



中国劳动关系学院“十三五”规划教材

21世纪 信息资源管理系列教材

Literature Retrieval and Review Training Course

文献检索与综述 实训教程

刘泰洪/编著

非
外
借

中国人民大学出版社



中国劳动关系学院“十三五”规划教材

21世纪 信息资源管理系列教材

Literature Retrieval and Review Training Course

文献检索与综述 实训教程

刘泰洪/编著

中国人民大学出版社

· 北京 ·

图书在版编目 (CIP) 数据

文献检索与综述实训教程/刘泰洪编著. —北京: 中国人民大学出版社, 2018. 2
21 世纪信息资源管理系列教材
ISBN 978-7-300-25436-4

I. ①文… II. ①刘… III. ①信息检索-教材 IV. ①G254.9

中国版本图书馆 CIP 数据核字(2018)第 008273 号

中国劳动关系学院“十三五”规划教材

21 世纪信息资源管理系列教材

文献检索与综述实训教程

刘泰洪 编著

Wenxian Jiansuo yu Zongshu Shixun Jiaocheng

出版发行 中国人民大学出版社

社 址 北京中关村大街 31 号

电 话 010-62511242(总编室)

010-82501766(邮购部)

010-62515195(发行公司)

网 址 <http://www.crup.com.cn>

<http://www.ttrnet.com>(人大教研网)

经 销 新华书店

印 刷 北京溢漾印刷有限公司

规 格 185 mm×260 mm 16 开本

印 张 12.75

字 数 288 000

邮政编码 100080

010-62511398(质管部)

010-62514148(门市部)

010-62515275(盗版举报)

版 次 2018 年 2 月第 1 版

印 次 2018 年 2 月第 1 次印刷

定 价 35.00 元

版权所有 侵权必究

印装差错 负责调换

作者简介

刘泰洪，中国劳动关系学院公共管理系副教授，2008年北京师范大学行政管理专业博士毕业。长期为中国劳动关系学院的行政管理专业、劳动与社会保障专业、行政与政治学专业教授文献检索与综述课程。在《天津社会科学》《人文杂志》《改革》等期刊公开发表学术论文30余篇。主持完成国家社科基金青年课题1项，主持完成教育部人文社科规划课题1项，主持完成北京市教育科学规划课题1项；作为主要成员参与国家社科基金课题、省部级课题多项。

内容简介

本书介绍了文献信息检索基础、文献的加工与管理、网络信息资源检索、中文全文数据库、电子图书、特种文献检索，以及文献综述与写作、学术规范与论文写作等知识。本书结合公共管理学科的常用检索词与检索特点，对文献检索过程进行介绍，着力提升学生的信息检索能力；同时本书将文献的检索过程与文献综述的写作进行结合，能使学生的动手与动笔能力都得到锻炼。



前 言

现代社会是信息社会，各学科、专业乃至社会上各行各业的发展中，信息都起着举足轻重的作用。“文献检索与综述实训教程”是根据教育部的文件精神开设的高等院校专业实验课程。该课程以文献信息资源及其相关检索系统的特点与使用方法为教学重点，其目的是使学生获得一定的文献信息搜集、整理、加工与利用能力。课程旨在增加学生的信息意识，提高其文献资源的获取能力，培养其独立的检索能力与论文撰写能力，为其以后的职业发展以及知识更新奠定良好的基础。

在学术研究的过程中，研究由问题开始，但研究问题的确定是一个复杂的思维过程。文献检索往往是发现学术问题的起点，并且检索贯穿学术研究的全过程。文献检索与论文写作相辅相成。基于上述原因，本课程融理论、方法、实践于一体，由两个部分组成，第一部分是文献检索，第二部分是综述写作，将“文献检索”与“文献综述写作”置于同一实验课程的教学框架中。通过第一部分的学习，使学生能够根据实际需求灵活运用各种检索技术，查询及获取所需要的文献资料；在具备了文献检索知识的前提下，通过第二部分的学习，掌握文献综述写作的基本技能，如写作前的准备、写作过程中的材料组织和书写规范，以及后期论文润色和投稿阶段的注意事项等。全书结构清晰，内容由浅入深、循序渐进，教材中附有丰富的公共管理专业检索案例，有较强的针对性和实用性。

本书的出版，是编者基于多年“文献检索与综述”实验课程教学的经验总结。北京建筑大学图书馆詹宏伟老师撰写了本书第二章的内容。由于时间紧迫，编者水平有限，书中不足与疏漏之处难免，敬请读者批评指正。在教材编写的过程中编者参考了大量国内外文献资料，他们的智慧是本书得以完稿的基础，在此向作者致谢。书中引注与书末参考文献难免有疏漏，敬请谅解。

感谢中国劳动关系学院学术委员会将本教材列为学校“十三五”规划教材予以立项资助，为本书的出版提供了经济支持。感谢中国人民大学出版社盛杰编辑及相关工作人员为本书出版付出的辛苦。

编者

2018年1月8日



目 录

第一章	文献信息检索基础	1
第一节	文献概述	1
第二节	文献信息检索	6
第三节	文献检索语言	9
第四节	计算机检索基本知识	18
第二章	文献的加工与管理	24
第一节	文献整序	24
第二节	图书馆文献信息资源利用	26
第三节	数字图书馆	33
第三章	网络信息资源检索	42
第一节	网络信息资源	42
第二节	网络检索工具	51
第三节	搜索引擎	55
第四节	学术搜索引擎	62
第四章	中文全文数据库	71
第一节	中国知网数据库	71
第二节	万方数据	84
第三节	维普信息资源系统	91
第四节	人大复印报刊资料数据库	101
第五章	电子图书	108
第一节	电子图书概述	108
第二节	读秀学术搜索	110
第三节	超星电子图书	118

第四节	方正 Apabi 电子图书	122
第六章	特种文献检索	126
第一节	会议论文检索	126
第二节	学位论文检索	131
第七章	文献综述与写作	139
第一节	文献综述概述	139
第二节	文献综述的写作	143
第三节	文献综述范文举例	155
第八章	学术规范与论文写作	170
第一节	学术规范	170
第二节	学术论文的写作	175
	参考文献	192



第一章 文献信息检索基础

第一节 文献概述

一、文献的定义

人类在改造自然和变革社会的伟大斗争中，需要各种知识。要获得这些知识，除了直接实践外，更多的是要学习和吸取前人的经验、思想和知识，即充分利用文献。文献有哲学、社会科学文献，也有自然科学文献。“文献”一词最早见于《论语·八佾》，南宋朱熹《四书章句集注》认为“文，典籍也；献，贤也”。这里的“文”指典籍文章，“献”指古代先贤的见闻、言论以及他们所熟悉的各种礼仪和自己的经历。《虞夏书·益稷》也有相关的引证说明“文献”一词的原意是典籍与宿贤。近现代的一些工具书又将其解释为“具有历史价值的图书文物资料”和“与某一学科有关的重要图书资料”。例如：《辞源》解释为“文献：……后指有历史价值的图书文物”；《辞海》解释为“文献：……今专指有历史价值的图书文物资料”；《现代汉语词典》的解释是“有历史价值或参考价值的图书资料”。

随着现代科学技术的发展，文献的生产和交流呈多样化态势，人们对文献的认识已不再停留在传统书籍、档案、典章、史料的藩篱里，而是集中在对文献本质属性及其构成要素的研究上。现代学者对“文献”一词有了新的注释和诠释。

国际标准化组织《文献情报术语国际标准》(ISO/DIS 5217)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。”我国1984年实施的中华人民共和国国家标准《文献著录总则》关于“文献”的定义是：“文献：记录有知识的一切载体。”文献不但包括现代图书馆的全部馆藏，也包括档案馆、博物馆等所收藏的全部馆藏。

随着社会的发展、研究的不断深入，以及现代科技的飞速发展，记录知识的载体和手段都发生了巨大变化。记录知识不再仅仅凭纸张这种单一载体，出现了各种各样的载体材料，包括胶卷（片）、磁带、磁盘和光盘等；产生了各种各样的记录信息和知识的方式，

不再仅仅是刻、写和印刷了，还包括录音、录像和其他数字化复制方式，多媒体的出现就是这种方式的较为集中的体现。

综上所述，我们认为，文献是用文字、图形、符号、声频、视频等技术手段记录人类知识的载体，或理解为固化在一定物质载体上的知识。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、传播和继承知识的最有效手段，是人类社会活动中情报的最基本、最主要的来源，也是交流传播情报的最基本手段。正因为如此，人们把文献称为情报工作的物质基础。在国内国外，常常可以看到有人把“文献”与“情报”、“文献学”与“情报学”等同起来，虽然这种等同未必适宜，但却反映了文献在情报活动和科学中的极为重要的地位。

从上述定义中可以看出文献包括4个要素。

- (1) 知识和信息，即文献的内容。
- (2) 物质载体，如竹简、纸张、胶片、磁盘等，是文献的外在形式。
- (3) 记录知识和信息的符号，如文字、图表、声音、图像等。
- (4) 记录的方式或手段，如书写、印刷、复制、录音、录像等，即将知识和信息固化到载体上的手段。

二、文献的属性

文献的内容最为重要，是文献的内涵和实质所在，是判定文献价值的主要依据。同时，物质载体形式、记录符号系统和文献的发展演变同样至关重要。

1. 知识传承性

文献信息是文献的本质属性，没有信息性，也就不会有文献。任何文献都传递或记录一定的信息知识，传递信息和记录知识是文献的基本功能。人类文明正是靠古今文献的不断继承才得以保存到现在。

2. 物质载体性

文献的存在一定要依附于某种物质载体，载体是文献内容的外壳，其材质随人类科学技术的发展不断演化。从古至今出现过甲骨、金石、简牍、兽皮、泥板、缣帛、纸张、胶片、音像、磁盘、光盘等类型。

3. 记录多样性

文献里的知识信息是通过各种人工方式将其记录在载体上的，而不是天然赋予在一定的物质实体之上的。人们采用各种记录手段，如写、刻、印、录等方式积累知识，传承人类文明，形成了形式多样的文献类型。

4. 发展变化性

文献从出现到现在始终处于运动状态，并且随着人类社会的科技进步和记录水平的提高不断发展演进。进入信息社会以后，信息交流更加频繁，文献的数量日益庞大，载体形式越发多样。同时，新的问题相继出现，文献的更新速度在加快，生命周期在缩短，发展变化速度难以预测。

三、文献的类型

文献作为知识的载体，科学的媒介，自然少不了类型划分。如果没有类型区分，图书馆里的数以万计的文献会乱得像一锅粥，许多的文献便失去了利用的价值。根据不同的划

分标准,文献可以分成多种类型。

1. 按照文献的出版类型划分

文献的类型大致可以分为图书、期刊、研究报告、专利文献、科技报告、学位论文、会议文献、标准文献、科技档案、产品技术资料、政府出版物、报纸、声像资料、电子出版物、网络出版物等。

(1) 图书。图书是以印刷方式单本刊行的,内容比较成熟、资料比较系统、有完整定型的装帧形式的出版物。图书可分教科书,科普读物,辞典、手册、百科全书等工具书。联合国教科文组织对图书的定义是,凡由出版社(商)出版的不包括封面和封底在内49页以上的,具有特定的书名和著者名,编有国际标准书号(ISBN),有定价并取得版权保护的出版物称为图书。图书的特点是内容较系统、全面、成熟、可靠,但出版周期较长,报道速度相对较慢。

(2) 期刊。期刊也称杂志,是由多位作者撰写的不同题材的作品构成的定期出版物。期刊有固定刊名,以期、卷、号或年、月为序。期刊是定期或不定期连续出版的印刷读物。在我国,期刊出版单位出版期刊,必须经新闻出版总署批准,持有国内统一连续出版物编号(ISSN)。期刊内容新颖,报道速度快,信息含量大,是传递科技情报、交流学术思想最基本的文献形式。期刊情报占整个情报源的60%~70%,它与专利文献、科技图书三者被视为科技文献的三大支柱,是检索工作中利用率最高的文献源。

(3) 学位论文。学位论文是学术论文的一种形式,是指学生为了获得所修学位,按要求所撰写的论文。学位论文有严格的格式要求,一般不公开出版,分为学士论文、硕士论文、博士论文3种。学位论文是就某一专题进行研究而做的总结,多数有一定的独创性。学位论文是非卖品,除极少数以科技报告、期刊论文的形式发表外,一般不出版,目前国内已有中国知识资源总库(CNKI)的中国博士学位论文全文数据库、中国优秀硕士学位论文全文数据库,万方数据公司的学位论文数据库等,可供查找学位论文使用。

(4) 会议文献。会议文献是指在会议等正式场合宣读的首次发表的论文,往往代表着某一领域内的最新成就。其主要特点是传播信息及时、论题集中、内容新颖、专业性强、质量较高,但其内容与期刊相比可能不太成熟。会议论文属于公开发表的论文,一般正式的学术交流会议都会出版会议论文集。

(5) 报纸。报纸是一种出版周期最短、发行量最大的出版物。报道内容极为广泛,和人们生活息息相关,是人们日常生活中最常接触的信息源。报纸是每日、每周或每隔一定的时间发行的一种连续出版物,以刊载新闻为主,包括评论文章、特写、广告和其他内容的文章,是重要的社会舆论工具和大众传播工具。报纸按范围级别分,有全国性报纸和地方性报纸;按内容性质分,有综合性报纸和专业性报纸;按出版时间分,有日报、双日报、周报和月报等。联合国教科文组织把每周至少出版四次的报纸称为日报,把每周出版三次及以下的报纸称为非日报。报纸的特点是出版速度快,发行数量大,传递迅速及时,读者对象面广,能以最快的速度报道世界各地发生的最新事件。报纸的信息具有极强的时效性,信息量大,但这也是造成报纸查找不便的原因。

(6) 专利文献。专利文献是指发明人或专利权人申请专利时向专利局所呈交的一份详细说明发明的目的、构成及效果的书面技术文件,是经专利局审查,公开出版或授权后的

文献。广义的专利文献还包括专利公报(摘要)及专利的各种检索工具。专利文献主要由专利说明书构成,内容比较具体,有的还有附图,通过它可以了解该项专利的主要技术内容。由于只有符合新颖性、创造性和实用性的发明创造才能获得专利权,所以专利说明书对于工程技术人员,特别是产品工艺设计人员来说,是一种切合实际、启迪思维的重要情报源。专利文献范围广泛,出版迅速,格式规范,文字简练、严谨,有助于科技人员借鉴国际先进技术,避免重复劳动并节省科研经费。

(7) 标准文献。标准文献是标准化组织或有关机构对产品、服务等的质量、规格、生产过程及检验方法等所做的技术规定,是不同生产、服务厂商应共同遵守的规范性文件。技术标准按审批机构和应用范围划分,有国际标准、区域性标准、国家标准、部颁标准、企业自订标准等,内容包括各种基础标准、产品标准和方法标准等。

(8) 政府出版物。根据联合国教科文组织的规定,政府出版物是指根据国家机关的命令,并且由国家负担经费而出版的一切记录。政府出版物是各国政府所属各部门发表、出版的文献的总称,其内容极为广泛。

2. 依照文献加工层次划分

由于文献出版量的急剧增加,再加上文献类型的多样化和分布的离散状态等因素,信息用户要想准确而快捷地从大量无序的原始信息中获取相关信息,就如同大海捞针一样困难。为了便于人们利用文献,文献根据加工深度的不同主要分为以下3种类型。

(1) 一次文献。一次文献又称原始文献,是人们对已创造的知识进行第一次加工(固化)而形成的文字记载。一次文献具有创造性、原始性、多样性的特点。

(2) 二次文献。二次文献是对一次文献进行加工、整理后形成的产物,它对一次文献的特征,如题名、责任者、出处、分类、主题等进行揭示和排序,或将一次文献的内容压缩成文摘。二次文献是用来报道一次文献的,类型有目录、索引、文摘等。二次文献具有集中性、工具性、系统性的特点。

(3) 三次文献。三次文献是按照特定的课题,利用二次文献选择有关的一次文献加以分析、综合而编写出来的文献类型,主要有综述、述评、专题报告、可行性报告、数据手册等。三次文献具有综合性、科学性、针对性的特点。

总之,从一次文献、二次文献到三次文献,文献形式的变化反映了文献由分散到集中,由无序到有序,由博而精地对知识信息进行不同层次的加工的过程。它们所含信息的质和量是不同的,对于改善人们的知识结构所起的作用也不同。一次文献是最基本的信息源,二次文献是一次文献的集中提炼和有序化,三次文献是把分散的一次文献、二次文献,按照专题或知识的门类进行综合分析加工形成的成果,是高度浓缩的文献信息。

3. 按照载体(外在)形式划分

(1) 刻写文献。刻写文献是雕版和活字印刷发明之前的历史文献和没有排印的文字记录类型。先人用手写或刻字的方式记录下来的文献,是最早的文献载体形式。如我国古代刻写于龟甲兽骨上的甲骨文献,刻铸在青铜器或石头上的金石文献,书写在竹、木片上的简牍文献,记录于丝织品上的缣帛文献,印刷术发明使用前的古写本和印刷术发明使用后的写本、抄本、稿本等纸质手写文献。这类文献往往具有重要的历史价值和文物价值。

(2) 印刷文献。印刷文献是以纸张为载体,采用雕版刻字、活字排印、石印、铅印、油印、胶印、影印、复印等方法形成的文献,是典型的传统文献形式,也是现代文献的主要形式。其优点是可直接、方便地阅读,阅读者可以随心所欲地在自己的书本上做各种阅读记号,可以在书本的眉上行间、篇头卷尾进行评点、批注,记录心得,并可长期保存、反复研读。还因其学术内涵、编著者的名望、独特的装帧形式、出版年代等因素,给某些收藏爱好者提供了实在的满足。其不足是体积大,存储密度低,纸张容易损坏变质。

(3) 缩微文献。缩微文献是指缩微复制品。它是以感光材料为载体,以缩微摄影为记录手段,将手写和印刷型文献缩微而形成的文献形式,包括缩微胶卷、缩微胶片等。缩微文献的优点是体积小,存储密度高,节省储藏空间,价格便宜,便于保存、转移、传递与检索,可以说是印刷型的变体。缺点是阅读时须用阅读器。

(4) 声像文献。声像文献又称直感型或视听型文献,是以磁性材料或感光材料为载体,借助特殊的机械装置,以声音和图像形式记录在载体上的文献。如唱片、录音带、幻灯片、电影片、录像片等。它脱离了或基本脱离了文字形式,只利用视觉、听觉直接接受文献信息,直观、生动,传播速度快,便于更新。但对于常规的书本学习,作为经常性的使用手段,它代替不了纸质印刷文献。

(5) 数字文献。数字文献是以数字代码方式将图、文、声、像等信息记录在磁性材料或光学材料上,由计算机或具有类似功能的设备输入和输出的文献形式。主要包括软磁盘、光盘、集成电路卡等。数字文献存储密度高、存取速度快,特别是磁盘,存储容量极大,传输速度极高,并可在网络上使用,使文献信息交流跨越时空的障碍,加速了知识信息的传播和利用。

(6) 网络文献。网络文献是借助信息技术而存在于互联网上的形式比较特殊的文献。随着科学技术的进步,网络文献作为一种新的信息载体,具有内容丰富、信息涵盖量大的特点;形式多种多样,有文本型的,有声音型的,有图像型的;分布广泛,检索途径多。它能够跨越时空的阻隔,使其所负载的信息内容既可世代地传递下去,也可在不同国家、不同民族、不同地区之间进行传递,成为信息时代联系世界和沟通全人类思想的纽带。^①

四、文献的社会作用

文献是人类思想与智慧的结晶,是人类社会文明发展到一定程度的产物,是人类文化的知识内容与载体材料有机结合而成的一种宝贵的智力资源。

1. 文献是记录和保存人类知识的宝库

文献是人类共同财富,是人类物质文明和精神文明发展的记录;是人类有史以来所从事的生产实践活动、科学技术实践活动和人们的相互交往活动的真实写照;是人类在一切活动中成功经验和失败教训的系统积累和总结;是人类文明和文化知识的最佳表现方式。文献只有被记载在某种物质载体上,才能长期保存。文献只有不断地积累,才能被今人和后人利用,从而创造出新的文明和文化。所以,文献是汇集和保存人类精神财富的最佳载体,是供全人类分享利用的知识宝库。

^① 赵大志. 地方文献建设研究. 成都: 西南交通大学出版社, 2012: 12-13.

2. 文献是获取和传播知识的重要媒介

在文字出现前的远古时代,人类的生产力水平极为落后,各方面都不发达,所获得的经验或知识也很少,人们只要用自己的大脑记忆就可以了,并不需要考虑是否重复以往的经验。后来,随着生产的发展,人们社会交往的增多,各种知识不断产生,需要记忆的知识也随之增加,光靠大脑既记不住也容纳不了,而且记住的东西也容易忘记,于是就产生了把这些经验或知识记录下来的想法。在纸发明以前,古代人只能在甲骨、简牍、缣帛上做记录。在雕版印刷发明以前,古人只能凭手工抄写来记录文献。正是在文献的初级阶段经验积累的基础上,古人才发明了纸与雕版印刷术,这使文献的记录方式更为便利,传播的范围更广,速度更快。人们又从文献中汲取、利用知识贡献于社会,从而极大地推动了社会文明的发展。人类的知识和文明只有通过文献才能突破时空的局限,在时间和空间上得以传播和交流,获取各个时期和各个地区的各种信息资料。可见,文献正是通过其知识获取和传播,推动人类创造发明,促进社会向前发展。

3. 文献是人类文明与进步的重要标志

文献的内容反映了一定社会历史阶段人类的知识水平。人类认识社会与自然界的各种知识,主要是通过文献的积累、总结、储存与提高来实现的。对文献的学习、继承与创造性运用,又反作用于社会,成为推动社会向前发展的有力因素。文献是记载前人成果的重要手段之一,是人类文化科学知识的体现,是创造性活动的基础,可以帮助人们认识客观事物,拓展思路,开阔视野,丰富知识。文献的质和量是某一学科领域、某一个人、某一群体乃至某个国家的学术水平和成就的重要标志。文献也是预测和决策的重要工具。对有关文献进行定量和定性分析,可以预测该事物的发展趋势,从而为决策提供依据。

据上所述,文献对人类文明、社会进步至关重要。正如英国科学哲学家卡尔·波普尔的思想实验得出的结论,如果我们人类所有的机器和工具都被破坏了,而图书馆的文献还存在,那么人类仍然能够重新发展起来;如果文献同所有的机器和工具一起都被破坏了,那么人类文明的重新出现就会是几千年以后的事了。卡尔·波普尔的思想实验生动而深刻地揭示了文献对我们人类的重大作用。

第二节 文献信息检索

一、文献信息检索的概念

文献信息检索中的“信息检索”一词出现于20世纪50年代。文献信息检索(Information Retrieval)亦称为情报检索,是指将信息按一定的方式组织和存储起来,并根据用户的需要找出有关信息的过程。即根据一定的检索目的,选择有关的检索工具和检索系统,按照一定的检索途径、方法和步骤,在海量文献信息源中迅速准确地查找所需的文献信息。它是用户提交的包含了特定的文献信息需求的检索式与文献信息检索系统中的信息集合进行相符性比较的过程,这里的相符性比较是建立在检索语言的基础之上的。

文献信息检索有广义和狭义两层含义。广义的文献信息检索包括文献信息的存储和检索两个方面。存储是指对一定数量的揭示文献特征的信息或从文献中摘出的知识、信息进行组织、加工、整序并将之存储在某种载体上,编制成检索工具或组织成检索系统。检索

就是根据需要,利用一定的检索工具和检索手段,把所需的文献线索或知识、信息从检索系统中查找出来的过程。

狭义的文献信息检索通常是指后一半,即文献信息查询或检索,是从任何信息集合、情报集合中找出所需的信息的过程。存储和检索从意义上讲是两个完全不同的概念,但存储是为了检索,而检索以存储为前提,它们是相互依存的关系。

信息检索就是指依据一定的方法,从已经组织好的大量信息集合中查出特定的相关信息的过程。文献检索是信息检索的一种类型,是指依据一定的方法,按照一定方式将文献组织存储在某种载体上,并利用相应的方法或手段从中查出符合用户特定需要的文献的相关过程。查找出来的只是关于文献的信息或文献的线索,如果要真正获取文献中所记录的信息,那么还要依据检索所取得的文献线索或关于特定文献的信息,去索取和查阅文献的原文。文献检索是文献信息工作的重要组成部分,是科学研究的前期工作,进行文献检索时,必须按照与存储一致的思路和方法,才能获得良好的检索结果。^①

二、文献信息检索的原理

文献信息检索的基本原理,就是通过对大量的、分散无序的文献信息进行收集、加工、组织、存储,建立各种各样的检索系统,并通过一定的手段和方法,使信息存储和检索这两个过程所采用的特征和表示达到一致,以便有效地获得和利用信息。文献信息检索是文献检索和信息检索两个概念的统一。文献信息检索的全过程包括存储和检索两个过程。其中,存储是检索的基础,检索是存储的目的。文献信息存储和检索的全过程如图 1-1 所示。

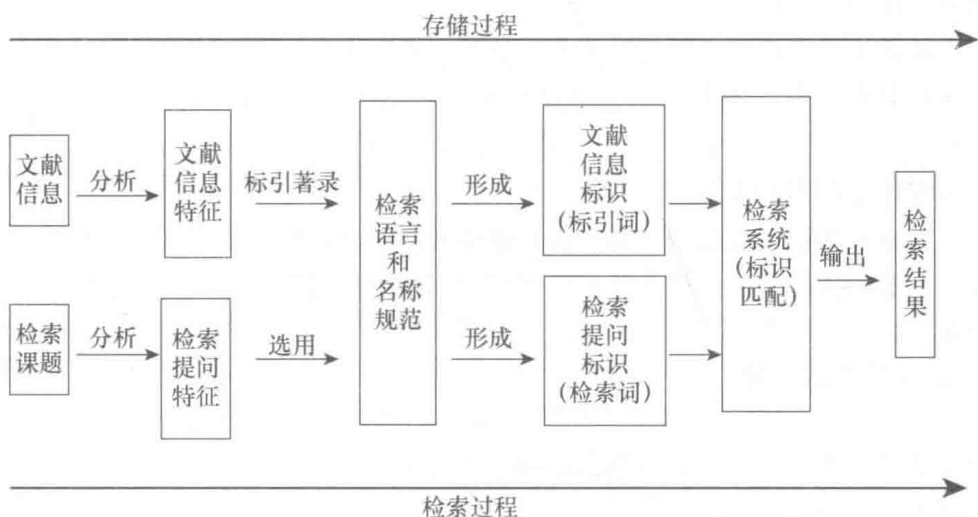


图 1-1 文献信息存储和检索过程

1. 存储过程

存储过程即选择特定内容或特定范围的文献信息,抽出文献信息的特征进行著录、标引和整序,形成一条条文献线索并使之成为有序化的文献集合体,即编制检索工具或建立检索系统的过程。该过程主要包括以下三个方面。

^① 高新陵,吴东敏.科技文献信息与科技创新.南京:河海大学出版社,2013.

(1) 著录。从文献信息中选择特定文献,并描述其外表特征(题名、著者、号码等)和内容特征(主题词、分类号等),形成一条条款目。

(2) 标引。对文献信息的主题内容进行概念分析,将每个概念转换成系统语言(检索语言),用表达文献内容特征和外表特征的最简练的词语或某种符号作标识,如主题词、分类号、著者姓名等。

(3) 排序。将各种标识按照一定的顺序系统排列(如按学科分类、字顺、数序、音序等),使所选文献成为有序化的、有共同存储和查找方式的、为检索者提供有规律的检索途径的文献集合体(检索工具或检索系统)。

2. 检索过程

检索过程指从文献集合体(检索工具或检索系统)中,查出(或输出)所需文献线索或浓缩物的过程。这主要是按照需要形成提问标识,按照一定的检索途径,查找与检索提问相符合的文献特征标识的过程。或者说文献信息检索就是通过文献特征标识与检索提问标识的匹配来实现的。为了实现这一匹配,标引人员的标引用语和检索者的检索用语必须采用共同的语言,这就是检索语言。检索语言沟通了文献的存储和检索两个过程,是检索工具或检索系统的重要组成部分。其检索过程包括以下三个方面。

(1) 分析检索课题,确定检索入口,找出能表达课题特征或主题要领的检索提问标识,如主题词、分类号、著者姓名等。

(2) 将检索提问标识与文献集中的文献标识进行相符性比较。标识一致时,标识所指的文献就是用户所需的文献。

(3) 在检索工具(或检索系统)中查找所需文献的线索,进而获取原文。

所以说,存储是检索的基础,检索是存储的目的,检索过程实际上是存储过程的逆过程。

三、文献信息检索的类型

检索是对未知知识最有效的获取方法,也是人们根据需要,利用人类有史以来积累的知识的唯一有效的方法。它的最大优势在于可以用最少的时间、最快的速度获得尽可能多的信息与知识。

1. 依据检索方式划分

(1) 手工检索。手工检索简称“手检”,指人们通过手工的方式检索信息。其使用的检索工具主要是书本型、卡片式的信息系统,即目录、索引、文摘和各类工具书。检索过程是由人以手工的方式完成的。

(2) 计算机检索。计算机检索简称“机检”,指人们利用数据库、计算机软件技术、计算机网络及通信系统进行的信息检索,其检索过程是在人机的协同作用下完成的。

(3) 综合检索。在文献信息检索的过程中,既使用手工检索方式,又使用计算机检索方式,也就是同时使用两种检索方式。

2. 依据检索内容划分

(1) 文献型信息检索。文献型信息检索指利用检索工具或检索系统查找文献的过程,包括文献线索检索和文献全文检索。文献线索检索指利用检索工具或检索系统查找文献的出处,检索结果是文献线索。它包括书名或论文题目、著者、出版者、出版地、出版时间

等文献外部特征。用于检索文献线索的检索工具有书目、索引、文摘及书目型数据库和索引、题录型数据库。文献全文检索是以文献所含的全部信息作为检索内容,即检索系统存储的是整篇文章或整部图书的全部内容。检索时可以查到原文以及有关的句、段、节、章等文字,并可进行各种频率统计和内容分析。文献全文检索主要是用自然语言表达检索课题,较适用于某些参考价值大的经典性文章,如各种典籍、名著等。文献全文检索是当前计算机信息检索的发展方向之一。

(2) 事实型信息检索。事实型信息检索以特定的事实为检索对象,其检索结果主要是客观事实或为说明客观事实而提出的数据。如某个字、词的查找,某一诗句的查找,某一年、月、日的查找,某一地名的查找,某一人物的查找,某一机构的查找,某一事件的查找,某一法规制度的查找,某一图像的查找,某一数据、参数、公式或化学分子式的查找等。完成事实型信息检索主要借助于各种参考工具书及事实型数据库。事实检索多利用词语性和资料性工具书,包括字典、词典、百科全书、类书、政书、年鉴、手册、名录、表谱、图录等。也利用某些线索性工具书,如索引、文摘、书目,以及利用学科史著作、科普读物等。

(3) 数据型信息检索。数据型信息检索是一种确定性检索,是以数值或图表形式表示的数据为检索对象的信息检索,又称“数值检索”。检索系统中存储的是大量的数据,这些数据既包括物质的各种参数、电话号码、银行账号、观测数据、统计数据等数字数据,又包括图表、图谱、市场行情、化学分子式、物质的各种特性等非数字数据。

3. 依据信息组织方式划分

(1) 全文检索。全文检索指检索系统中存储的是整篇文章乃至整本书。用户根据个人的需求从中获取有关的章、节、段、句等信息,并且还可以做各种统计和分析。

(2) 超文本检索。超文本结构类似人类的联想记忆结构,它采用了一种非线性的网状结构组织块状信息,没有固定的顺序,也不要求读者必须按照某个顺序来阅读。采用这种网状结构,各信息块很容易按照信息的原始结构或人们的“联想”关系进行组织。

(3) 超媒体检索。由于把多媒体信息引入超文本里,产生了多媒体超文本,也即超媒体。它是对超文本检索的补充,其存储对象超出了文本范畴,融入了静态、动态图像及声音等多媒体信息。信息存储结构从单维发展到多维,存储空间范围不断扩大。

第三节 文献检索语言

一、文献检索语言概述

文献检索语言就是文献信息检索系统中的标识系统,能提供多种多样的检索点,如著者名、分类号、主题词、关键词等。文献检索语言是人工创制的或者由人工设计并由计算机自动生成的,它可以是从自然语言或专业文献中抽取出来并予以规范化的一套词汇(如主题词表),也可以是遵循某种分类体系的一套分类代码,或者是代表某一类事物的某一方面特征的一套代码(如代表化合物的多种代码)等。它们可用于对文献和网络信息的内容进行逻辑分类、主题标引或特定信息的描述和提示,所以又称为文献存储与检索语言、标引语言、索引语言等。

文献检索语言在各种文献检索系统中无处不在,它种类繁多,各具特点,各有优势又或多或少存在缺陷。在实际应用中常有两种或多种检索语言用于同一检索系统以供选择使用或者相互取长补短。近年来在强大的计算机信息技术支持下研制开发的新型检索语言集成系统,已使网络文献信息智能化检索初现端倪。用户的检索提问可以用短语甚至句子等自然语言形式输入,系统能够进行自动分析形成检索策略进行检索。检索技术的长足进步,很大程度上得益于检索语言研究成果的应用。

检索语言就是为沟通文献标引与文献检索而编制的人工语言,也是连接信息存储和检索两个过程中标引人员与检索人员双方思路的渠道,是用于文献标引和检索提问的约定语言。如果没有检索语言作为标引人员和检索人员的共同语言,就很难使标引人员对文献信息内容的表达和检索人员对相同文献信息内容需求的表达取得一致,信息检索也就不可能顺利实现。因此,编制检索语言的目的是不但能够保证不同的标引人员描述文献特征的一致性,而且能够保证检索提问词与文献标引词的一致性。要把存储和检索联系起来,检索语言所表达的概念应该是唯一的。这就是说,表达的概念同所要表达的事物一一对应,尽量减少一词多义或多词一义的现象,要使其在该检索系统中具有单义性。文献检索语言是规范化的自然语言,归纳起来有以下几方面的特点。

(1) 检索语言具有严密性(单一性)。一词一义,每一个词只表达一个概念,为文献的标引和检索提供了共同语言的依据。

(2) 检索语言具有系统性。在标引文献时可以将同一主题概念的文献集中存储在一起,便于用户检索。

(3) 检索语言具有可控性。在标引文献的主题概念时,可以将文献包含的情报内容及其外表特征简明地表示出来。

二、文献检索语言的基本要求

文献检索语言主要有以下几方面的要求。

(1) 具有必要的语义和语法规则。能准确地表达某一学科技术领域的任何文献和任何提问的中心内容、主题、特征。

(2) 具有表达概念的单一性。用这种语言表达的每一文献或提问,只能有一种解释。

(3) 具有文献检索标识和提问特征进行比较和识别的方便性。既可用于手工检索,又可用于计算机检索。

(4) 检索语言体系应科学合理。

三、文献检索语言的功能

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,使概念的表述模糊不清。检索语言的这种单一性保证了表述概念的唯一性,保证了标引与检索的一致性,使信息检索人员能够又全、又准、又快地检索到含有所需信息的文献。

检索语言的功能是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所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,以便进行有规律的检索。在信息存储过程中,用它来描述信息的内容和外部特征,从而形成检索标识;在检索过程中,用它来描述检索提问,从而形成提