

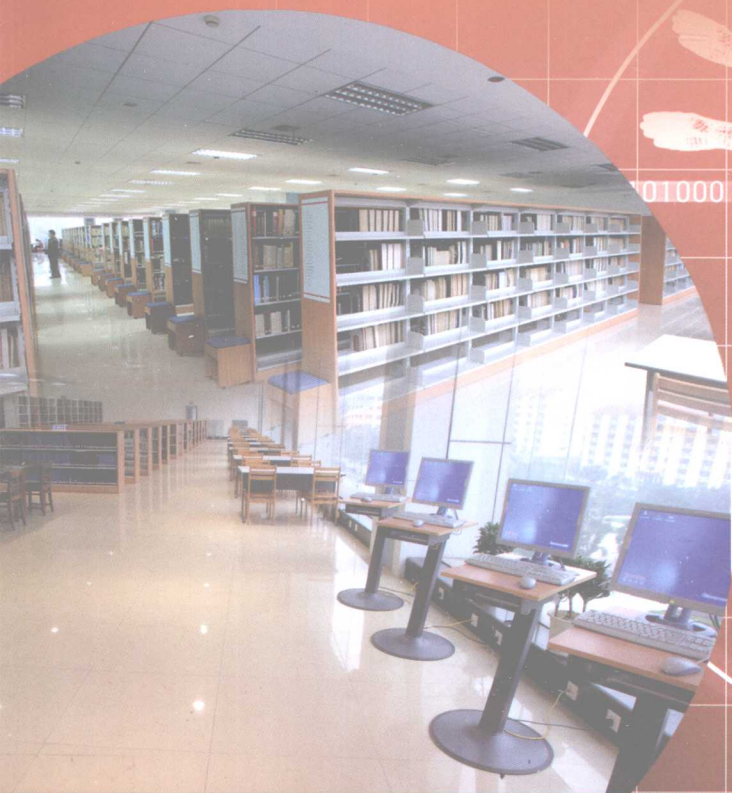


中国科学院教材建设专家委员会规划教材

全国高等医药院校规划教材

医学文献检索

李彭元 何晓阳 主编



科学出版社

www.sciencep.com

中国科学院教材建设专家委员会规划教材

全国高等医药院校规划教材

医学文献检索

主 编 李彭元 何晓阳

副主编 徐成兵 耿 鹏 武海东

编 者 (以姓氏拼音排序)

丁云轩 耿 鹏 何晓阳 李彭元

连丽红 刘小谦 龙卫东 唐德羽

文思敏 吴治蓉 武海东 徐成兵

叶春芳 张精理 赵海荣

科 学 出 版 社

北 京

· 版权所有 侵权必究 ·

举报电话:010-64030229;010-64034315;13501151303(打假办)

内 容 简 介

本书是在吸收和消化大量医学文献检索理论和实践成果的基础上编写的,内容主要有医学文献检索的基本理论、基本知识和基本技能,图书馆利用的基本知识和技能以及生物医学领域具有代表性的重要数据库、检索系统和网络信息的检索利用,还为适应军医大学医学文献检索课程教学实践的需要,特别增加了有关军事医学信息检索的内容。同时考虑到参考文献管理软件对科学论文写作和投稿的重要意义,选取了国内外和医学领域具有代表性的参考文献管理工具予以介绍。另外,为了读者学习和利用教材的方便,还在附录部分增加了内容索引。

本书内容丰富、语言精练,是进行医学文献信息检索的实用工具书,适合高等医药院校各专业的学生和从事医学等领域工作的人员使用。

图书在版编目(CIP)数据

医学文献检索 / 李彭元,何晓阳主编. —北京:科学出版社,2010.8
(中国科学院教材建设专家委员会规划教材·全国高等医药院校规划教材)
ISBN 978-7-03-028450-1

I. 医… II. ①李… ②何… III. 医学-情报检索-医学院校-教材
IV. G252.7

中国版本图书馆 CIP 数据核字(2010)第 148491 号

策划编辑:李国红 许贵强 / 责任编辑:许贵强 李国红 / 责任校对:陈玉凤
责任印制:刘士平 / 封面设计:黄 超

版权所有,违者必究。未经本社许可,数字图书馆不得使用

科学出版社出版

北京东黄城根北街 16 号

邮政编码:100717

<http://www.sciencep.com>

新蕾印刷厂印刷

科学出版社发行 各地新华书店经销

*

2010 年 8 月第 一 版 开本:787×1092 1/16

2010 年 8 月第一次印刷 印张:19 1/2

印数:1—5 000 字数:462 000

定价:36.00 元

(如有印装质量问题,我社负责调换)

前 言

语云“知识的一半,就是知道要到哪里去寻找它。”文献检索是讲授如何在浩如烟海的信息和知识的集合中,准确而无重大遗漏地寻找特定知识的一门技能,这在知识和信息成几何级数迅猛增长的今天,其重要性不言而喻。可以说,文献检索的技能也就是在知识的海洋中检索和获取特定知识的能力,在很大程度上决定了一个人开始起步的高度。文献检索对任何一个有志于学习的人,特别是正在学校学习的莘莘学子,不论他将来是从事教学、科研、临床,还是管理,如果希望自己的工作能够事半功倍,能够站在巨人的肩膀上看得更远,医学文献检索都是一门需要认真学习和掌握的基本技能。

本书在吸收和消化大量医学文献检索理论和实践成果的基础上,除了编写有关医学文献检索的基本理论、基本知识和基本技能,有关图书馆利用的基本知识和技能,以及生物医学领域具有代表性的重要数据库、检索系统和网络信息的检索利用外,还为适应军医大学医学文献检索课程教学实践的需要,特别增加了有关军事医学信息检索的内容。同时考虑到参考文献管理软件对科学论文写作和投稿的重要意义,本书选取了国内外和医学领域具有代表性的参考文献管理工具予以介绍。另外,为了方便读者学习和利用教材,还在附录部分增加了内容索引。这些新增加的内容具有一定的探索性,希望这些内容的增加对热爱学习的医学生和已经参加工作的医学工作者,在医学文献信息的检索、管理和利用方面能有所裨益。

由于参加本书编写人员的学养和水平有限,教材中不足和错漏之处在所难免,恳请学界同仁不吝赐教、惠予指正,以便将来再版时修订。

编者

2010年4月

目 录

前言

第一章 绪论	(1)
第一节 文献的基本知识	(1)
第二节 信息检索语言	(7)
第三节 文献检索基础	(15)
第四节 计算机信息检索系统	(21)
第二章 图书馆资源利用	(25)
第一节 图书馆服务简介	(25)
第二节 馆藏图书的分类与排架	(28)
第三节 馆藏目录和联合目录	(30)
第四节 学术期刊及其利用	(34)
第五节 图书馆数字资源利用	(37)
第三章 常用医学文摘数据库检索	(40)
第一节 中国生物医学文献数据库	(40)
第二节 MEDLINE	(49)
第三节 PubMed	(57)
第四节 Embase	(67)
第五节 BIOSIS Previews	(76)
第四章 全文数据库检索	(85)
第一节 电子图书全文数据库	(85)
第二节 中文期刊全文数据库	(99)
第三节 外文期刊全文数据库	(114)
第四节 网络免费期刊全文资源	(134)
第五章 引文数据库检索	(139)
第一节 引文及引文数据库	(139)
第二节 中国生物医学期刊引文数据库	(145)
第三节 中国科学引文数据库	(149)
第四节 美国科学引文索引	(153)
第五节 期刊引证报告	(162)
第六章 特种文献信息检索	(165)
第一节 学位论文检索	(165)
第二节 会议文献检索	(170)
第三节 专利信息检索	(180)

第七章 网络医学信息检索	(190)
第一节 搜索引擎概述	(190)
第二节 通用搜索引擎	(192)
第三节 医学搜索引擎	(200)
第四节 常用中文生物医学网站	(206)
第五节 常用外文生物医学网站	(209)
第八章 药学信息检索	(217)
第一节 药学信息资源概述	(217)
第二节 国际药学文摘数据库	(222)
第三节 Drug Information Fulltext	(225)
第四节 网络药学信息资源	(228)
第九章 军事医学信息检索	(235)
第一节 军事医学信息资源的特点和分布	(235)
第二节 军事医学信息检索	(238)
第十章 文献管理工具	(260)
第一节 EndNote	(260)
第二节 NoteExpress	(276)
第三节 医学文献王	(283)
参考文献	(293)
附录	(295)
附录一 国内外主要图书分类法简介	(295)
附录二 MeSH 范畴表主要类目	(297)
附录三 MeSH 副主题词等级表	(301)
附录四 AHFS 药理学/治疗学分类	(302)
索引	(305)

第一章 绪 论

第一节 文献的基本知识

一、文献的概念

文献(documents, literature)是“记录有知识的一切载体”(《中华人民共和国国家标准:文献著录总则》, GB 3792. 1-83), 是指以文字、图像、公式、声频、视频、代码等手段, 记录或描述有信息、知识的一切物质载体。人类在漫长的生产实践、科学实践和社会实践中逐步地认识客观世界, 并产生和积累了大量的知识和信息。为了更好地把这些知识和信息存储起来, 便于传播、交流和利用, 需要用一定的方法和手段, 将这些知识和信息记录在一定的物质载体上, 这就形成了文献。文献由三个基本要素构成, 一是内容上的知识或信息, 这是文献的实质内容; 二是记录知识和信息的手段, 即揭示和表达知识或信息的标识符号, 如文字、图像、公式、声频、视频、代码等; 三是载体, 即供记录知识或信息符号的物质材料, 举凡泥版羊皮、龟甲兽骨、金石泥陶、竹木缣帛、纸张胶片、磁性材料等。

知识和信息是文献的实质内容。知识是经过人的思维整理过的信息、数据、形象、意象、价值标准以及社会的其他符号化产物, 不仅包括科学技术知识——知识中最重要的部分, 也包括人文社会科学知识, 还包括商业活动、日常生活和工作中的经验和知识, 人们获取、运用和创造知识, 以及面临问题做出的判断和提出解决方法的知识。显然, 了解信息里面包含的知识是人们利用信息的主要目的。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件的不确定性减少。

与知识和信息密切相关的概念还有情报, 情报是指人们为一定的目的而收集的有使用价值的知识或信息。文献与信息、知识和情报之间有密切的联系。信息、知识、情报必须固定在一定的物质载体上, 形成文献后才能长期进行传递, 文献是信息、知识和情报存储、传递和利用的重要方式。因此, 有的文献称文献检索为情报检索、知识检索或信息检索。

二、文献的特征

文献的内部特征和外部特征是文献检索的入口, 是设计情报检索语言的基础, 充分认识文献的特征对文献检索具有重要意义。从文献检索的角度来考虑, 科技文献的特征分为内部特征和外部特征。文献的内部特征主要是指作为文献所记录的知识和信息的科学分类和研究主题。以文献的内部特征为依据形成的检索语言是文献检索的重点和难点。文献的外部特征主要包括文献的题名、责任者、摘要、关键词、特有的序号(如专利号)、参考文献等。

科学分类, 是根据一定的原则, 全面研究某个时代各门学科之间的区别与联系, 用一定的结构形式, 把它们连成一个整体, 从而确定每门学科在整个科学体系中的地位, 揭示各学

科领域之间的联系,其结果就是科学分类体系。文献是记录科学知识和信息的,因此文献分类以科学分类为基础,结合文献特有的属性,把相关的文献纳入既定的分类体系,完成对文献的分类整理,建立起文献分类体系。以科学分类为基础的文献分类,为分类检索文献提供了可能。科学分类法虽然能很好地体现现有学科的知识体系,但不能处理大量的非学科文献。

“主题”一词源于德国,最初是一个音乐术语,指乐曲中最具特征并处于优越地位的旋律,即主旋律,它能表现一个完整的音乐思想,是乐曲的核心。后来这个术语被广泛用于文艺创作和科学研究中。主题词是指能代表文献内容实质的、经过严格规范的专业名词术语或词组。文献检索是将表达文献主题内容主题词按字顺排列,并利用参照系统和其他方法间接显示概念之间的关系,提供按照反映文献主题的主题词字顺检索途径。文献的主题有主要主题和次要主题之分,因此,描述文献主题内容主题词也有主要主题词和次要主题词之分。

利用文献的外部特征作为标识和检索点可以设计检索语言。如把文献的著者收集在一起,按字顺排列,形成倒排文档,可以提供从著者途径检索文献的方法。再如把化学物质的分子式按元素符号字顺排列,元素符号相同再比较原子数目,分子式相同的物质,按母体名称排列,若还相同,再比较取代基前缀,形成倒排文档,可以提供从化学物质的分子式出发检索文献的方法。有关文献的特征与文献检索之间的关系和对文献检索所起的作用,请参见本章第二节。

三、文献的社会职能

(一) 文献是人类的精神财富和人类文明的重要结晶

自从有了文字,人类在长期的改造自然和改造社会的实践过程中,一代接一代地把自己获得的实践结果,包括成功的经验和失败的教训,记载在一定的物质载体上形成文献,并传之久远。文献是人类所从事的社会活动、科学活动和生产活动的真实反映,是人类的精神财富和人类文明的重要结晶。文献在保存、积累和传承文化知识及科学技术方面的作用是不可低估的。

(二) 科技文献是衡量科学技术发展水平的重要标志

从现代社会的观点来考察,人们所从事的每一项科学研究和技术发明,都必须撰写和发表反映科研成果、发明创造的文献。科技文献是记录、揭示和传播最新科学成果和科学知识的重要手段和媒介。因此,科技文献的产出和质量成为衡量一个国家、地区的学科或机构科技发展状态及学者学术水平的重要标志,是体现个人或团体科学创造力、确认其科学地位的一个公认指标。

(三) 科技文献是传播科学技术知识的基本手段

人类已有的知识、信息,都存储在文献中,文献成了人类社会最重要的情报源。文献从纵向保存和继承了人类社会从古至今的文化科学知识,横向形成了连接知识创造和利用的纽带,这为继承和传播前人和同代人的科学研究成果提供了物质基础。自19世纪以来,先后出现了电话、广播、电视、互联网等各种先进的现代化通信技术,知识信息的传播手段有

了巨大进步。人们可以通过书信、各种学术会议、网络等多种形式来进行学术交流,但最重要、最基本,也是最规范的手段和途径仍然是科技文献。文献不仅是展示科研成果的园地,同时也是获取最新信息,实现知识更新的最主要信息来源。

(四) 科技文献是确认科学发现和技术发明优先权的基本依据

科学荣誉与优先发表相关联是学术界的惯例。科学发展史上的任何一项发现或发明总是和第一个发表相应文献的作者姓名联系在一起的,即科技文献成为认定科学发现和技术发明优先权的重要依据。因此,在获得新知识、新发现等研究成果时,都要尽快、尽早地以科学文献的形式发表或申请专利,以获得社会的承认。日本京都岛津制作所的工程师田中耕一因在质谱学、分裂蛋白质分子领域研究出新的方法而获 2002 年诺贝尔化学奖,当时在学术界引起激烈争论。部分科学家认为,另外两位德国科学家虽然比田中耕一发表类似研究成果晚两个月,但是方法比田中耕一更见成效,因此更应该获得此项大奖。对此,诺贝尔奖化学委员会主席本特·努登予以严厉驳斥,他认为该奖项本来就是要授予第一个改变人们思维方式的发明者,田中耕一获奖理所当然。

(五) 科技文献是进行科学评价研究的基本素材

科技文献与科学技术发展水平密切相关,是科技成果的真实描述和客观反映。科学技术能够通过科技文献反映出来。对于科学技术来说,科技文献具有多方面的评价功能,这种评价具有很高的准确性和可靠性。采用数学与统计学方法来描述、评价与预测科学技术的现状与发展趋势的文献计量学,对科学活动过程与管理实施量化评估、刻画与预测的科学计量学,采用定量方法来描述和研究信息现象、过程和规律的信息计量学,以及以科学为研究对象的科学学,无一不以科技文献为研究的直接取材,离开了科技文献,这些关于科学评价的研究就无法进行。

四、文献的分类

根据不同的标准,可以将文献划分为不同的类型。

(一) 按文献的出版形式划分

1. 图书 图书是各种文献中最为古老、利用率最高,至今仍在频繁使用的文献之一。图书的主要特征是:主题突出,内容系统,论述全面,观点成熟,可读性、知识性、学术性较强。它是人们系统了解和掌握某一学科知识或某一创作内容的基本文献。联合国教科文组织将篇幅(封面除外)不少于 49 页的非定期出版物称为图书;5~48 页的则称为小册子;4 页以下的称为零散资料,用来与期刊等连续出版物区别。常用的图书有教科书、专著、参考工具书等。

每一种公开出版的图书在其版权页都标有一个由 12 位数字组成的国际标准书号 (ISBN),这是一种国际通用的出版物代码,代表某种特定图书的某一版本,具有唯一性和专指性。读者可以通过 ISBN 检索或查询特定的图书。

2. 期刊 期刊是连续出版物的一种,具有比较固定的名称和出版周期,比较一致的开本、稳定的栏目,并以分期形式逐次刊行。其中以报道最新科技知识和科研成果为主的期

刊为科技期刊。学术性的期刊刊名中常有杂志(journal)、学报(acta、bulletin)、纪事(annals)、文献(archives)、会刊(proceedings)、评论(reviews)、进展(progress、advances in)等字样。

与图书类似,每种期刊均有国际标准连续出版物号(ISSN)。ISSN号同样具有唯一性和专指性,因而成为文献检索中查询某一种期刊的检索途径之一。除期刊外,连续出版物还包括以刊载新闻、时事评论为主的报纸,汇编一年内各方面或某方面的情况、资料、数据的年鉴。

3. 学位论文 根据国家标准(GB 7713-87),学位论文是表明作者从事科学研究取得创造性的结果或有了新的见解,并以此为内容撰写而成的,作为提出申请授予相应的学位时评审用的学术论文。简而言之,学位论文就是学生为了获取学位向高等学校或其他学术研究机构提交的学术研究论文。学位论文有学士论文、硕士论文和博士论文之分,水平参差不齐,其中以博士论文的价值最高。

4. 会议文献 一般指在各种学术会议上发表和交流的学术报告、会议录和会议论文集。通常学术会议只涉及某个学科领域的一个或几个相关主题,与会者大多为该领域的研究人员,他们对会议主题的历史、现状及发展趋势有着不同程度的研究和了解。会议论文代表了各自的最新研究成果,因此与学术会议密切相关的会议文献具有专业性、新颖性、前瞻性和导向性的特征。按发表时间的不同,会议文献可划分为会前文献和会后文献。

5. 专利文献 是指实行专利制度的国家及国际性专利组织在审批、公布专利过程中所产生的官方文件及有关出版物的总称。此文献一般按功能分为三类:①详细描述发明创造具体内容及其专利保护范围的各种专利说明书;②刊载专利题录、专利文摘和专利索引的专利公报、专利年度索引;③专利分类表等,此类文献是集技术、法律、经济信息于一体的科技文献,专利说明书是其核心部分。绝大多数医学研究成果属于科学发现,而不是技术发明的范围,因而不能享受专利保护,但医学科研中的医疗设备、医疗仪器、化学物质、微生物菌种、药品开发等则属于专利保护的范畴。

6. 其他文献 除了上面提到的图书、期刊等主要的文献类型外,还有科技报告、政府出版物、技术标准、产品资料、技术档案、产品说明书等文献类型。这些文献都是我们获取有关信息的重要途径。

(二) 按文献的载体形式划分

1. 纸质型文献 又称印刷型文献,是以纸张为记录载体,以书写或印刷技术为记录手段而产生的文献。此类文献的优点是携带方便、不需要借助其他特殊的阅读设备,符合传统的阅读习惯,是目前人类交流信息和传递知识的最重要、最常用的手段;缺点是存储密度小、占用空间大,现代的纸质文献因纸张的理化特性,不易长期保存。

2. 缩微型文献 缩微型文献是以感光材料为载体,以照相技术为记录手段而形成的一种文献,包括缩微胶卷、缩微平片、缩微卡片等。此类文献的优点是占用空间小、便于收藏和长期保存、价格便宜、复制性能好,缺点是需要借助复杂的阅读设备才能进行阅读,保存和使用需要一定的条件。目前,在整个文献体系中,缩微型文献只占较小的比例,一般图书馆较少入藏。

3. 声像型文献 又称视听型文献,是以感光材料、电磁材料为存储介质,以电磁或光学手段将文字、图像和声音记录下来而形成的一种动态型文献,必须通过相应的视听设备才能“阅读”。此类文献主要包括录音带、录像带、唱片、光盘等。其优点是有良好的直观性能、

存储密度高、便于远距离传输。

4. 电子型文献 又称数字型文献,是将文字、声音、图形、图像等信息,以数字编码的方式存储在磁盘、光盘等载体上,通过电子计算机等设备进行阅读利用的文献类型,主要包括电子图书、电子期刊、各种数据库等。与传统的文献类型相比,电子型文献具有占用空间小、传递速度快、便于检索、易于复制、利于共享等特点。此类文献的大量问世是信息时代来临的重要标志之一。随着计算机技术和网络技术的迅速发展和日益普及,电子型文献越来越受到人们的重视。

上面的四种文献形式,印刷型文献和电子型文献的比例占了整个文献体系的绝大多数,而缩微型文献和声像型文献所占比例则相对逐渐减小。

(三) 按文献的加工层次划分

1. 一次文献 又称原始文献,是文献的基本类型,是著者根据自己的研究成果(如科学实验、调查研究、观察分析的结果等)为基本素材撰写而成的文献,包括期刊论文、学术专著、科技报告、专利说明、会议论文、学位论文、技术标准等。译文属于一次文献。一次文献是对知识的第一次加工,其特点是新颖性、原创性和系统性。文献检索就是要通过二次、三次文献等检索工具进行查询,最终获取一次文献,即原始文献。

2. 二次文献 是为了更好地报道和便捷地检索一次文献,由相关机构对一次文献进行收集、分析、加工、整理后,并抽取一次文献的内部特征和外部特征,按照一定的规则加以编排而形成的文献,具体包括各种书目、索引、文摘等出版物。它将大量分散、无序、庞杂的原始文献转变为有序的、便于管理和检索的系统。因此,二次文献浓缩了大量同一内容的一次文献,是文献检索工具的主体。

3. 三次文献 是在利用二次文献的基础上,选择一次文献加以分析、综合而编写出来的专题报告,如文献综述、述评报告、技术预测、数据手册、百科全书、年鉴、进展、指南等。此类文献具有综合性特点,一般说来综述蕴涵的信息量大,可使读者不必阅读大量一次文献就能了解当前某一课题领域研究水平与动态,从而节约时间。文献后面往往附有大量的重要参考文献,为科研人员开展研究提供非常重要的线索,因此备受科研读者的欢迎。

4. 零次文献 零次文献即形成一次文献之前非正式出版的文献,如手稿、笔记、信函、发言稿、实验数据、观察记录、调查材料、统计数字、各种口头交流的信息、经验等。这些未正式出版的文献,往往反映的是科研工作中的新进展、新设想或是遇到的新问题,具有新颖性、及时性、启发性等特点。由于零次文献不公开发表,查找和获取比较困难。

此外,学术界还有关于灰色文献和黑色文献的概念。灰色文献指非公开发行的内部文献或限制流通范围的文献,主要包括一些控制在内部范围发行的刊物、教材、技术报告、会议资料等。黑色文献指尚未破译或辨识其中信息的文献,属于保密状态或不能公布其内容的文献,如考古发现的考古文献、内部档案、私人日记等。零次文献、灰色文献和黑色文献的内涵和外延具有一定的重叠。

五、文献的发展趋势

(一) 数量增长迅速,内容交叉重复

据联合国教科文组织统计,20世纪50年代,全世界每年出版图书约20万种、期刊约2

万种,现在每年出版图书约 100 万种、期刊约 20 万种,图书平均 20 年增长一倍,期刊 10 年就翻一番。除正常的科学技术进步和新兴学科、边缘学科的诞生带来的文献增长外,同一篇文献用不同的形式、不同的语种、在不同的地区发表,也使得文献总量大大增加。比如,一篇会议论文可以在科技期刊上发表,也可以写进学位论文,或将已经发表的期刊论文集结出版。再版、重版的文献大量增加,这些现象导致文献数量增长迅速,载体和出版形式交差重复,大大增加了文献的冗余。

(二) 老化速度加快,分布聚散有序

现代科学技术的迅速发展使得新理论、新技术、新方法、新成果不断地取代原有的理论、技术、方法和成果,知识更新和老化的速度相应地加快。作为科学知识载体的文献的更新速度也越来越快,更新周期越来越短。1958 年美国科学家贝尔纳在其发表的《科技情报的传递:用户分析》一书中,借用放射线元素衰变过程中的术语“半衰期”来描述文献的老化速度。文献的半衰期是指各学科被利用的文献总量中,一半文献失去利用效率所经历的时间。各种文献形式和各个学科文献的半衰期是不同的,而且呈现出越来越短的趋势。在各个学科中,医学文献的半衰期是比较短的。

科技文献分布聚散有序是说现代科技文献呈现出既集中又分散的不均匀现象。1930 年英国文献学家布拉德福率先提出描述文献分布聚散规律的经验定律——布拉德福定律。其文字表述为:如果将科技期刊按其刊载某学科专业论文的数量多少,以递减顺序排列,那么可以把期刊分为专门面对这个学科的核心区、相关区和非相关区。各个区的文章数量相等,此时核心区、相关区、非相关区期刊数量成 $1:n:n^2$ 的关系。

(三) 质量参差不齐,无用文献增加

文献内容交叉重复和文献老化速度加快,大大增加了文献的冗余度,导致文献质量参差不齐和无用文献增加。据统计,约有 35% 的期刊论文发表后从未被人引用过,有 49% 的文献仅被引用过一次。学术腐败、科研工作低水平重复对无用文献的增加推波助澜。此外,要求发表的文献太多,技术发明为了抢占首创权等原因,许多文献未经严格审查就发表出来也导致无用文献的增加。计算机和互联网技术的普及,在为学术交流和文献获取带来极大方便的同时,也使得互联网上的电子文献薰莠同器、良莠不齐。

(四) 发表时滞严重,传播速度加快

科学技术飞速发展,科技论文大量增加,导致科技论文从投稿到发表的时间越来越长,有的达到一年以上,而且时滞仍在不断延长中。特别是核心期刊的稿源丰富,编审校人员工作认真,编辑质量较高,从收稿到论文发表时间更长,有的论文从投稿到经编辑加工、作者修改到发表有的达到两年之久。

但是,随着现代交通工具和通信手段的进步,文献的传播速度越来越快,甚至瞬间即至。科学文献一旦公开发表,进入学术交流系统,读者通过网络获取文献,几乎没有时间和地域的差别。而在 15 世纪末,哥伦布发现新大陆的消息过了半年才传到西班牙女皇那里;100 多年前美国总统林肯遇刺的消息在 12 周以后才传到英国王宫;互联网普及以前,一篇科技论文在美国出版的科技期刊上发表,到中国读者阅读到这篇论文,至少需要两个月的时间。真可谓是“事易时移,不可同日而语”。

(五) 文献载体多元,语言障碍增加

文献的载体种类繁多,有传统的纸质文献,也有缩微胶片、缩微胶卷、缩微卡片等缩微型文献,录像带、录音带、电影、幻灯片等各种视听型文献,大量的电子文献,更是以惊人的速度增加。一篇科技论文在科技期刊上发表后,又被收录进数据库,成为电子文献,供用户使用。

发表文献的语种也越来越多。20 世纪初,科研人员只要掌握英语、德语和法语,就能够阅读全世界 92% 的化学化工文献;20 世纪 50 年代以后,这三种语言的化学化工文献只占到 66.6%;目前,科技文献常用的语言有六七十种之多。美国的《化学文摘》收录的语种达 70 多种,《医学索引》收录的语种也有 45 种。在医学文献中,英语发表的文献占了 70%,其余的医学文献都是用英语以外的语种发表的。发表文献的语种增加,阻碍了科技信息的交流。

正因为现代文献具有以上的发展趋势,要在浩如烟海杂乱无序的文献集合中检索到自己需要的特定的文献,必须学习和掌握文献检索的基本知识和基本技能。

第二节 信息检索语言

科技文献浩如烟海,每个用户的信息需求又各不相同,我们应当如何从浩如烟海的文献中找到所需信息呢? 我们需要一个能将文献与需求联系起来的“桥梁”——用户和系统的共同约定或规范,以此描述文献信息及用户需求的特征。这种把信息存储与检索联系起来、把信息组织人员与信息用户联系起来以便取得共同理解、实现交流的语言,叫做信息检索语言。

信息检索语言的作用表现在两个方面:首先,检索语言用来描述文献及文献中信息内容的特征,把文献转换为一定的文献标识,构成信息检索系统的各种检索途径,这个过程称为信息组织;同时,它也用来描述检索提问及需求内容的特征,把提问转换为一定的提问标识或检索标识,以便在检索系统中查找特定的文献信息,这个过程则称为信息检索。信息组织是信息检索的前提和基础,信息检索是信息组织的出发点和归宿,二者相辅相成。信息检索语言则是将信息组织与信息检索联系起来的纽带。

一、信息检索语言的种类

现在的信息检索系统,由于信息组织方式、用户群体、学科覆盖范围等不同,往往采用不同的检索语言,以适应不同的检索特性要求。据统计,单就检索语言来看,目前世界上已存在数千种,在目前常用的文献检索系统中,我们常能见到系统内采用多种检索语言,形成多种不同的检索途径的情况。尽管如此,我们仍然需要对其分类做一个简单介绍。

信息检索语言有多种分类方式,本教材从文献揭示的角度来划分,将其分为两大类。第一大类是揭示文献自身特征的检索语言。它又可分为两类,一是揭示文献自身外部特征的检索语言,如题名、责任者、出版单位、代码序号等,二是揭示文献内部特征的检索语言,主要有分类语言和主题语言。分类语言是历史最为悠久的一种检索语言,它按学科性质进行分类和排列。主题语言则以事物为聚类核心,它虽然仅有百余年的历史,但在信息检索领域占据了极其重要的地位,尤其是检索性能较为完善的叙词语言,在计算机检索领域得

到了广泛的认可和应用。第二个大类是揭示文献之间结构关系的引文检索语言,它利用文献之间的引用与被引用关系,作为检索中文献主题之间的相关关系,无需词表,也不必标引文献,检索简单而有效。引文语言以其独特的检索方式、索引结构等在传统目录学基础上产生了重要突破,从而获得国际信息界的认同。下面将对目前国内外影响最大的分类语言及医学领域内的主题语言作重点介绍,引文语言请参见第五章第一节。

二、分类语言

分类语言是将表示各种知识领域的类目按知识分类原理进行系统排列并以代表类目的数字、字母符号作为文献主题标识的一类情报检索语言,亦称分类法。文献内容属于某个类目的范围,就用该类目的分类号标引,并反映在与分类体系一致的序列中的特定位置,分类语言的优点是逻辑结构清晰。分类语言也可以分为等级列举式和分面组配式两种。等级列举式分类语言属于先组式语言,分类体系明显,容易理解,但因其采用列举式分类方法和类目的单线排列方式,所以存在着不能无限容纳概念的局限性和集中与分散的矛盾。分面组配式分类语言就本质而言属于后组式语言,它基本上克服了等级体系分类语言的缺点,但其分类体系不够明显,理解起来有一定困难。使用分类检索语言建立的文献情报检索系统能够使检索者鸟瞰全貌、触类旁通,对系统地掌握和利用一个学科或专业范围的知识更为便捷有效。

国内外比较重要的分类语言有《中国图书馆分类法》、《国际十进分类法》、《杜威十进分类法》、《国际专利分类表》等。

(一) 中国图书馆分类法

《中国图书馆分类法》是我国编制出版的大型综合性分类法,是目前国内图书馆使用最为广泛的分类法体系,简称《中图法》,现为第四版。它是目前我国应用最为广泛的图书分类法,不仅用于图书馆图书的分类、排架、组织目录等,也用于其他文献,如期刊论文的分类、数据库的检索等。

《中图法》由类目表、注释和说明、标记符号、索引 4 部分构成,类目表是其主体,共分为 5 个基本部类,22 个一级类目。此分类法采用汉语拼音字母与阿拉伯数字相结合的混合号码,用一个字母代表一个大类,以字母顺序反映大类的次序,在字母后用数字作标记。为适应工业技术发展及该类文献的分类,对工业技术二级类目,采用双字母,各类目按概念之间的逻辑关系逐级展开,当分类号的数字超过三位时,为了醒目加上小圆点。其一级类目名称如下。

- A 马克思主义、列宁主义、毛泽东思想、邓小平理论
- B 哲学、宗教
- C 社会科学总论
- 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
N 自然科学总论
O 数理科学和化学
P 天文学、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
Z 综合性图书

医药卫生类基本类目表

R1 预防医学、卫生学
R2 中国医学
R3 基础医学
R4 临床医学
R5 内科学
R6 外科学
R71 妇产科学
R72 儿科学
R73 肿瘤学
R74 神经病学与精神病学
R75 皮肤病学与性病学
R76 耳鼻咽喉科学
R77 眼科学
R78 口腔科学
R79 外国民族医学
R8 特种医学
R9 药学

在类目表展开时,不少类目多需要采用相同的划分标准,从而得到相同的子目。为了缩小类表的篇幅,同时也为了增强分类表类目设置的规范性和伸缩性,一般分类法均会采用复分表的形式,《中图法》也不例外,它的复分表可分为通用复分表和专用复分表两种。通用复分表是供整个文献分类法有关类目共同使用的复分表,如总论复分表、世界地区表、中国地区表、国际时代表、中国时代表、世界种族与民族表、中国民族表、通用时间地点表等。专用复分表只适用于某一基本大类或专门学科,一般设置于相应门类之中。《中图法》(第四版)共设有 67 个专用复分表,如在 R25/278 中医各科及中医急症学类目下,设置了专门的复分表,如下。

- 01 预防和控制
- 02 病理和病因
- 03 免疫
- 04 诊断
- 05 治疗
- 06 并发症
- 07 预后
- 08 诊疗器械
- 09 急症

注释和说明是图书分类法的重要组成部分,它可以帮助分类人员正确理解图书分类法,明确类目的含义,保证文献资料归类的准确性。注释是对类名的补充说明,主要用以解释类目的内容范围,指明类目之间的关系,规定了类目的细分方法及某一类的特殊分类规则。《中图法》(第四版)的说明分为编制说明和使用说明两种:编制说明包括编制原则、体系结构、编号制度、标记符号等方面的问题;使用说明则阐述了各大类的结构原理、一般分类规则、特殊分类规则等。

(二) 国际十进分类法

《国际十进分类法》,即 UDC(Universal Decimal Classification),又称为通用十进制分类法,是目前世界上规模最大、用户最多、影响最广的一部文献资料分类法,现有 20 多种语言的各种详略版本。近百年来,UDC 已被世界上数十个国家的 10 多万个图书馆和情报机构采用,已成为名副其实的国际通用文献分类法。目前,国内的硕士、博士论文系统均要求使用 UDC 分类号,便于信息的交换和处理。

《国际十进分类法》是通用型分类法,包括所有学科,其版本依照详略程度,可分为四类:足版、节版、缩版和口袋版。其中,足版类目最为完整,多作为校对的基础;节版用得最多,其内容为足版的 1/3 左右;缩版约占足版内容的 1/10;口袋版则纯供教学用。

UDC 由主表、辅助表及索引组成。

1. 主表 分为以下十大类。

- 0 总类、科学和知识
- 1 哲学、心理学
- 2 宗教、神学
- 3 社会科学
- 4 (语言)
- 5 数学和自然科学
- 6 应用科学、医学、技术
- 7 艺术、娱乐、体育
- 8 语言、语言学、文学
- 9 地理、传记、历史

其中的第四类现已被归并到第八类,空出位置拟作扩充科技类目之用。

下面以医学为例作说明。

- 6 应用科学、医学、技术

61 医学

611 解剖学,人类和比较解剖学

612 生理学,人类和比较生理学

613 卫生学总论,个人健康和卫生

613.6 职业危险,工业健康卫生

614 公共健康卫生,意外事故防止

614.7 空气卫生学,水,土壤污染和控制

614.71 空气卫生,空气污染

614.715 尘土、烟造成的空气污染

614.76 土壤空气共同污染,动植物分解,动物粪便

614.777 水,水卫生

614.8 事故,危险,灾害,事故防止,个人保护,安全

615 药理学,治疗,毒物学

615.3 药剂

616 病理学,临床医药

617 外科,整形,眼科

618 妇科,产科

619 兽药

UDC 采用单纯阿拉伯数字作为标记符号。它用个位数(0~9)标记一级类,十位数(00~99)标记二级类,百位数(000~999)标记三级类,以下每扩展(细分)一级,就增加一位数,每三位数字后加一个小数点。

2. 辅表 UDC 的辅助表有语言、文献类型、地点、民族和种族、时间、观点、材料、人物等 10 个,如下所示。

1a 联结-附加

1b 关系-子类-固定顺序

1c 语言复分

1d 形式复分

1e 地理复分

1f 种族与国籍复分

1g 时代复分

1h 非 UDC 标记

1i 属性复分

1k 特性复分(材料及人物)

辅表用加号、冒号等符号连接,表达类号之间的关系。通用辅助表适用于整个主表,表示地方、语言、时代、媒体等出现于各主题的概念。专属辅助表适用于特定的主题。

《国际十进分类法》是全球公认的分类标准之一,它在等级列举制的基础上采取了多种符号进行组配,因而发展成为一部等级列举与组配相结合的混合式分类法,具有不受特定语种限制、类号组合方式灵活、足以表达文件或资源的复杂性及详细程度等优点。当然,它也有其自身缺点,如更新速度不够快、部分主题的类目设置太过简略、使用太复杂等。