



第 1 章 文献检索概述

1.1 文献的基本概念

1.1.1 信息

1.1.1.1 信息的定义

信息是物质的一种普遍属性。信息的定义并不统一，不同领域的学者从各自的专业角度定义信息。在信息检索的概念中，信息主要是指物质的本质，即世界上一切事物的运动状态、规律的表现或者反映。现行国家标准 GB/T 4894—2009《信息与文献·术语》中将信息定义为被交流的知识，涉及事实、概念、对象、事件、观念、过程等，信息也指在通信过程中为了增加知识用以代表“被交流的知识”的一般消息。任何可标识的对象，包括媒介和记录信息的组合、以物质或数字形态表达的知识成果等均可称为信息资源。信息资源是有组织的有序信息，没有经过组织信息，即使达到一定的丰度也不能称为信息资源。

信息存在于自然界、人类社会和思维领域中，它与客观事物共存，是人们认识世界、改造世界中取之不尽、用之不竭的宝贵资源。随着信息社会的发展，信息已经成为重要的战略性资源，是衡量社会科技进步、经济发展、文化和教育水平的标准之一。

1.1.1.2 信息的特性

虽然信息的定义繁杂，但是各类型信息具有一些共同的特性。①客观性与可扩充性：事物存在或运动即会产生信息，因此信息是客观存在的，不以人的意志为转移。随着客观事物的状态和特征的不断变化，可以不断地产生新信息，具有可扩充性。②传递性：信息可以从信息源通过一定载体传递到接收器，因此信息是可传递性的。信息不仅可在人与人、人与计算机之间传递，也可在生物界的植物或动物之间传递。信息通过文献、光盘等载体存储，随时间而传递，称为时间传递；如果通过通信传输系统不断地传递，称为空间传递。③多态性与可转化性：信息可以以声音、文字、图像等多种形式存在，也可以从一种状态转换为另一种状态。④共享性：同一个信息可以同时被许多人在不同的空间共同使用，在共享的过程中还有可能会产生新的信息。⑤可加工性：通过加工整理（包括分类、归纳、精简等），信息可以成为便于识别、效用更高的信息，因此，信息是具有可加工性的。



在新媒体和新技术环境下，信息资源不仅种类丰富、数量庞大，而且其载体形态和传播方式发生了很大变化，数字化、网络化、移动化、碎片化已成为常态，导致的结果是人们对信息资源的检索和获取已变得前所未有的容易，通过轻点鼠标或轻触手机屏幕即可获取大量所需的数据，各种数据库系统还会根据检索结果自动生成各种统计图表。此时，对信息资源的探索、发现与研究远比单纯的信息检索和获取更加重要。信息的探索、发现能力得靠平时的积累，搜索行为多了，自然对各种资源的产生、传播、会出现在什么地方以及各种检索工具和检索方法有所了解。而思考和了解得越多，就越能高效地搜索和挖掘信息内容。

1.1.1.3 信息素养

1974年，美国信息产业协会主席 Paul Zurkowski 提出信息素养的概念：信息素养是一种通过各种信息检索工具获取信息和解决问题的能力。随着信息化时代的发展，信息素养的概念也在改变，但其核心仍是如何理解信息、选择信息、评估信息和交流信息。信息素养是大学学生的必备技能，也是信息社会所需人才的重要能力。

信息素养包括信息意识、信息能力和信息道德，其中，信息意识是信息素养的基础，信息能力是信息素养的核心，信息道德是信息素养的规范性原则。人们尤其是大学生应了解信息的重要性，遵循法制观念和 information 道德规范，快速、有效地运用各种工具以及信息资源获取有价值的信息，并对其加工产生新的信息。

高信息素养不仅体现在“检索能力”上，而且体现在思维方式、思维水平和思维层次上。信息素养教育不单纯是传授信息检索知识、提高信息能力和信息道德，更应该着眼于培养知识的融会贯通和开放灵活的思维方式，学习者可将信息素养的相关知识与其自身的知识结构进行融合，在知识、能力、综合素质三个方面同时提高。

1.1.2 知识

《辞海》2020年版对知识的解释为：人类认识的成果或结晶。知识是在社会实践中形成的，是对现实的反映。根据反映层次的系统性，可以分为经验知识和理论知识。经验知识是知识的初级形态，系统的科学理论知识是知识的高级形态；社会实践是检验知识的标准，知识是人类共同的精神财富。知识借助于一定的形式，如语言、劳动产品等进行传递和交流。国家标准 GB/T 4894—2009《信息与文献·术语》中将知识定义为基于推理并经过证实的认识。

知识来源于信息，属于信息的一部分。信息被有选择地收入人们的思维系统，经过大脑的储存、挖掘、甄别、提炼、加工、处理、转换而形成知识。知识随社会实践、科学技术的发展而发展，是人们对客观世界物质形态和运动规律的认识的总结。知识按照内容一般可以分为自然科学知识、社会科学知识和思维科学知识。知识可以共享、传递和再生，也可以反复使用而不损耗，其价值并不会因为使用而减少。

当前人类科学知识每3~5年增加一倍,生命科学的发展和知识更新速度较其他学科更为迅速。创新是时代发展的必然要求,但也只有在新的起点或者说处在知识的前沿才能谈得上开拓、创新,所以说信息检索是科研的基本功,是科研的第一步,同时也贯穿了科研的全过程。

1.1.3 情报

情报是以各种方式进行传递或者交流的有用的知识,具有明显的目的性和时效性。情报是激活的知识,最开始情报多与战争中敌情的报告有关,后来随着战争的减少,社会经济活动越来越占据主导,情报的概念逐渐泛化,向经济和社会方向延伸,内涵扩大到获得的他方有关情况以及对其分析研究的结果。目前情报的概念仍具有一定的针对性,常用来表示机密的针对性信息。情报是活化的知识,信息要成为情报,一般要经过选择、综合、分析和研究加工过程,即经过知识的阶段才能成为情报。知识只有在被需要的情况下才能成为情报。因此,情报具有传递性、知识性、时效性、针对性和非公开性等特征。

情报已经渗透到人们的经济生活和社会活动中,有目的地收集情报有助于正确决策。按服务对象不同,情报可分为军事情报、科技情报、战略情报、经济情报、体育情报等;按传递媒介或传播方式不同,情报分为口头情报、文字情报、实物情报、声像情报;按传递范围不同,情报可分为大众情报和专门情报;按传递内容不同,情报可分为科技情报、经济情报、体育情报、市场情报和政治情报。按照公开程度不同,情报可分为公开情报、内部情报、秘密情报、机要情报等。

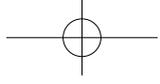
信息、知识和情报三者的关系密切。信息包含知识,知识属于信息的一部分,情报是对交流对象来讲具有重要功能的知识,其逻辑关系是“情报包含于知识,知识包含于信息”,而信息、知识和情报是文献的实质性内容。

1.1.4 文献

1.1.4.1 文献的定义

文献原指典籍与贤者,朱熹注:“文,典籍也;献,贤也。”后专指具有价值或与某些学科相关的图书文物资料,如历史文献、医学文献等。《辞海》2020年版对文献的解释为:记录知识的一切载体的统称,如历史文献、医学文献等。如今,文献的定义为利用文字、图像、符号、声频、视频、网络等手段将人类知识记录在各种载体(如甲骨、金石、竹帛、纸张、胶片、磁带、光盘、网络等)上。

国际标准化组织《文献情报术语国际标准》(ISO/DIS 5127)对文献的解释是:在存储、检索、利用或传递记录信息的过程中,可作为一个单元处理的,在载体内、载体上或依附载体而存储有信息或数据的载体。现行中华人民共和国国家标准《文献



著录总则》GB/T 3792.1—1983 定义文献为：记录有知识的一切载体。

从文献的定义可知，构成文献必须具备 4 个要素，分别为文献的内容，即知识或信息；记录知识的物质载体，如竹帛、纸张、胶卷、光盘等；记录知识的符号，如文字、公式、图表、声音、图像、编码、电磁信息等；一定的记录方式和手段，如书写、印刷、复制、刻录、录音、录像等。由此可见，文献应具备知识性、物质性和记录性这 3 个基本属性。

在收集、整理、交流和利用文献的过程中，对文献的特点、功能、类型、生产方式和整理方法进行研究，总结文献发展规律，逐渐形成文献学这一学科。文献学是一门识别和描述文献的理论、活动和技术的科学。

1.1.4.2 文献的功能

(1) 存贮知识的功能

自古以来，人类认识世界、改造世界所取得的各种知识，主要是靠文献来存贮的。虽然文献不能把人类的知识全部都存贮起来，但是它却能记录保存人类知识的精华。因此，文献一直是人类了解过去、认识现在和预测未来的重要工具。正因如此，文献早已成为存贮人类知识的重要的形式。在漫长的历史长河中，人类积累了大量的文化知识，这是人类的宝贵财富。随着社会的发展和科学的进步，这笔财富正在不断地剧增。怎样保存它，虽然不能全部借助文献，但是长期的实践证明历史上许许多多珍贵的文化遗产，大多数是依赖文献才保存流传至今的。因此，文献是具有保存文化遗产的功能。

(2) 传递和交流信息的功能

文献能记录人类一切精神文明、物质文明的历史和现状，是传递人类社会知识的最佳工具。如果说，古代文献以“藏”为主，那么，当代文献则是以广泛交流为主，因此，传递和交流文献信息是当前图书馆的一项重要使命。

知识是信息的一部分。情报是为解决特定问题而进行交流的知识。文献记录了部分知识。人们通常讲的技术信息、市场信息等信息，已不是自然信息，而是经过人脑加工后进行传递并发生效用的信息（即情报）。

信息、知识、情报、文献之间的关系如下：事物存在即发出信息；信息经人脑加工变为知识；知识被记录形成文献；文献经传递应用变为情报；情报应用与实践产生新的信息；知识经传递应用为情报；情报经过一定时效后，又可还原为知识，情报经提炼，升华还可成为新的知识。

科技工作者在科研工作之前，在制订科研计划时必须查阅文献，并贯穿于科研工作的始终，以便证实工作中所发生的各种问题是否正确，有无错误，特别是实验方法、理论数据、历史事例与前人的研究事例是否相符，这样可以随时改进工作方法，改进仪器设备，使科研工作减少差错，少走弯路，提高成功的可能性，因此文献在促进科研和教学中起着非常重要的作用，推动着科研不断向前发展。

1.2 文献的分类

正是因为有了文献,知识才能得到较好的保存和传播。文献的范围非常广泛,古代的甲骨文、碑刻、竹简、帛书,现在的图书、报纸、期刊、机读资料、缩微制品、电子出版物等都是文献。殷墟甲骨文的重大发现在中华文明乃至人类文明发展史上具有划时代的意义。甲骨文是迄今为止中国发现的年代最早的成熟文字系统,是汉字的源头和中华优秀传统文化的根脉。

根据不同的分类标准,文献可分为多种不同类型。

1.2.1 按文献载体的形式分类

1.2.1.1 纸质型文献

以纸张为知识的存贮载体,通过书写、油印、铅印、胶印、静电复印等手段,将文字固化在纸张上所形成的文献。自印刷术发明至今,印刷型文献一直是最重要的文献类型。印刷型文献是指以传统纸张为载体,通过各种印刷技术制作出来的文献。它一直是文献最主要的存在方式,在图书情报机构的馆藏中占有较大的比例。其优点是便于传递和阅读,缺点是文献体积大,存贮密度低,需要很大的存贮空间,很难实现自动化管理和提供自动化服务。各种图书、期刊、报纸、自传、信件、会议记录、备忘录、笔记等纸质文献都属于此类型。

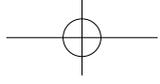
据史料载,现今公认的人类最早的印刷术是中国古代的雕版印刷术,产生于隋末唐初,距今已有1400多年历史。此后,毕昇在此基础上发明了泥活字印刷术。毕昇发明的活字印刷术是对趋于鼎盛时期的雕版印刷术的质的突破,大大推进了中国和世界印刷术的近代化进程,是世界印刷技术史上一项伟大的创举。

1.2.1.2 视听型文献

视听型文献又称声像型文献,它是磁性材料或光学材料为载体,借助特定的机械设备通过录音、录像、摄影、摄像等手段,记录声音、视频图像等信息而形成的文献,如录音带、录像带、电影胶片、唱片、幻灯片、网络视频、激光视盘等。视听型文献的优点是形象直观,人闻其声,物观其形,存储密度高。缺点是制作成本较高,且使用时需要借助一定的设备。

1.2.1.3 缩微型文献

缩微型文献是以感光材料为载体,通过缩微摄影技术将文献的体积浓缩而形成的文献,如缩微卡片、缩微平片、缩微胶卷等。缩微型文献的优点是体积小、存储密度高,不易损坏和变质,保存期较长。但缩微型文献需借助缩微阅读器才能阅读,使用不方便。



1.2.1.4 电子型文献

电子型文献又称机读型文献、数字文献。它以磁、光、电为载体，通过数字代码方式记录图像、文字、声音、影像等信息。电子图书、电子期刊、电子报纸、网络数据库、联机数据库、网络新闻、光盘数据库等都属于电子型文献。电子型文献的优点是一次加工，多次使用，出版周期短，更新快，存贮容量大且存取速度快，节省存放空间，易于复制和分享，是一种很容易应用的文献类型。电子型文献也需要借助计算机阅读，使用费用略高。随着电子计算机和网络技术的发展，电子型文献的主导地位逐渐凸显，各省市、高校图书馆越来越重视电子型文献的收藏，网络数据库的购买数量和质量也成为衡量图书馆水平的标准之一。

数字对象标识符（digital object identifier, DOI）是在数字环境中内容对象的永久性标识号。2007年3月，中国科技信息研究所和万方数据得到授权成立中文DOI注册机构。在已知的DOI号前面添加“<http://dx.doi.org/>”即可检索到原始文献。例如，已知一篇文献的DOI号为10.2108/zs200121，在网页的地址栏输入<http://dx.doi.org/10.2108/zs200121>，即可检索到文献“Two New Brackish-Water Species of *Macrostomum* (Platyhelminthes: Macrostomorpha) from China and Their Phylogenetic Positions”。

1.2.2 按文献的出版形式分类

1.2.2.1 图书

图书通常是分页并形成一物理单元的，以书写、印刷或电子形式出版的知识作品。图书是一类历史悠久的文献类型，反映了某一时期某一学科的水平。图书是对知识系统全面的概括和论述，内容成熟，适用于系统学习知识或对某个问题进行一般性了解。图书包括专著、教科书、工具书、小册子、单卷书、多卷书、丛书等。图书的优点是内容系统、完整、可靠；缺点是出版周期较长，内容更新慢。图书是用于传播知识的一类出版物，有完整定型的装帧形式，具有特定的书名和编者，公开出版的图书有国际标准书号（international standard book number, ISBN）。联合国教科文组织提出图书的页数不包括封面和封底需达到49页以上。国际标准书号是国际通用的图书或者出版物的代码，具有唯一性和特指性，一个国际标准书号仅对应一个或者一份出版物。2017年，ISBN书号从10位升级为13位。《信息检索技术》的ISBN为978-7-302-37465-7，其中包含978为EAN·UCC前缀，表示图书。7为组号，代表国家代码。302是出版者号，37465是出版序号或书序号，最后一个7为校验码或校验号，其作用是检验ISBN后面的数字是否正确。

1.2.2.2 期刊

期刊又称杂志，指采用统一名称，定期或不定期发行的连续出版物，每期内容含有多篇文章，由多个作者分别完成。期刊与图书类似，每个达到国际标准的期刊均具

有专属的国际标准期刊号 (international standard serial number, ISSN)。ISSN 是由国际标准 ISO3297 制定的, 由八位数字组成, 分为两组, 每组四位数, 两组之间用“-”相连, 最末一个数字为检查号。例如, *Nature* 的 ISSN 号为 0028-0836, 《中国科学: 生命科学(英文版)》的 ISSN 号为 1674-7305。经我国国家新闻出版署批准在中国大陆发行的期刊, 具有国内连续出版物刊号, 国内刊号以中国国别代码“CN”作为前缀。国内刊号由 2 组数字组成, 中间以“-”连接, 第一组 2 位数字代表省份或者直辖市, 采用《全国行政区划代码》的前两位表示, 第二组数字为序号, 而最后一个字母代表学科范畴, 根据中图分类号, 字母 Q 代表的就是生物科学。例如, 《中国科学: 生命科学(英文版)》在国内的刊号为 CN11-5841/Q, 《中国细胞生物学学报》的刊号为 CN31-2035/Q, 其中 CN 为中国大陆代码, 11 或者 31 为地区代码, 分别代表北京和上海。

按照出版规律, 期刊可分为定期和不定期, 定期又包括周刊、半月刊、月刊、双月刊、季刊、半年刊等。与图书相比, 期刊因为出版周期短, 所以内容新颖, 报道速度快。此外, 期刊报道的内容涵盖学科范围广、文章种类多且数量巨大, 据统计, 期刊信息占总信息量的 70% 左右。由于上述特点, 期刊已经成为当前科研人员学习和交流学术思想的重要信息来源, 适用于追踪学术前沿、了解新技术或新方法、掌握某一专业领域研究动态等。

1.2.2.3 报纸

报纸与期刊都属于连续出版物, 报纸每期版式基本相同。例如, 《人民日报》包括要闻、评论、理论、国际、综合、专题等栏目。报纸内容多为新闻与评述。报纸定期出版, 周期更短, 很多都是日报, 少数为双日报、周报或者旬报, 因此报纸传递信息更及时。国内报纸也有统一刊号, 比如《人民日报》的刊号为 CN11-0065。报纸信息量繁杂, 也是十分重要的情报信息来源之一。

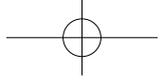
1.2.2.4 会议文献

会议文献是指在国内外各种会议上宣读或者发表的论文或报告。会议论文经过整理后编辑成会议文献集出版, 属于公开发表的论文。会议文献的优点是内容新颖可靠、专业领域比较集中、信息数量大、学术水平高。会议文献适用于了解与自己专业领域相关的学术进展和动态等情况。会议文献是重要的信息源之一。

1.2.2.5 学位论文

学位论文是高等院校或科研院所的本科生和研究生为取得学位资格, 在导师的指导下完成科研设计或实验的学术性较强的研究论文。学位论文分为学士学位论文、硕士学位论文和博士学位论文。根据涉密程度, 学位论文一般可分为公开、内部、秘密和机密四个等级, 保密的学位论文会根据保密等级, 在一定的年限后公开。学位论文主题专一、内容系统而深入, 理论性强, 具有一定的独创性, 所以也是一种重要的信息源。

近年来, 有关大学生学术不端的社会事件屡屡见诸于媒体。2017 年, 教育部公布



的《普通高等学校学生管理规定》规定“抄袭伪造、代写买卖论文将被开除学籍”；2019年，清华大学新修规定“论文抄袭将被开除学籍”。这些举措反映了当前高校大学生的学术不端行为的普遍性和严重性，亟须严肃处理。学术研究是为了追求真理，不可以有半点虚假，然而，近几年科研学术不端行为时有发生，为学术共同体、国家财力和科研人员学术生命带来极大的负面影响。只有研究者自身提高学术诚信的自觉性，树立正确的学术道德观，从内心杜绝学术失信的想法，才能从根本上解决问题。研究者平时要努力提高自己的学术科研实力和学术修养，强化自律意识，端正学习动机，认真学习学术诚信的相关规范和正确的学术科研方法，养成实事求是的学术习惯。

1.2.2.6 科技报告

科技报告是由研究单位或国家政府部门完成的描述研究技术、研究内容、研究进展或成果的特种文献。科技报告独立成册，报告内容专业具体、完整可靠，多具有保密性，报道速度比期刊快。科技报告对科技创新具有举足轻重的作用。美国政府最具代表的科技报告分别是美国国防部的 AD (armed services technical information agency documents) 报告、行政系统的 PB (publication board) 报告、国家航空航天局的 NASA (national aeronautics and space administration) 报告和国际能源部的 DOE (Department of Energy) 报告，被称为美国政府四大科技报告。由于四大科技报告含有较多生物、医学和环境科学的内容，是生物学文献检索不可忽视的来源之一。我国科技报告多由国家或地方财政资金支持的科技计划项目产生，中国国防科技部出版的 GF 报告 (国防科技报告) 是我国科技报告的重要组成部分，科技报告可为落实科技成果、驱动社会和科技发展提供重要参考资料。

1.2.2.7 专利文献

《信息与文献·术语》指出，专利是对发明、实用新型专利或设计给以一个指定时间段保护的工业产权所有权。发明人或专利权人向专利局申请保护时需呈交一份详细的技术说明书，经专利局审查，公开出版并授权。在这一过程中形成的专利说明书、专利公报、专利分类资料、专利检索工具等统称为专利文献，但多数情况下特指专利说明书。只有具有创新性的技术、方法、设计才可能获得专利保护，因此专利新颖性强。通过专利文献可了解某领域的技术发展水平和最新动态，科研项目开题或开发新产品之前，应多了解专利情况，以防重复研究或侵权。专利申请后，国家知识产权局颁发专利申请受理通知，其中包含专利申请号。在专利申请人获得专利授权之后，国家知识产权局颁发专利证书，同时在申请号前面加上 ZL。因此，ZL+ 申请号表明是授权专利。根据保护内容不同，专利可分为发明专利、实用新型专利和外观设计专利。

1.2.2.8 政府出版物

政府出版物又称“官方出版物”，是指各国政府部门及其所属机构出版的信息资料，是基于政府信息的一种有形化和公开化的产物。政府信息或者政府文件经过适当的编

辑整理,制作成图书、期刊或者电子出版物等形式传播给公众即可称为政府出版物。政府出版物一般有出版物编号。政府出版物具有权威性、公开性、公益性、及时性和指令性的特点。政府公报、政报政刊、政府白皮书、法规单行本等均属于政府出版物。按照学科来分,政府出版物可以分为社会科学和自然科学两大类。借助于政府出版物,可以了解国家的科技和经济政策,可以帮助企业和科研人员选择适当的项目和课题。

1.2.2.9 标准文献

标准文献是由某一权威机构或者主管机关批准并颁发的对工农业新产品、工程建设的质量、规格、技术参数及其检验方法所作的各种规范性文件,并具有一定的法律效力,简称标准。标准文献规范了从事产品设计和生产的技术指标,具有计划性、协调性和法律约束性的特点。标准文献有助于产品质量标准化,对提高产品质量、合理利用资源、促进产业创新、推广应用研究成果等有着非常重要的意义。标准文献可以反映一个国家的技术水平和实力。国家标准的编号由国家标准的代号、标准发布顺序号和年号组成。GB代表强制性国家标准,GB/T代表推荐性国家标准。《信息与文献·术语》为现行国家标准,编号GB 4894—2009,由全国信息与文献标准化技术委员会提出并归口,国家图书馆、清华大学图书馆、中科院文献情报中心和中国科学技术信息研究所起草,取代了之前的GB/T 4894—1985《情报与文献工作词汇·基本术语》。标准包括技术标准和管理标准,其中技术标准包括基础标准、产品标准、方法标准等。按照适用范围,标准可又分为国际标准、国家标准、区域性标准和企业标准。

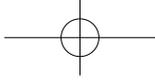
1.2.2.10 产品样本

产品样本又可称为产品目录、产品资料、产品手册或产品说明书,对产品的规格、技术参数、性能构造、原理用途、用法和操作步骤等做出具体说明。厂家和代理商常制作图文并茂、形象直观的产品资料进行宣传和推广,免费赠送、传播给消费者。产品样本是消费者了解产品的性能、用法和优缺点的主要途径,也是客户购买或订货的主要依据,被称为“印刷形式的推销员”,兼具了促销的功能。产品样本具有技术成熟、数据可靠的特点,出版发行周期短,更新速度快,随着产品的更新迭代而被淘汰。

产品样本数量大,内容庞杂,分散在企业或者展会,人人可得,是公开的信息。产品样本能反映产品生产加工水平和发展动向,内容简练易懂,技术成熟可靠,是产品内涵的直观体现。因此,产品样品的收集整理可为企业研发方向提供参考数据和支撑。

1.2.2.11 技术档案

技术档案是科学研究和生产建设部门在从事科研和生产活动中形成的真实记录,例如科技文件、审批报告、技术合同、研究计划、任务书、协议书以及设计图纸、照片、建筑图纸、实验图表、原始记录的原本以及复制件等记录资料。技术档案内容真实可靠,反映生产和科技活动的结果及过程,是科技传承的主要手段,也是后期技术提升和产品维修的依据。



1.2.3 按文献的级次分类

1.2.3.1 零次文献

零次文献大多没有公开发表或者传播，是在科研活动或者工业生产过程的原始记录或者草图等。口头交谈、演讲汇报、发言稿、实验记录、笔记、设计草图、会议记录、书信、内部档案等均属于零次文献。零次文献内容不够成熟系统，但是新颖性较高，具有直观性和很强的针对性。由于零次文献多由私人或单位内部保存，很难通过检索工具获得。

1.2.3.2 一次文献

一次文献又称原始文献或者一级文献，是著者以本人的生产实践经验或科研实验结果为依据撰写的文献。一次文献包括论著、论文、科技报告、会议论文、专利说明书、学位论文、技术标准等。一次文献数量巨大、分散，是主要的科技信息来源。一次文献内容新颖，具有创新性，是科技查新的主要比对文献来源，也可反映一个国家的科学技术水平和科研发展方向。

1.2.3.3 二次文献

二次文献又称检索性文献或者检索工具，是指人们把大量分散、无序的一次文献收集起来，按照一定的方法进行加工整理，用于检索的文献。二次文献并不是新的信息，只是按照学科范围、时间等特征对一次文献进行整理、编排、加工和汇总。主要包括目录、题录、索引、文摘和数据库等。二次文献内容没有新颖性，它只是查找一次文献的工具和线索，可提高一次文献的检索效率，有助于读者快速获得有用的信息，在情报检索和科技查新等工作中具有重要地位。

1.2.3.4 三次文献

三次文献又称参考性文献，是指围绕某一内容、目的或者主题，利用二次文献，对一次文献进行全面系统的筛选、分析、综合和论述的文献。三次文献包括：综述、述评、学科年度总结、参考工具书、百科全书、年鉴、数据手册等。三次文献内容系统全面，具有知识性、概括性和针对性等特点，是一次文献内容的总结与浓缩，具有很好的参考性，有助于读者快速了解某一学科或技术的发展历史与进展状态。

综上所述，随着文献分级的增加，内容逐渐集中、有序。零次文献和一次文献组成了最基础的信息源，创新性强。二次文献将一次文献有序化，是进行信息资源检索的工具。三次文献是以二次文献为手段，对一次文献进行整理与评述，内容高度浓缩，既可以作为信息源，又可以为文献检索提供线索。

1.2.4 按内容的公开程度分类

1.2.4.1 白色文献

白色文献是指正式出版,通过常规渠道可以获得,在社会上公开流通的文献,如图书、报纸、期刊等。文献检索多是针对白色文献资源进行检索。

1.2.4.2 灰色文献

灰色文献是指通过常规的出版流通途径或者一般的检索方法难以获取的文献。灰色文献的公开程度介于白色文献和黑色文献之间。政府机关的内部报告、政策性文件、技术档案、未发表的稿件、书信、实验数据等均属于灰色文献。

1.2.4.3 黑色文献

黑色文献是指非公开发行、内容保密的文献,包括军事情报资料、保密的技术资料等。

1.3 信息检索基础知识

1.3.1 信息检索的概念

文献记录了人类社会活动和各个学科领域的发展状况,大量的文献信息组成了知识的宝库。为了便于进行查找文献和学术交流,一些学术团体和机构创办了检索性刊物,并逐渐形成了系统的检索工具,如目录、文摘和索引等,随着计算机技术的普及,目前以计算机和网络技术为基础的现代化文献检索手段已经占据了文献检索的主流。

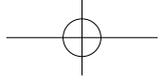
信息检索的概念有广义和狭义之分,广义的信息检索是指将信息按一定的方式组织和存储起来,并根据用户的需要查找出有关信息的过程,包括信息的存储与检索两个过程。狭义的信息检索仅指该过程的后半部分,即用户根据需求从信息集中找出目标文献的过程。

1.3.2 信息检索的类型

信息检索是大学生应该培养的基本技能之一,掌握信息检索可提高人们获取信息的效率。根据检索内容的不同,信息检索可分为文献检索、事实检索和数据检索。

1.3.2.1 文献检索

文献检索是从大量文献集合中,查找出特定的相关文献线索或者全文的过程,即所检索到的是关于文献的信息或文献全文,它的检索对象是文献,结果是导出与某一课题有关的论文、图书及其出处等信息,如所回答的是“关于新型冠状病毒有哪些文



献？”之类的问题。检索结果不直接回答技术问题，只提供与之相关的文献外部信息或者全文。文献检索根据检索的结果可分为书目检索和全文检索。

1.3.2.2 事实检索

事实检索是对事实型数据进行存储和检索的过程，通过对数值型数据、事实、概念、思想、知识等进行查询、运算、比较、演绎和推导，获得关于某一事物发生时间、地点和过程的检索。事实检索直接回答用户问题，如“如何获得转基因拟南芥？”“谁最先完成转基因植物的研究？”

1.3.2.3 数据检索

数据检索是将观察或实验得到的数据，经过筛选、分析、整理和鉴定，存储在某一载体上，然后采用适当的方法或手段从中找出符合用户所需的数据过程，它具有数量的性质，并以数值形式表示。数据检索的结果是确定的，直接回答用户提出的问题，如“第一个胚胎晚期富集蛋白是哪一年被发现的？”

1.3.3 信息检索的意义

随着现代科学技术突飞猛进的发展，文献数量与品种急剧增加，科研课题日趋专门化与综合化，信息检索工作迅速发展起来，在社会中具有越来越重要的地位。信息已经成为人类社会的重要财富，迅速、准确、全面地查找出与研究课题有关的资料，可节约大量时间与人力，避免科研中的重复劳动。

党的十八大提出实施创新驱动发展战略，强调“科技创新是提高社会生产力和综合国力的战略支撑，必须摆在国家发展全局的核心位置。”信息检索是获悉前人成果、了解学术动态的重要渠道之一。在开题前，科研人员需对研究领域进行文献调研和课题查新，分析国内外的研究趋势，避免科学研究工作中的低水平和重复劳动，在前人研究基础上有所突破和创新。从选题、确定方案、具体实施、结题和发表论文，文献检索贯穿了科学研究的全过程。

信息的存贮与检索是信息资源公开和交流传递的基础，促进了科技、经济和社会的快速发展。面对大量无序的文献，快速有效地查找有用的资料可为科研人员、教师和学生节省大量时间。

在整个科技情报工作中，信息检索占有极其重要的地位。科技情报工作的主要内容，就是大量地整理报道现期的和积累检索过去的文献，在此基础上开展定期服务，并进行情报的分析综合研究，以提供对生产、科研、决策、规划工作有用的情报。因此，文献的搜集、研究、存储、检索和传播是科技情报工作的五个环节。把搜集到的情报进行有序的积累（建立检索系统）和有组织地报道（通过文摘、索引、目录等）是情报工作的基本内容，情报工作的重要一环就是文献检索。检索系统的建立是情报工作的一项基础工作。

综上所述,信息检索有利于先进经验和先进技术的迅速推广,摸清科学技术发展水平与动向和加速科学技术前进的步伐。

1.4 检索语言

在文献检索工作中,文献标引工作者和文献检索者之间,需要共同遵守和使用语言词汇。检索语言即为连接信息存贮和信息检索两个过程的桥梁。

1.4.1 检索语言的概念

检索语言又称标引语言、索引语言、文献检索语言、信息存贮与检索语言、概念标识系统等。

检索语言是应文献信息的加工、存贮和检索的共同需要而编制的专门语言,它是表达一系列概括文献信息内容和检索课题内容的概念及其相互关系的一种概念标识系统。文献信息检索人员使用检索语言表达检索课题,有助于准确、全面、迅速地检索系统获得所需要的文献信息。

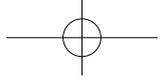
目前,世界上的信息检索语言有几千种,如《中国图书馆图书分类法》《杜威十进位分类法》《NASA叙词表》《汉语主题词表》计算机检索指令等,国际通用的医学相关的检索语言有《医学主题词表》等。我国现有的生物学相关的主题词表有《生物学主题词表》《生物分类叙词表》《林业汉语主题词表》《农业科学叙词表》和《水科学和渔业叙词表》等。

随着生物学科学的飞速发展,新术语和新概念层出不穷,新的主题词表也将逐渐增加。“君子之学必日新,日新者日进也。不日新者必日退,未有不进而不退者。”研究者应进一步丰富已有的知识,努力拓展思维空间,在培养科研能力的同时不忘提升自身的信息素养。

1.4.2 检索语言在文献检索中的作用

非规范化语言(自然语言)的词义有时不明确,常出现“多词一义”或者“一词多义”。检索语言通过对相同或者相关的信息进行系统化和组织化处理,揭示信息的相关性,直接影响信息检索的效果。信息的存贮和检索均需通过检索语言对信息进行标引与表达。因此,信息检索语言需具有必要的语义和语法,一词一义。

检索语言是沟通信息存贮和信息检索两个过程的桥梁,也是沟通标引者和检索者的桥梁,在信息检索中处于核心地位,达到最高的查全率和查准率。通过分析信息存贮和检索的全过程可以看出,检索语言的功能体现在对文献的情报信息内容及其外表



特征加以规范化的标引,保证标引者表述的一致性;对内容相同及相关的文献信息加以集中或揭示其相关性;使文献信息的存贮集中化、系统化、组织化,便于检索者按一定的排列次序进行有序化检索;便于将标引用语和检索用语进行相符性比较,保证标引者和检索者或者不同检索者之间表述的一致性。

检索语言是动态变化的,需要根据社会和科学技术的发展不断地加以更新和改进。检索语言只有在使用过程中可以被长期地及时维护,才能在检索效率方面发挥最佳功能。

1.4.3 检索语言的类型

检索语言是用来标引和检索情报的规范化的人工语言。根据标识的性质与原理或表达文献的特征,可将检索语言分为不同的类型。

1.4.3.1 按照标识的性质与原理分类

(1) 分类语言

分类语言是将各种概念按学科范畴和知识的相关性进行分类,用数字、字母或者符号对类号和类目进行标识的检索语言。分类语言将概念依据学科性质和逻辑层次进行系统排序,可反映事物的从属、派生、交叉和并列等关系。

分类语言的具体表现形式主要是分类表,它由类目表、分类号、类目注释和分类法索引组成。类目表又由基本大类、简表、详表和辅助表组成。目前国际常用的分类法有《美国国会图书分类法》(*The Library of Congress Classification, LCC*)、《杜威十进分类法》(*Dewey Decimal Classification, DDC*)、《国际专利分类表》(*International Patent Classification, IPC*)。我国图书馆分类三大法分别是:《中国图书馆分类法》(原称中国图书馆图书分类法,简称中图法)、《中国科学院图书馆图书分类法》(简称科图法)和《中国人民大学图书分类法》(简称《人大法》)。

《中图法》第一版于1975年出版,是我国各图书馆使用最广泛的分类法体系,1999年正式更名为《中国图书馆分类法》,2000年电子版问世,2010年出版第5版,被《师姐图书及情报机构百科全书》列为世界九大分类法之一。《中图法》包括7种形式的分类法,分别是详本、资料法、简本、中小学儿童馆本、期刊本、索引使用手册。《中图法》采用字母和数字混合标记,将知识分为5大部类和22个大类(表1-1)。

(2) 主题语言

主题语言以描述事物或概念的名词术语作为概念标识,将概念标识按照字顺排列,并用参照系统等方法间接显示概念之间相互关系。主题语言以自然语言的字符为字符,以名词术语为基本词汇,以概念之间的形式逻辑作为语法和构词法,用一组语词作为文献检索标识。主题词的规范化处理就是对同义词、近义词、多义词进行规范。把主题词按照一种便于检索的方式编排起来,就是主题词表,它是主题标引的主要工具。主题词表揭示了同义词、近义词、反义词之间的语义关系,展示了同一族系中各主题

表 1-1 《中图法》(第五版)分类表

1. 马克思主义、列宁主义、毛泽东思想、邓小平理论	
A	马克思主义、列宁主义、毛泽东思想、邓小平理论
2. 哲学	
B	哲学、宗教
C	社会科学总论
3. 社会科学	
D	政治、法律
E	军事
F	经济
G	文化、科学、教育、体育
H	语言、文字
I	文学
J	艺术
K	历史、地理
4. 自然科学	
N	自然科学总论
O	数理科学和化学
P	天文学和地球科学
Q	生物科学
R	医药、卫生
S	农业科学
T	工业技术
U	交通运输
V	航空、航天
X	环境科学、安全科学
5. 综合性图书	
Z	综合性图书

词的等级结构,限定了较含糊主题词的含义或确定其意义与范围。主题语言分为标题词语言、关键词语言、叙词语言和单元词语言。

标题词语言又称标题法,也是最早使用的一种主题语言。标题词语言是从自然语言中选取的、经过规范处理的、表示事物概念的词、词组或短语,是一种规范化的检索语言。标题词按字顺排列,词间语义关系用参照系统显示,并以标题词表的形式体现。如美国《国会图书馆标题表》(Library of Congress Subject Headings, LCSH),它是



国内外用于组织西文图书的主题检索工具，是提供主题检索途径的一种著名标题语言。标题词语言通常使用主标题词和副标题词来构成检索标识，以达到一定的专指度。主标题标示事物本身的概念，副标题说明限制主标题涉及的方面。一个主标题，可能汇集多个不同的副标题。例如：“个体—发育”这个标题中，“个体”一词用于表达事物的概念，是主标题，“发育”作为副标题词，说明它只限于事物的一个方面。以标题词来标引和编制索引的典型检索工具是美国的《工程索引》。

关键词语言是适应目录索引编制过程自动化的需要而产生的。关键词法是从科学技术文献的篇名、摘要、正文中选出具有实质性意义并能表达文献主题内容的科技名词术语来作为检索标识的。这些具有实际意义的科技名词，对揭示和描述文献主题内容来说是重要的，带关键性的词语称为关键词（key words）。关键词未经规范化处理，即对同义词、近义词等不进行严格优选，甚至对同一个词的单数、复数等词形不加统一，即可以保持原形。概括地说，关键词就是将文献原来所用的，能描述主题概念的那些具有关键性的词抽出，不加规范或只作小量的规范化处理，按字顺排列，以提供检索途径的方法。

叙词语言是用规范化科技名词作为基础的主题法检索语言。所谓叙词（descriptor）是指从文献内容中抽出的、能概括表达文献内容基本概念并经过规范化的名词或术语。因为叙词也是标引文献主题的词，故国内也称主题词。叙词语言是专门为文献标引与检索而设计的人工性后组式语言，即文献的主题概念可以用多个叙词表示，因此可以形成任意合乎逻辑的组配和众多的检索途径，更适用于计算机情报检索。用叙词语言编写的词表称为叙词表（或主题词表），常用的叙词表有：《INSPEC词表》《NASA词表》《汉语主题词表》等。叙词语言是多种检索语言的原理和方法的总和，它具有较优越的检索功能，是发展最快，应用最广的检索语言。

单元词语言也是一种规范化检索语言，是在标题词语言基础上发展起来的。它是不能再分解的概念单元的规范化名词（单元词）作为文献主题概念的标识。单元词是从文献内容中抽出来的最基本的、具有独立概念的，即在字面上不能再分的标识文献主题的词，它可进行概念组配。如“肺”“脓肿”两个单元词组配表达肺脓肿。目前实际上已被叙词语言所取代。

（3）代码语言

代码语言是对事物的某方面特征，用某种代码系统来表达和排列事物概念，从而提供检索的检索语言。例如，根据化合物的分子式这种代码语言，可以构成分子式索引系统，允许用户从分子式出发，检索相应的化合物及相关的文献信息。

1.4.3.2 按照表达文献的特征分类

（1）表述文献外表特征的检索语言

表述文献外表特征的检索语言通常是指文献的篇名（题目）、作者姓名、出版者

或研究机构、报告号、专利号等。将不同的文献按照篇名、作者名称的字序进行排列，或者按照报告号、专利号的数字进行排列，这样就形成了以篇名、作者及号码的检索途径来满足读者需求的检索语言。因此，表述文献外表特征的检索语言可以分为题名语言、著者语言、号码语言等。描述文献外表特征的检索语言比较直观，容易理解，以此编制的检索工具（如题名索引、著者索引、报告号索引等）使用起来比较简单，容易掌握。

（2）表述文献内容特征的检索语言

表述文献内容特征的检索语言通常指所论述的主题、观点、见解和结论等。大家都知道，任何一篇文献的内容，无非是论述某个观点事物，任何客观事物都有一定的概念。因此，标引文献内容，进行主题分析后，以形成文献的主题概念。表述文献内容特征的检索语言可以分为分类检索语言和主题检索语言。

1.5 信息检索技术

1.5.1 检索工具

检索工具就是人们用来报道、存储和查找文献情报的工具。它是在不同学科范围内对某一阶段出版的有关文献进行收集、报道并提供检索途径的二次文献。检索工具应具备三个条件：一是详细记录所著录文献线索，读者可根据这些线索查找所需文献；二是根据检索标识，如分类号、主题词、文献号、代码等查找所需文献；三是提供检索的必要手段，如分类索引、主题索引等，便于读者检索。检索工具按加工文献和处理信息的手段可分为手工检索工具和机械检索工具；按出版形式可分为期刊式检索工具、单卷式检索工具、卡片式检索工具、胶卷式检索工具和网络检索工具；按载体形式可分为书本式检索工具、卡片式、缩微式、胶卷式检索工具、光盘式检索工具和网络载体；按著录格式可分为目录型检索工具、题录型检索工具、索引型检索工具和文摘型检索工具。

1.5.2 检索效率

1.5.2.1 查全率

查全率（recall ratio, R）是指从数据库内检出相关文献的量与检索系统中相关文献总量的比率，是测量检索系统检出相关文献能力的尺度，但是系统中相关文献的总量一般是不确定的，只能估计，因此查全率是估算值。查全率受数据库收录文献信息、索引词汇的专指性等因素影响。扩大检索课题的内容、扩大检索途径、放宽限定条件、



英文单词采用截词符号等方式均可提高查全率。

1.5.2.2 查准率

查准率 (precision ratio, P) 又称精确率和求准率, 是指检出的相关文献与检出的全部文献的百分比, 是评估检索成效的一项指标。如果检索结果非全文, 用户很难判断所查到的信息是否与课题相关, 会限制查准率的计算。索引词不准确、组配规则不严谨、选词及词间关系不准确等是影响查准率的主要因素。通过选择主要概念和基本概念、提高主题词的专指度、选择规范专业用语和采用逻辑“与”组配都有利于提高查准率。

查全率与查准率是衡量文献检索效果的主要技术指标。需要引起注意的是, 查全率与查准率是一对矛盾的指标, 当查全率高时, 查准率一般很低; 当查准率高时, 查全率一般很低。因此, 查全率与查准率相互结合使用才能说明系统的检索效果。

1.5.3 现代信息检索技术

信息检索目的即利用检索语言和检索途径达成检索意图。在进行信息检索时, 单独用一个检索词进行检索往往无法获取满意的检索结果, 这时候就需要构造复杂的检索式。现代信息检索技术就是利用当前联机数据库、光盘数据库和网络数据库等检索工具查找目标信息时所采用的技术, 主要有布尔逻辑检索、位置算符检索、截词检索和限制检索等。

1.5.3.1 布尔逻辑检索

所有网络资源系统都支持布尔逻辑检索技术, 常用的布尔逻辑运算有 3 种: 与 (AND)、或 (OR)、非 (NOT)。优先顺序为: 非 > 与 > 或。

(1) 逻辑“与”

逻辑“与”表示概念交叉或者概念限定的组配, 常用算符“*”“AND”或者“并且”表示。如图 1-1 所示, A AND B 表示查找包括 A 和 B 的记录。例如: 植物 AND 动物, 要求植物和动物出现在同一记录中。由此可见, 逻辑“与”缩小了检索范围, 可提高检索信息的查准率。

(2) 逻辑“或”

逻辑“或”是一种具有概念并列关系的组配, 用算符“+”或者“OR”表示。如图 1-1 所示, A OR B 表示查找包含 A 或者 B 的记录。例如: 植物 OR 动物, 要求记录中出现植物、动物两者之一, 或者记录中同时出现植物和动物。用逻辑“或”检索技术可扩大检索范围, 提高检索结果的查全率。

(3) 逻辑“非”

逻辑“非”是一种具有概念排除关系的组配, 常用算符“-”或“NOT”算符表示。如图 1-1 所示, A NOT B 表示查找包含 A 但是不包含 B 的记录。例如: 植物学 NOT 动物,

要求记录中出现植物，但不出现动物。逻辑“非”排除了不需要的概念，因此可提高检索的查准率。

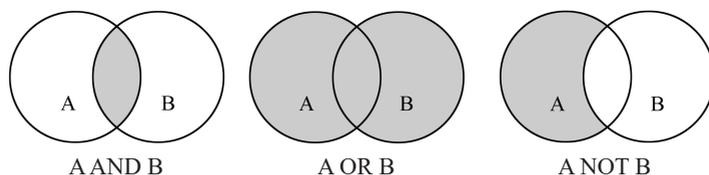


图 1-1 布尔逻辑技术检索示意图

1.5.3.2 位置算符检索

位置算符又名全文查找逻辑算符或相邻度算符，用来规定符号两边的词出现在文献中的位置。位置算符可表示词与词之间的相互关系和前后词序，一次增强选词指令的灵活性，提高检索的查全率与查准率。常见的位置算符有：(W)算符、(N)算符和(F)算符。

(1) (W) 邻近算符

(W)算符是 With 的缩写，A(W)B 表示两个检索词 A 和 B 顺序不能颠倒，A 与 B 之间不能有其他词，但是可以有一个空格或者标点符号或者连接号。

(2) (N) 邻词位置算符

(N)算符是 Near 的缩写，A(N)B 表示两个检索词之间不能有其他词，但是可以有一个空格或者标点符号，A 和 B 顺序可以颠倒。

(3) (F) 邻词位置算符

(F)算符是 Field 的缩写，A(F)B 表示两个检索词必须出现在同一字段内，但是位置和顺序不限，A 和 B 之间可以有任意词。

1.5.3.3 截词检索

截词检索是利用检索词的词干或者不完整的词形进行检索。因为英语单词常存在各种单复数和动词词尾变化等形式变形，截词检索可有效避免由此引起的漏检，提高文献查全率。常用的截词符号有“?”和“*”。其中“?”代表有限截词，代表 ≤ 1 个字符。例如 plant? 可检索出含 plant 或 plants 的记录，plant?? 可检索出含 plant 或 plants 或 planted 的记录。“*”代表无限截词，在绝大部分检索系统中都使用这一截词技术。如 plant* 可检索出含 plant 或 plants 或 planted 或 planting 的记录。

截词技术可分为四种，分别是后方截词、前方截词、前后截词和中间截词。

(1) 后方截词

后方截词又称前方一致检索，是指将截词符号放在检索词的右边。前面列举的 plant? 和 plant* 均属于后方截词。

(2) 前方截词

前方截词又称后方一致检索，是指将截词符号放在检索词的左边。例如：?biotic



可检索出含 abiotic 或 biotic 的记录。

(3) 前后截词

前后截词又称中间一致检索,是指将截词符号同时放在检索词的左边和右边。例如:
* 检索 * 可检索出含文献检索技术或信息检索技术或信息方法等词汇的记录。

(4) 中间截词

中间截词又称前后一致检索,是指将截词符号放在检索词的中间。例如:
summariz?e 可检索出含 summarize 或 summarise 的记录。一般来说,中截词仅使用有限截词,不可使用无限截词。

1.5.3.4 限制检索技术

限制检索是指检索系统提供的,通过附加检索条件减少检索结果的一种检索技术。通常对篇名、文摘、作者姓名、文献出版日期等进行限制检索。

(1) 字段限制检索

字段限制检索是将检索词在特定的字段中进行检索,同一个检索词在不同的字段中进行检索,获得的文献数量不同。检索系统常默认为在“所有字段”也就是全文中进行检索。例如,如果中国知网选择在全文检索倪嘉缙,检索结果共 2273 条,如果限制在作者检索倪嘉缙,检索结果为 358 条,如果限制第一作者检索倪嘉缙,检索结果为 13 条。

(2) 范围限制检索

检索系统一般提供对检索文献的范围进行限制,如限定文献的发表时间、国别、语种和文献类型等。例如,在知网上搜索“新型冠状病毒”,一共可获得 2.32 万条结果,如果限制文献类别为学位论文,则有 206 条结果,如果进一步限制 2020 年发表的文献,则为 44 条结果。

1.6 文献检索的策略

1.6.1 文献检索的方法

1.6.1.1 常用法

常用法又称直接法,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

顺查法直接按照时间的顺序,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,它是用于较大课题的文献检索。例如,已知