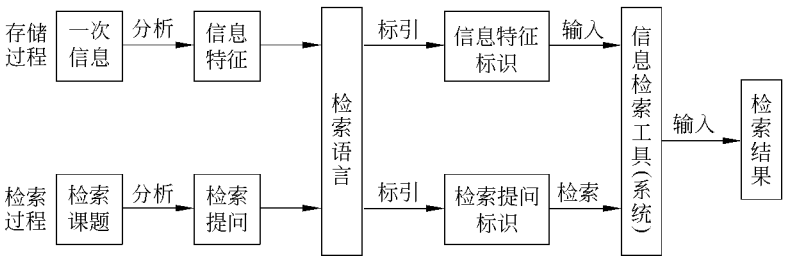

图 3-1 信息检索原理

判断计算机信息检索效果好坏的三个指标是：检索速度、查准率和查全率。其中，检索速度是指检索出信息所消耗的时间，查准率是被检出相关信息量与被检出信息总量的百分比，查全率是被检出相关信息量与相关信息总量的百分比。

(2) 信息检索意义

信息检索的主要意义包括终身学习、获取知识和科学研究。

信息检索是终身学习的前提。学校知识教育的目标是培养学生的自学能力、思维能力、表达能力、研究能力、组织管理能力等。现代教育已经扩大到人的一生，唯有实现终身教育才能成为适应现代社会的公民。通过信息检索可以防止知识老化，完善人的知识结构，较好地融入信息社会。

信息检索是获取知识的捷径。1972 年美国普林斯顿大学物理系的大学生约瀚·菲利普，在图书馆里用 4 个月时间查阅有关制造原子弹的资料，画出制造原子弹的设计图。他设计的原子弹，体积为棒球大小，重量为 7.5kg，威力接近广岛原子弹爆炸的能量，造价为 2000 美元。从这个例子可以看出，通过信息检索可以获取各种有用的知识。

信息检索是科学研究的基础。在美国实施“阿波罗登月计划”过程中，对阿波罗飞船的燃料箱进行压力实验时，发现甲醇会引起钛应力腐蚀，为解决该问题支付了数百万美元。事后，有人通过十多分钟的信息检索就发现前人早在十多年前就解决了这个问题，方法是在甲醇中加入 2%的水。所以，在科研领域有许多科学家在进行重复劳动。据国外资料统计，美国每年由于重复研究所造成的经济损失，约占全年研究经费的 38%。日本在化学化工方面的重复研究比例，大学占 40%、民间占 47%、国家研究机构占 40%，平均重复率在 40%以上。这种重复研究，在我国也是普遍存在的。

(3) 信息检索类型

按照文献出版情况，计算机信息检索可以分为回溯信息检索和定题信息检索。其中，回溯信息检索是根据时间范围进行文献信息检索的方法，通常在着手课题鉴定和专利查新时使用该方法；定题信息检索是根据指定题目进行文献信息检索的方法，具体操作是将用户提问预先存储在存储器中，由检索系统按照提问要求定期检索数据库中的最新文献信息，并将检索结果发送给用户。使用该方法的优点是用户不必进行联机检索，就能定期获取文献信息。

2. 信息检索要素

信息检索的四大要素分别是信息来源、信息意识、信息获取和信息利用。

(1) 信息来源

信息来源是信息检索的物理基础，信息来源的构成如表 3-1 所示。

表 3-1　信息来源

分　类	说　明
文献载体	包括印刷型、机读型、声像型、缩微型等
文献加工程度	包括零次文献、一次信息、二次信息、三次信息、高次信息
出版形式	包括图书、报刊、研究报告、会议信息、专利信息、统计数据、政府出版物、档案、学位论文、标准信息等，其中后 8 种属于特种文献

(2) 信息意识

信息意识是人们利用信息系统获取所需信息的内在动因，这是信息检索的前提，具体表现为人们对信息的敏感程度、选择取向、理解识别能力等。信息意识包含信息认知、信息情感和信息行为 3 个层面，也是人们学习信息知识并运用信息解决实际问题的基础。信息意识的评价内容包括对信息的社会作用与经济价值的认识，对特定信息需求的自我识别，对信息科学的认识，能充分准确地表达出对特定信息的需求。

在现代信息社会中，信息意识至关重要。由于信息浩如烟海，其中凝聚着无数的思想、数据、方法、事实、科研成果、商机等，人们只能通过检索和处理获得信息的真正价值。所以，强烈的信息意识已经成为不可或缺的能力之一。

(3) 信息获取

信息获取是信息检索的目标，要达到这个目标则需要了解各种信息来源，掌握信息检索方法，熟练使用检索工具，对信息检索效果进行正确评估。具体体现在人们对信息存储机构，如图书馆、Internet、各种光盘数据库等的应用能力。

信息获取的评价内容包括：确定信息需求，选择经济的获取信息的方法，采取适当的策略并检索信息，记录与管理信息并能根据情况来适时调整检索策略或检索的信息源，评估检索获取的信息和信息源。

知道信息在什么地方，这本身就是知识中最重要的部分。在信息社会，人们被大量有用或无用信息包围，但又不知道在何处查找所需信息。据美国国家基金会在化学工业部内的调查统计表明，科研人员的全部工作时间分配是：收集信息占 50.9%，实验论证占 32.1%，数据处理占 9.3%，计划与思考占 7.7%。所以，掌握信息获取技术，可以让研究人员以最快速度、最精确的途径获得所需信息。

(4) 信息利用

检索信息资源的目的在于利用，“利用”的终极目标在于创新，“创新”的最佳效果在于贡献。实际上，信息资源是一种再生资源，在工程和科技领域中可以根据不同的目标利用有关的信息。一方面，利用信息将能够扩展视野，避免重复别人的研究工作；另一方面，也能够将已有信息转换成新知识。因此，如何充分利用信息资源并进行科学研究工作是应

该具有的能力。

信息利用的评价内容包括提炼与综合主要观点，提出新结论，对新进行评价，评估信息成果对决策问题的支持程度，将信息成果以适当形式组织到原始问题环境中以解决原始问题，反思信息成果产生过程，有效传播信息成果。

3. 信息加工程度

根据文献加工程度可将信息来源分为零次文献、一次文献、二次文献、三次文献和高次文献，如图 3-2 所示。

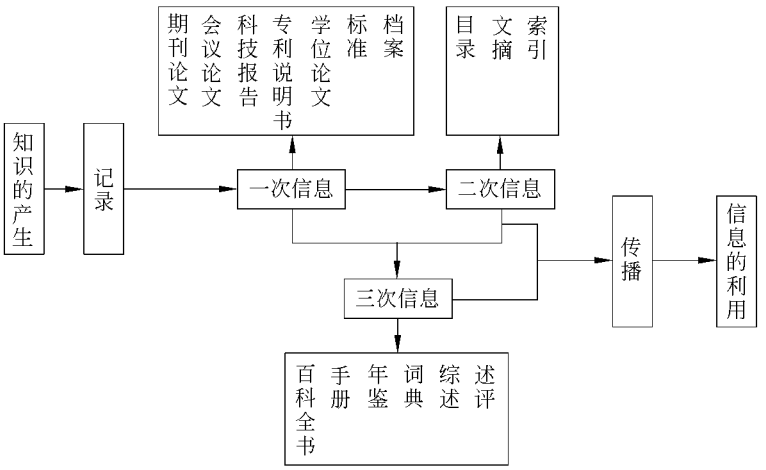

图 3-2 信息来源及其利用

零次文献。零次文献是没有经过系统整理并形成正式文献的零散资料或原始记录，如手稿、笔记、照片、书信、实验数据等。零次文献的主要优点是内容的新颖和原始，缺点是内容欠成熟、结构分散、检索困难等。

一次文献。又称为原始文献，这是人们研究成果的直接记录，如期刊论文、科技报告、会议文献、公开出版的著作、学位论文、发明专利等。一次文献的主要优点是新颖、原创、系统、参考与使用价值较高等，缺点是零碎、分散和无序。

二次文献。这是通过科学方法进行筛选、分析和整理，并按内容特征和外部特征进行提炼浓缩后编制而成的文献，包括目录、文摘、索引等。形成二次文献的过程就是从分散、无序到集中、有序的控制过程。这种文献的主要优点是易于存储与传播、有较高的检索与传递价值等。

三次文献。这是通过系统组织、综合研究与分析一次文献和二次文献的结果，如述评、百科全书、专题报告、年鉴、手册、字典等，优点是实用和技术含量较高。

高次文献。许多文献还要在一次文献息、二次文献和三次文献的基础上进行深入分析和整理才具有创造价值，这就是大量高次文献产生的主要原因。

综上所述，零次文献是最原始的信息资源，尽管没有公开发表，但还是产生一次文献信息的主要素材；一次文献是最主要的信息资源，是信息检索和利用的主要对象；二次文献是对一次文献信息的集中提炼和有序化后的信息资源，是检索一次文献信息的有效工

具；三次文献是按知识门类或专题将文献信息重组和浓缩而成的信息资源，是人们检索数据信息和事实信息的信息资源；高次文献对信息的提炼和浓缩更深入，是人们宏观理解信息并进行创新的重要依据。实际上，信息具有“可再生性”，要充分利用信息资源，信息工作者应该将信息加工处理成不同的等级。

4. 信息出版形式

根据信息出版形式的不同，可以划分为图书、期刊、会议论文、科技报告、学位论文、专利文献、技术档案、技术标准、产品样本、报纸、政府出版物等。

印刷型图书是历史悠久的传统信息类型，它具有全面、系统、成熟、综合等优点。但出版周期过长，传递信息的速度较慢，信息老化速度快。近年来增长迅速的电子图书可以部分弥补印刷型图书的不足。

期刊也是历史悠久的传统信息类型，它具有出版周期短、刊载文献速度快、数量多、内容新等优点，能及时反映某一研究领域的学术发展水平。目前，我国正式出版的期刊杂志近万种。随着电子期刊的大量涌现，用户将可以通过期刊全文数据库进行检索。

会议论文是将学术会议或专业会议上交流的论文编辑出版的信息类型，科学工作者通常通过参加会议或阅读论文来了解学科发展动向。它的优点是内容新颖、专业性强、信息传递迅速等，并能够及时反映某个专业领域的最新研究水平。缺点是参加路途过远的会议很不方便且费用较大。

科技报告又称为研究报告或技术报告或政府报告。其中，科技报告的特点是内容新颖详尽、专业针对性强等，报道科技成果的速度要远远快于会议论文、期刊和图书。但是，部分科技报告是保密或控制发行的，普通用户不容易获取原文，因此它又被称为“灰色文献”，如表 3-2 所示是美国的科技报告；技术报告是记录研究工作和开发调查工作的成果或进展情况的一种信息类型；政府报告是政府部门制定有关科学发展与工业研究的一种信息类型。

表 3-2　美国的四大科技报告

科技报告	说　　明
PB 报告	包括环境污染、土木建筑、城市规划、生物医学等领域
AD 报告	包括物理、航空航天、材料工程技术、军事等领域
NASA 报告	包括航空、空间科学等领域
DOE 报告	包括能源保护、核能、太阳能、矿物燃料等领域

我国收藏科技报告较全面的单位是国家图书馆、中国科技信息研究所、上海图书馆和国防科技信息研究所等。其中，国家图书馆的主页如图 3-3 所示。

学位论文是研究机构和高等院校的学生为获得各种学位而撰写的科学论文，如学士论文、硕士论文和博士论文。但是，学位论文的质量参差不齐，仅硕博士学位论文的参考价值较大。学位论文一般不会出版发行，通常只有学位授予单位（如研究机构和高等院校的图书馆）才会保存。

专利文献又称为专利说明书，是一种将科学技术发明与经济和与法律结合起来的一种信息类型。它是由专利申请人向专利机构提交的说明该项发明的技术原理、优点、专利

图 3-3　国家图书馆的主页

权限等书面文件组成的。当然，专利文献包含丰富的技术情报，含金量极高。据统计表明，全世界高新技术中的 90%以上都是通过专利文献公布发行的。

技术档案一般是通过技术活动和实践形成的一种信息类型，包括某个工程对象的图样、图表、技术文件、照片、原始记录等。它具有内容真实、详尽、准确可靠、参考价值大等特点，可作为科研和工业生产的重要依据。技术档案具有明显的保密性，普通机构和用户不容易获取原文，但不是“灰色文献”。

技术标准是对各种产品、元件、工程建设质量、规格、检验方法等作业进行技术规定的一种信息类型，很明显具有约束性、时效性和针对性。技术标准对于改进产品工艺、提高产品质量、加强市场竞争等，都将起到非常重要的作用。技术标准涉及内容非常多，可按不同的方式进行划分，如表 3-3 所示。

表 3-3　各种技术标准

组　成	说　　明
组成内容	分为产品标准、基础标准、方法标准、安全与环境保护标准等
成熟程度	分为正式标准、指导性技术文件、试行标准、标准化规定等
使用范围	分为国家标准、国际标准、区域标准、部颁标准、行业标准、企业标准等
约束程度	分为强制性标准和推荐性标准两种

产品样本又称作产品目录、产品说明书、产品手册等，它是厂商对指定产品的用途、规格、性能、构造、原理、使用方法等所做的详细说明。它的特点是形象直观、数据可靠、图文并茂等，有助于了解有关工业领域的生产动态和发展趋势，是提高企业形象、进行技术革

新、开发与设计产品、订货等方面不可或缺的信息资源。

报纸是每期版式相同的、以报道新闻与评论为主的一种定期出版物，包括日报、双日报、三日报、周报、旬报等。它的特点是出版周期短，信息传递及时。目前我国正式出版的报纸共计 1943 种，比如成都商报电子版，如图 3-4 所示。

图 3-4　成都商报电子版

政府出版物是政府部门及其机构发布或出版的文献，分为行政性文献和科技性文献两大类。其中，行政性文件包括法令、条约、决议、规章制度、政府报告、会议记录、调查统计资料等；科技性文件包括科普资料、科技政策、科研报告、技术法则等。它的主要特点是具有正式性和权威性，方便人们了解国家的有关科技、经济发展政策以及研究现状，有助于正确地选择课题和科研方向。

3.1.2　文献分类法

1. 文献分类法概念

所谓文献分类法是根据文献内容与形式按照指定体系系统地组织并区分文献的方法。为实现文献检索的快速与高效，文献分类法必须做到科学合理，它的表现形式是各种分类表。这些分类表的组织结构非常重要，最终将确定文献检索的效率。

文献分类法是通过由许多类目根据指定原则组织起来的分类体系，并使用许多标记符号来表示各级类目并固定相应的先后顺序。文献分类法是图书馆和情报部门日常进行文献分类、组织藏书、管理资源等工作的重要工具。具体的文献分类法多种多样，若按文献编制方法划分则有等级列举法、分面组配法和列举组配混合法三种。若按文献内容划

分则有专业分类法和综合分类法两种。

2. 具体文献分类法

中国目前正在使用的文献分类法包括中国科学院图书馆图书分类法、中国图书馆图书分类法、中国图书资料分类法和国际图书集成分类法。

中国科学院图书馆图书分类法。简称“科图法”，由中国科学院图书馆发起并编制，于1958年开始出版发行。科图法将全部文献分为5个基本部类和25个大类，使用阿拉伯数字的标记号码。其中，标记号码又分为两部分：大类与主要类目和细分类目，前者使用00～99的顺序数字，后者使用小数制形式。

中国图书馆图书分类法。简称为“中图法”，由北京图书馆等单位发起并编制，于1975年开始出版发行。中图法将全部文献分为5个基本部类和22个大类，使用汉语拼音字母与阿拉伯数字相结合的混合码标记符号，并按照小数形式进行排列，如表3-4所示。

表3-4 中国图书馆图书分类法(第四版)

字母	说　明	字母	说　明
A	马克思主义、列宁主义、毛泽东思想、邓小平理论	TD	矿业工程
B	哲学、宗教	TE	石油、天然气
C	社会科学总论	TF	冶金工程
D	政治、法律	TG	金属学与金属工业
E	军事	TH	机械、仪表工业
F	经济	TJ	武器工业
G	文化、科学、教育、体育	TK	能源与动力工程
H	语言、文字	TL	原子能技术
I	文学	TM	电工技术
J	艺术	TN	无线电技术、电信技术
K	历史、地理	TP	自动化技术、计算机技术
L	自然科学总论	TQ	化学工业
O	数理科学和化学	TS	轻工业、手工业
P	天文学、地球科学	TU	建筑科学
Q	生物科学	TV	水利工程
R	医药、卫生	U	交通运输
S	农业科学	V	航空、航天
T	工业技术	X	环境科学、安全科学
TB	一般工业技术	Z	综合性图书

中国图书资料分类法。由中国科技情报研究所等单位发起并编制，于1975年开始出版发行。该分类法是在中图法基础上，进行适当修订与细化而编制的，所用的文献体系结构、标记符号等与中图法基本相同。

中国人民大学图书馆图书分类法。简称为“人大法”，由中国人民大学发起并编制，于1953年开始出版发行。人大法将全部文献分为4个基本部类和17个大类，使用阿拉伯数字的标记号码。

国际图书集成分类法。由武汉理工大学图书馆的胡昌志老师发起并编制，他仔细分析了中图法从理论体系到实际运用过程中存在的种种弊端，首次将形式逻辑思想作为文献分类的理论基础，提出文献传承律、交叉学科定位律等分类思想，从而使文献分类从专业领域向大众化方面进行转变。

3.1.3　信息检索的对象和手段

1. 信息检索的对象

信息检索的对象包括文献、数据和事实，对应的信息检索就是文献检索、数据检索和事实检索。

文献检索是以文摘、题录、全文等二次信息为检索对象的检索，用于查找指定的课题、著者、地域、机构、事物等信息以及这些信息的出处和收藏单位，检索结果一定是文献信息，例如检索“设计体育馆的参考文献”。一般分为全文检索和书目检索两种，查找用户指定的文献，如法律文献。

数据检索是利用数据库或参考工具来检索包含在文献中的指定数据、参数、图表、公式、化学分子式等，检索结果一定是数据信息，例如检索“体育馆的承重量?”。一般分为数值检索和数据检索两种，用于查找用户指定的数值型数据，包括统计数据、调查数据和特性数据，如销售额、物品数量等。

事实检索是以一个客观事实为检索对象，对该事物发生时间、地点、过程等进行检索，查找用户指定的描述性事实，如政府机关、企业、事业或人物的基本情况，其中人物情况包括生平事迹、受教育过程等。

综上所述，文献检索属于相关性检索，检索结果只是所需文献的线索，还必须进一步查找才能检索到有关的一次信息；数据检索与事实检索则是确定性检索，检索结果是可供用户直接利用的信息，所以只有获得三次信息才能知道真正有效的“数据”和“事实”。

2. 信息检索的手段

信息检索的主要手段包括手工检索、光盘检索、联机检索和网络检索。这些手段可以概括为手工检索和计算机检索。其中，手工检索是利用印刷型检索书刊(纸质的书刊和资料)来检索信息的过程，主要优点是回溯性好、没有时间限制、不收费等，缺点是费时且效率较低；计算机检索是利用机器检索数据库的过程，包括光盘检索、联机检索和网络检索，如表3-5所示。主要优点是速度快，缺点是回溯性差、有时间限制和费用较高。

表 3-5 计算机检索

检索手段	说　　明
光盘检索	存储媒介使用光盘，由于光盘网络能使用户非常方便地检索大型数据库，所以检索效率得到很大提高。优点是检索费用低、多用户资源共享、数据存取快速、大容量存储、光盘数据库与联机系统并存等
联机检索	存储媒介使用硬盘，并可以利用通信手段将多台计算机连接起来，实现联机检索。优点是计算机和联机通信的广泛应用
网络检索	互联网用户是通过网络搜索引擎或浏览器获取信息的。Internet 的兴起使信息检索方式出现革命性变化，极大地提高信息检索和处理能力。但是，计算机信息检索有较多技巧，需要用户具备较高的“机检”水平

3.1.4 信息检索的方法、途径和步骤

德国柏林图书馆内有这样一段话：“这里乃知识宝库，若您拥有宝库的钥匙，则您将拥有这里的全部知识。”这里所说的“钥匙”就是信息检索的各种方法。

1. 文献检索方法

文献检索方法与文献检索的课题、性质和类型紧密相关，主要方法可以分为三大类：工具检索、追溯检索和综合检索。

(1) 工具检索方法

工具检索方法又称为直接检索方法，即直接利用文献检索工具来查找文献的方法。在选择检索工具时，一方面应该根据课题内容充分利用综合性检索工具；另一方面需要使用专业性检索工具。两者的综合利用可以提高文献的查准率和查全率。根据时间顺序和范围，工具检索方法可以分为 3 种：顺查方法、抽查方法和倒查方法。

① 顺查方法。该方法以检索课题的开始年代为检索起点，按时间顺序由远到近地进行检索，直到检索到所需的文献信息为止。该方法的优点是查全率高，缺点是需要耗费大量的时间和精力。

② 抽查方法。该方法将针对研究课题本身的发展特点，选择学科发展迅速、发表文献多的时期，进行逐年检索的方法。该方法的优点是以较少检索时间获得较多文献信息，缺点是必须非常熟悉学科发展特点，即查全率较低。

③ 倒查方法。该方法按逆时间由近及远地进行文献检索，常用于新课题或具有新内容的旧课题，需要的是近期发表的文献资料，以便掌握最新的研究动向和课题水平。该方法的优点是节约时间，缺点是查全率较低。

(2) 追溯检索方法

追溯检索方法又称为回溯检索方法，即利用已有文献中提供的参考文献，由近到远地进行追溯检索。该方法的优点是直观方便，能够不断追溯并检索到相关的大量文献，并在检索工具的功能弱时能够扩大检索信息源，缺点是查全率较低和漏检率较高。

(3) 综合检索方法

综合检索方法又称为分段检索方法、循环检索方法或交替检索方法，为当前使用最广