

信息检索 方法与实践

主编 汪楠 张炎



東北大學出版社
Northeastern University Press

© 王楠 炎 2007

图书记录编目 (CIP) 数据

信息检索方法与实践 / 王楠, 张炎. — 沈阳: 东北大学出版社, 2007.2

信息检索方法与实践

中国版本图书馆 CIP 数据核字 (2007) 第 093929 号

主编 汪楠 张炎
副主编 成鹰

本书是信息检索专业的一门专业基础课教材。本书以信息检索的基本概念、基本原理、基本方法、基本工具、基本流程、基本应用为主线, 结合大量的实例, 系统地介绍了信息检索的基本知识、基本方法、基本工具、基本流程、基本应用。本书可作为高等院校图书馆学、情报学、档案学、信息管理专业及相关专业的教材, 也可供从事信息检索工作的工程技术人员参考。

出版 东北大学出版社

地址: 沈阳市和平区文化路 3 号 11 号

邮编: 110004

电话: 024-83687331 (编辑部) 83680587 (社务室)

传真: 024-83680180 (编辑部) 83680582 (社务室)

E-mail: newsp@nupress.com

http://www.nupress.com

印刷 沈阳中科印刷有限公司

发行 东北大学出版社

幅面尺寸: 184mm×260mm

印 张: 10.5

字 数: 289 千字

出版 2007 年 2 月第 1 版

印刷 2007 年 2 月第 1 次印刷

责任编辑: 王楠 张炎

封面设计: 王楠

版式设计: 王楠

文字编辑: 王楠

东北大学出版社

元 18.00 : 价 宝

· 沈 阳 ·

ISBN 978-7-81105-390-9

© 汪楠 张炎 2007

图书在版编目 (CIP) 数据

信息检索方法与实践 / 汪楠, 张炎主编. — 沈阳: 东北大学出版社, 2007.5

ISBN 978-7-81102-390-9

I. 信… II. ①汪 ②张… III. 情报检索 IV. G252.7

中国版本图书馆 CIP 数据核字 (2007) 第 063959 号

内容提要

本书是通用性很强的有关信息检索方法和技能的著作。书中阐述了在现代信息技术环境下, 信息检索的相关概念、基本原理以及常用技术和方法, 从多种信息资源检索出发, 详细介绍了国内外常用的学术性文献数据库以及其他信息资源数据库, 系统介绍了国内外著名检索工具的结构和使用技巧, 阐述了信息检索的意义和信息资源的再生利用方法。同时, 借助大量的检索范例进行讲解, 以使读者能有效地掌握现代网络环境下数字化信息资源的检索和利用。本书结构合理、内容全面、重点突出、注重实用, 可作为高等院校信息检索课程的教学用书, 也可作为教学、科研、工程技术和社会各界信息用户检索信息的参考书。

出 版 者: 东北大学出版社

地址: 沈阳市和平区文化路 3 号巷 11 号

邮编: 110004

电话: 024—83687331 (市场部) 83680267 (社务室)

传真: 024—83680180 (市场部) 83680265 (社务室)

E-mail: neuph @ neupress.com

http: //www. neupress. com

印 刷 者: 沈阳中科印刷有限责任公司

发 行 者: 东北大学出版社

幅面尺寸: 184mm×260mm

印 张: 10.5

字 数: 269 千字

出版时间: 2007 年 5 月第 1 版

印刷时间: 2007 年 5 月第 1 次印刷

责任编辑: 牛连功 向 荣

封面设计: 唐敏智

责任校对: 张淑萍

责任出版: 杨华宁

ISBN 978-7-81102-390-9

定 价: 16.00 元

前言

以计算机、网络技术、通讯技术为代表的信息技术的迅猛发展,加快了信息交流与传递的速度,扩展了可利用的信息空间,使人们完全进入一个经济信息化、社会信息化的知识经济时代,信息日益成为社会各领域最活跃、最具决定意义的因素。

信息社会对人的挑战表现为对人的综合能力的挑战,而信息素质是信息社会中人的综合能力的重要组成部分。国民是否具备相当的信息素质和掌握足够的信息技术,已成为影响一个国家竞争力的重要方面。信息素质是一种能够发现信息需求、查寻、分析、判断、加工、筛选、综合利用、创造信息的各方面能力的总和。培养信息素质,就是培养人的独立自主学习的态度和方法,使之具有追求新知识、运用新信息的能力,善于从瞬息万变的事物中捕捉信息、创造新信息。信息素质已成为每个社会成员的一种基本生存能力,更是学习型社会及终身学习的必备素质。信息素质的培养,日益成为世界各国教育界乃至社会各界所关注的理论与实践的重大课题。

信息资源是人们进行学习、生活、科学研究和事业发展的基础,而信息资源的有效存储和检索作为人们获取信息资源的主要手段,也已经成为社会所有成员必须掌握的基本知识和技能。目前,“掌握资料查寻、文献检索及运用现代信息技术获取相关信息的基本方法”已被规定为大多数专业的人才培养要求。可见,加强信息检索课的教学,增强学生的信息意识,重视培养学生的收集处理信息的能力和获取知识的能力,已经成为当今人才培养的重要任务之一。

全书共9章,内容包括信息检索基础、常用中文数据库检索系统、经济信息检索、社会科学信息检索、科技信息检索、国外科技信息资源、特种文献检索、信息资源的利用和检索实例。本书从检索基本原理出发,针对不同信息资源的检索需求,注重检索方法和常用检索工具的介绍,并为读者提供了大量的检索实例,详细介绍了检索分析、检索策略的制定、检索结果的甄别以及检索系统评价等方法,同时,进一步为读者提供了信息资源的收集、整理以及科技文献的写作方法。其中第1, 5, 6, 7章由张炎编写;第2, 3,

4, 8章由汪楠编写;第9章由成鹰编写。汪楠完成全书的统稿工作。王妍、李雪飞、刘飒、陈敬参与了本书的校对和审读。

本书编撰过程中参考了大量信息检索文献,借鉴了许多网上资源。如万方数据资源系统、维普中国科技期刊数据库系统、中国知网等,在此恕不一一注明,谨向相关机构和作者致以诚挚的谢意。

由于作者水平有限,书中如有疏漏和不足之处,恳请专家和读者批评指正。

作者

2007年春

目 录

第 1 章 信息检索基础	1
1.1 信息检索概述	1
1.2 检索工具	6
1.3 检索语言	10
1.4 计算机信息检索基础	12
第 2 章 常用中文数据库检索系统	18
2.1 万方数据资源系统	18
2.2 维普中国科技期刊全文数据库	27
2.3 中国知网 (CNKI)	34
第 3 章 经济信息检索	44
3.1 经济信息检索的内容和工具	44
3.2 网络经济信息资源	49
3.3 中国经济信息网	56
3.4 国务院发展研究中心信息网	61
第 4 章 社会科学信息检索	64
4.1 社会科学信息检索的内容	64
4.2 社会科学信息数据库	72
4.3 社会科学网络信息检索	75
第 5 章 科技信息检索	82
5.1 科技信息检索的内容	82
5.2 国内科技信息资源	93
第 6 章 国外科技信息资源	98
6.1 美国工程索引 (EI)	98
6.2 (ISI)	103
6.3 美国《金属文摘》	110

6.4	美国《生物学文献》	112
6.5	英国《科学文摘》	114
第7章	特种文献检索	117
7.1	专利文献	117
7.2	中文专利	124
7.3	标准文献	128
第8章	信息资源的利用	135
8.1	文献信息服务系统	135
8.2	科技查新	139
8.3	信息资源应用	143
第9章	检索实例	150
实例 1:	关于水污染问题的研究	150
实例 2:	“玉米燃料乙醇产业化”项目研究	153
实例 3:	一个利用网络搜索的实例	153
	检索思考题	155
附录: 图书馆网址		158
参考文献		160
02		3.3
10		4.3
40		章 1 第
40		1.4
55		2.4
55		3.4
58		章 2 第
58		1.2
59		2.2
89		章 3 第
89		1.0
103		0.2
111		3.0

第1章 信息检索基础

1.1 信息检索概述

1.1.1 信息检索及其类型

1.1.1.1 信息的特征

①客观性。信息既不是物质，也不是能量，是客观事物普遍性的表征。信息是无处不在、无时不有的普遍社会现象。

②传递性。信息在事物之间的相互联系必定在信息的流动中发生。信息的传递性表现在人与人之间的信息交换，人与计算机之间的信息交换，动物界及植物界之间的信号交换。同时，人类进化过程中的细胞选择、遗传也是信息的传递与交换。

③广泛性。信息在不同的领域具有多种不同的特性或表现形式。例如：客观事物中的各种自然属性；人工设备的技术特征；人类社会的各种社会特征；人脑中反映客观事物认识的思想、知识；人类交流信息过程中的声音、文字、图像以及用各种编码形式记录下来的数据、新闻、情报、消息等。各种形式的信息又常常以综合的方式表现事物的特征，所谓“多媒体”正是信息多样性和综合性的集中表现。

④相对性。从信息作为事物相互联系的角度看，信息源不确定的程度或者信息源接受信息量的多少，均与信息接受者（信宿）的状态有关。同一信息对具有不同认知水平的人所产生的作用和有效性是不相同的。

⑤政策性。信息的政策性较强。信息使用者对国家、部门、地区制定的国民经济、社会发展、科技工作等方面的方针政策应认真领会，并作为工作中的指导思想贯彻执行。否则，在工作中就会走弯路。

⑥积累性。信息通过人脑思维或人工技术设备的综合、加工和处理，不断积累、丰富，提高其质量和利用价值。正是由于文献的累积，使得全世界每一个国家的科学发现、发明得以保存，并流传至今。基于此，科学家、社会学家可以将全世界各个时期的科学成就作为研究基础，开展自己的科技、社会工作。

⑦价值性。信息的质量和价值，实际上是对客观事物属性反映的深度和真实程度的认识。信息的价值具有相对性，同一信息因使用对象的不同而具有不同的价值。信息只有被利用才会产生价值。

1.1.1.2 信息检索的概念

信息检索，目前尚无公认一致的定义。国内外有关专家从不同角度解释信息检索，较有代表性的观点主要有以下几种。

①信息检索的范围较为广泛。动态信息、静态信息、音频信息、视频信息及各种数值信

息均属信息检索范围。如果将信息检索技术作为一门学科,它应该包括矩阵记数法、概率论、最优化理论、模式识别及系统分析技术等各学科领域的内容。

②信息检索主要是文献检索。信息检索是从大量的文献中查寻与情报提问所指定的课题(对象)有关的文献,或者是包含用户所需事实与消息的文献过程。这里所谈到的文献,不仅指文献线索,也包括文献的片断,如章、节、段落以及与事实有关的直接情报等。

我们认为信息检索是将信息按一定的方式组织起来,并根据用户需求找出相关信息的过程。这里的信息检索是指信息的存储与检索,是针对信息工作者和用户来定义的。如果仅针对用户,信息检索是指在信息集合中找出所需信息的过程。

1.1.1.3 信息检索的类型

(1) 按存储和检索的内容划分

信息检索按存储和检索的内容可划分为文献检索、数据检索和事实检索。

①文献检索。通常指的是在检索系统中检索以二次文献为对象(目录、索引、文摘)的信息,它们是文献信息的外部特征与内容特征的描述集合体。信息用户通过检索获取的是原文的“替代物”。

②数据检索。指在检索系统中检索的数值型数据,如科学技术常数、各种统计数据、人口数据、气象数据、市场行情数据、企业财政数据等,即事物的绝对值和相对值的数字。检索系统提供一定的运算推导能力,例如外推、内插、填补空缺数据,甚至具有列出曲线图或进行各种分析等功能。信息用户可以利用通过检索获得的经过核实、整理的数值信息再作定量分析。

③事实检索。指在检索系统的原始文献中检索关于某一事物(事件、事实)发生的时间、地点和过程(情况)等方面的信息。它是数值信息和系统数据信息的混合。一般先从系统中检索出所需信息后,再加以逻辑推理才能给出结论。这类信息主要是用于管理决策。

(2) 按系统中信息的组织方式划分

信息检索按系统中信息的组织方式可划分为全文检索和多媒体检索。

①全文检索。是指检索系统中存储的是整篇文章乃至整本书。检索时,用户可以根据需要从中获取有关的章、节、段、句等信息,还可以进行各种频率统计和内容分析。随着计算机容量的增大与运算速度的提高,全文检索正迅速扩大到各学科、专业。

②多媒体检索。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(形)以及声音等多种媒体信息。信息的存储结构从单维发展到多维,存储空间范围在不断扩大。

1.1.2 信息检索工作

1.1.2.1 信息检索的目的

现代社会信息丰富多样,知识的存储过于庞大和无序,出版物的泛滥使人们难以获得大量准确的信息。知识存储的无序化,加剧了“信息污染”的程度,堵塞了通向知识之门的道路,耗费了大量探索自然和社会规律的精力,以至于人们不得不认真学习和研究获取信息的方法,掌握从大量无序知识中搜索对自己有用的、准确的知识的技能。因此,信息检索已经构成知识体系中一个不可缺少的部分。

信息检索可以在信息用户与信息源之间充当媒介的作用。信息检索的任务就是将用户与

信息源经济地、有效地结合在一起；其目的是促进资源开发，协助管理者作出正确决策，帮助研究人员继承和借鉴前人成果，避免重复研究或走弯路，节省研究人员查找文献的时间。

1.1.2.2 信息检索工作

(1) 信息存储

信息的存储方式直接影响信息检索的速度和效率，影响查全率、查准率、误检率等评价检索质量指标的好坏。

信息存储可以利用自然语言或规范化语言。自然语言是指作者的书面用语。在传统的文献检索系统中，为了提高文献的检索效率，文献工作者创建了规范化语言。规范化语言是文献标引和文献检索共同使用的语言，它使文献描述和表达概念具有唯一性和专指性，便于提高查准率，标引时可以集中相关文献，提高查全率；它能显示概念间的各种关系，有利于及时调整检索策略等，检索时能达到较高的查准率，同时提高检索速度。

(2) 数据库与文档

信息的存储基本上是靠建立系统的单位自己来组织和构建各类信息的文档。文档的集合叫做文献数据库，即通常所指的文献型、数值型、事实型及全文信息的集合。如何合理构建数据库，是有效进行检索的关键。有关数据库的研究，包括文档编制时对文献的选择，根据检索语言对文献进行标引，数据库文档结构的描述、组织与更新，数据库的质量、使用效果及其评价方法、评价原则，数据库市场及版权问题，等等。

(3) 检索系统

检索系统是一定的目标、工具（印刷型或计算机）、信息资源（数据库）及检索方法与策略的集合体，一切检索都是通过检索系统实现的。检索系统的研究内容包括存储信息的广泛性和检索的迅速准确性。网络信息检索系统研究的内容主要包括系统硬件配置、软件的设计与分析、系统的评价、系统的智能化应用等方面。

(4) 检索策略

检索策略是在处理信息需求的基础上确定检索途径、检索词语，并明确各词语之间的逻辑关系、查找步骤、系统输出顺序等方面的科学安排。检索策略与检索效果密切相关。正确的检索策略可以优化检索过程，以最小的花费、最短的时间获得最佳的检索效果。

检索策略的研究主要包括以查全为目标的检索策略的调节与控制，以查准为目的的检索策略的调节与控制，以最小投入为目标的检索策略的调节与控制，检索策略失误的分析，系统对检索策略的自动修改机制的研究，等等。

(5) 检索工具

信息，资源浩瀚，形式多样。要快速、准确地查询到所需信息，必须借助各种检索工具，并根据不同的检索目的确定不同的检索途径和检索方法。

(6) 检索服务

检索服务是信息服务的重要环节，包括手工检索服务、机检服务、国内外联机检索服务、光盘检索服务、网络检索服务、定时检索服务、定题检索服务和周期检索服务等。

1.1.2.3 信息检索程序

无论是手工检索还是计算机检索，都是一个经过仔细地思考、并通过实践逐步完善查找方法的过程。整个检索过程通常包括以下几个步骤。

(1) 分析课题

在确定研究课题后,研究者应知道所需信息的学科性质、特点和水平层次,明确检索的学科范围——越具体越有利于检索。如果问题属于多学科或交叉学科,则可分别列出多学科或交叉学科的相关部分,确定所需信息在复杂的学科体系中的位置,从分类途径检索时选好上下位类(上位类俗称大类,下位类俗称小类,在多极分类系统中,上位类可分成若干下位类。主题途径检索时分上位关键词、下位关键词,一个上位关键词下通常列出多个下位关键词),从主题途径检索时搞清上下位关键词之间的关系,来扩大或缩小查找范围。

(2) 选择检索工具

检索工具成千上万,除了要知道哪类问题该用哪类工具书解决之外,还要熟悉一些具体工具书的内容、范围、特点、编排结构,工具书之间的相互关系,包括内容和时间的联系,有的放矢地进行查找。

确定工具书类型之后,还要根据编撰目的、材料收录范围和编排方法在同类型工具书中选择有希望获得所需信息的具体的检索工具。任何检索工具都是根据特定的目的和读者群编制的。检索工具收录材料的学科范围、地区范围、语种范围、时间范围和文献类型,在每一种检索工具书的用户指南(或导言、前言、编例、User's Guide、Helps for the reader等)中都有说明。这是首次使用某一检索工具前必读的内容。

(3) 确定检索途径及检索式

检索工具选定以后,应确定检索途径。检索途径往往不止一种,使用者应根据“已知”信息特征确定检索途径。所有文献的特征可分两大类,即外部特征(著者、序号、文献类型、出版时间、语种、文献号、代码等)和内容特征(分类、主题、文献篇名)。检索时,如果已知文献的外部特征,尽量使用该外部特征进行检索,以便达到信息检索更高的专指度。在不知文献外部特征的情况下,通常采用内容特征进行检索。通常采用内容特征进行检索时,希望泛指度高时通常选择分类途径,希望专指度高时采用主题途径。有些检索工具仅采用一种方式进行标引,用户只能根据已有的标引方式进行检索。

检索式是计算机检索中用来表达检索提问的一种逻辑运算式,又称检索表达式或检索提问式。它由检索词和检索系统允许使用的各种运算符组合而成,是检索策略的具体体现。构造检索式就是把已经确定的检索词和分析检索课题时确定的检索要求用检索系统所支持的各种运算符连接起来,形成检索式。

(4) 进行检索

在检索系统中将检索标识与系统中存储的文献标识进行匹配,查出相关文献,并对所获结果进行分析,看其是否符合需要。如果试检结果满意,可进行正式检索;否则,要分析原因,修改和调整检索策略。调整检索策略包括修改检索式、调整检索词、重新选择检索系统等。

(5) 获取原文

从检索工具上获得所需文献线索后,利用馆藏图书目录或报刊目录获取原文。馆藏目录或联合目录一般提供文献的索取号和收藏地址,获取原文时只要履行借阅手续或馆际互借手续即可。有时还可从著者处获取原文。许多网络检索工具可以提供原文的链接,只要用户支付相关费用便可方便地获取原文。如果不能在网络上获取原文,应收集以下信息,通过这些信息获取原文。

①识别文献类型。不同类型的文献,收藏地点不同、外表特征不同,在索取原文时首先

就要区别文献的类型。

②充分利用刊名。检索工具通常采用刊名缩写,并附有引用出版物一览表,供查对收录的出版物全称用。

③论文著者的地址。获取会议论文、学位论文、公司报告以及一些尚未公开发表的文章的原文,必须获得论文著者(包括团体著者)的详细地址。大多检索刊物的著录项中附有作者的工作单位名称缩写,可以据此查阅机构名录全称。

④识别语系。在检索工具中,俄文、中文、日文等的文献作者、出版物名称通常采用音译法转换成英文进行著录,索取原文前,要将这些音译的人名、出版物名称还原成原来的语种。

⑤其他外部特征。文献的其他外部特征包括文献篇名、编号(索书号、专利号等)、著者姓名、出版时间、分类号等。

了解文献的上述特征后,采用由近及远的方式进行查找,可通过以下途径获取原文。

其一,本单位图书情报部门。这是获取原文最方便、最经济的途径。应首先立足于本馆(或本所),其次是附近的图书馆或信息中心。如果读者目录有缺省,应利用公务目录、典藏目录或各院系分馆的目录,这类保存较好的、能完整反映馆藏的图书期刊目录,或通过馆际协作获取原文。

其二,从著者获取原文。给国外的论文著者写信索取复制件,已是国际上通行的学术交流方式。美国有一些情报机构开展这方面的服务,如研究图书馆中心(CRL)、国家技术情报服务中心(NTIS)、美国专利商标局(PTO)等。亦可以从学术团体指南这一类工具书中获得更详尽的情况。

其三,从检索刊物出版机构获取原文。国外一些著名检索机构,如美国化学文摘社(CAS)、情报科学研究所(ISI)等,都可向用户提供原文。ISI设有原迹论文服务,可使用户看到彩色版图、照片及复杂的图像资料。有些非盈利性出版物是商业书目所不收录的,要获取这些文献,可以通过学术机构或信息服务名录获得有关信息,然后通过函索方式获取出版物。

其四,利用国际联机检索终端向国外订购原文。如果读者急需原文,而一般订购渠道又很慢,则可以采用这个途径。订购时需填写“联机订购原文申请单”,注明文档名称及代号、原文题目、著者、出处等。这是一种较快的订购办法,一般半个月左右,但费用昂贵。

1.1.3 信息和文献的分类

1.1.3.1 信息的分类

按信息的内容划分,可分为科学技术信息和社会科学信息。科学技术信息包括自然科学信息和工程技术信息等;社会科学信息包括政治信息、军事信息、经济信息、技术经济信息和社会生活信息等。

按产生信息客体的性质来划分,信息可分为自然信息(瞬时发生的声、光、热、电、形形色色的天气变化、缓慢的地壳运动、天体演化等)、生物信息(生物为繁衍生存而表现出来的各种形态和行为,如遗传信息、生物体内信息交流、动物种群内的信息交流)、机器信息(自动控制系统)和社会信息(人类)。社会信息是指人与人之间交流的信息,既包括通过手势、身体、眼神所传达的非语义信息,也包括用语言、文字、图表等语义信息所传达的

一切对人类社会运动变化状态的描述。

按信息所依附的载体来划分,可分为文献信息、口头信息等。文献信息,就是以文献为载体的信息,以文字、符号、音像信息为编码信息,是经人们筛选、归纳和整理后记录下来的,它与人工符号本身没有必然的联系,但要通过符号系统实现其传递。口头信息是原始的没有形成文字、符号、音像的信息,不如文献信息易于流传。

1.1.3.2 文献的分类

文献信息是信息存在的一种重要形式,其优点是易识别、易保存、易传播,可使人类精神文明千古流传。文献是信息的主要载体,大多规范化的重要信息都以文献形式出现,所以文献的分类非常重要。

(1) 按加工层次分

文献是信息的主要载体,根据对信息的加工层次可将文献分为一次文献、二次文献和三次文献。

①一次文献。指人们对自然和社会信息进行首次加工而成的文字记载,是文献信息源的主要部分。例如专著、报纸、期刊、专利文献、标准文献、会议文献、样本等成品文献,都属于一次文献(Primary literature)。一次文献数量极为庞大,由于在内容上是分散的、无系统的,不便于管理和传播。

②二次文献。为了控制文献,便于人们查找,对一次文献进行再加工,通过整理、提炼和压缩,并按其外部特征(题名、作者、文献物理特征)和内容特征进行有序化管理,形成另一类文献形式就是二次文献(Secondary literature),例如目录、书目、索引、文摘等。二次文献不是一次文献本身的汇集,而是一次文献特征的汇集,通过它们可以很方便地找到一次文献,或了解一次文献的内容。二次文献内容相对集中,系统性强,便于管理和传播。

③三次文献。利用二次文献,选择有关的一次文献再加以分析、综合而编写出来的第三个层次的文献形式就是三次文献(Tertiary literature),例如专题报告、综述、进展以及手册、百科全书、年鉴等工具书。三次文献具有系统性、综合性、知识性和概括性的特点,它从一次文献中汲取重要内容,便于高效率地了解某一领域的状况、动态、发展趋势和有关信息。因此,要在浩瀚的一次文献中查找所需资料,往往离不开二次文献和三次文献。

(2) 按文献的载体分

①缩微文献。指以感光材料为载体、用缩微照相技术制成的文献复制品。按其外形可分为卷片型和平片型;按对它的穿透力可分为透明体和不透明体,前者需用透射式阅读机阅读,后者指缩微印刷品,用不透明感光纸印制而成,用反射式阅读机阅读。

②音像文献。指以磁性材料、光学材料为记录载体,利用专门的机械电子装置记录与显示声音和图像的文献,又称音像资料、视听资料、音像制品。音像文献脱离了传统的文字记录形式,直接记录声音和图像,给人以直观的感觉,又称直感型资料。

③硬磁盘和光盘文献。这类文献的最大优点是能够随机存取所需数据,存取速度快,适用于大容量检索。

1.2 检索工具

检索工具是指为检索需要而组织的文献检索系统。它是收集、加工、标引、组织文献的

结果,是检索工作赖以生存的手段和条件。检索工具包含目录、索引和文摘等。

1.2.1 检索工具的结构

检索工具的结构与检索工具的功能和效率密切相关。典型的检索工具通常由主体、辅助索引、使用说明和附录四部分组成。

(1) 主体

主体即检索工具的主要部分,由所收录文献的著录部分组成,是文献储存与检索的实体。包含文献的内部特征与文献的外部特征。内部特征如分类系统、主题系统和文摘等,外部特征包含书名系统、著者系统、编码系统等。

(2) 辅助索引

辅助索引是在文献检索工作中,从主体部分的文献款目编排系统以外的角度增加检索途径、扩大检索效果的辅助手段,它不能离开检索工具的主体单独存在。

(3) 使用说明

使用说明是检索工具的重要组成部分。任何一种检索工具都是通过使用说明向读者介绍本检索工具的性质、内容范围、收录年限、读者对象、著录方法、分类体系和检索方法等,帮助用户使用检索工具。

(4) 附录

附在检索工具后面的有关资料被称为附录。它包括出版物对照表、代码表等。

1.2.2 目录

1.2.2.1 目录的定义

目录是图书以及其他单独出版物的外部特征的解释和系统化记载,是将一批文献款目按照一定的次序或规则编排而成的一种文献工具。

1.2.2.2 目录的著录款目

目录的著录包括如下款目。

- ①书名(刊名)项;
- ②著者(编者)项;
- ③出版项:包括出版地点、出版者、出版期、版次等;
- ④稽核项:包括页数、图表、开本、装订形式、定价等;
- ⑤附注项:对上述各项的补充或说明;
- ⑥提要项:内容简介;
- ⑦业务注记:指图书馆或保存单位在目录上做的业务记载,包括索取号、登录号、分类号、主题词和存储地点等。

1.2.2.3 目录的排检

(1) 书名目录

①中文图书或期刊:按中文字顺排检,其方法有汉语拼音法、笔画法、部首法等。

②外文图书或期刊:西文书名或篇名按字顺排检。日文书名或篇名按汉字或假名排检,如按汉字排检,用笔画法;如按假名排检,利用假名顺序。俄文按俄文字母顺序排检。

(2) 著者目录

中文著者按中文字顺排列。外国著者按先姓后名排检。欧美著者一般名在前姓在后,编排时需将其倒置。

(3) 分类目录

分类目录是按分类法规定的图书、资料所属的学科、专业内容对应的分类号编目,并按分类号的顺序排列。分类号可以有阿拉伯数字、字母或两者结合组成。

(4) 主题目录

主题目录是按图书、资料的主题内容编排的目录。主题目录使用主题词的字顺排列。

1.2.2.4 目录的作用

目录能反映一定历史时期科学文化发展的概貌,是人们对浩如烟海的文献加以控制的有效手段,是查阅和利用文献必不可少的工具。目录的具体作用概括起来主要有如下几个方面。

①推荐作用。针对特定读者,按知识的连续性和发展的阶段性编纂和推荐文献,指导阅读,成为治学的门径,因此有人将目录比作在海洋中航行不可缺少的导航图。

②检索工具。帮助读者从特定角度去查找资料,例如题名、著者、主题等。理想的目录应具有分析的功能,能满足读者检索某一著作中最小的特殊单元。专业目录根据不同的研究需要选编不同学科内容、不同水平和各文种的资料,分门别类加以编排以供查阅,为专业研究提供方便。

③文献工作。用以核对各个著录项目,如著者、题名、版本、日期等。

④提供出处。有助于文献资料的收集选择、编目加工、报道推广和翻译评介等。

1.2.2.5 主要目录简介

(1) 国家书目

国家书目(National bibliographies),登记书目的主要类型之一,是全面揭示与报道一个国家出版的所有文献的系统目录。它有如下特征:所收录的材料主要是这个国家出版、印刷、发行的;用这个国家通用的语言编纂;收录文献所揭示的内容大部分是同这个国家相关的;大部分内容是不断更新的。

(2) 综合性书目

综合性书目是将各个学科门类的图书汇总编成的一种图书目录。其内容广博、包罗万象,既有哲学社会科学方面的书,又包括自然科学和应用技术方面的书;其层次不同,既有普及性读物,又有学术性著作。国家书目属于综合性书目。

(3) 专科学目

专科学目是反映某一学科或某一专题文献的工具。有的对所收图书还加以说明,作为指导阅读和提供查找资料的线索。

(4) 专题书目

专题书目是为某学科专业或某一研究课题编制的书目。它集中揭示某一方面的全部图书信息,具有较大的使用价值,从事科学研究应尽量使用专题书目。

(5) 机读目录

将目录数字化并记录在磁盘等载体上,通过计算机进行检索的目录称为机读目录。最初的机读目录产生于图书馆流通和编目系统,实际上是传统卡片目录的计算机化,其检索较手工方式快捷,使目录不仅具有二次文献功能,还能像一次文献那样满足用户多方面的需求。

(6) 联机目录

联机目录 (On-line catalog) 已经在很多图书馆得到运用, 如 OPAC 系统 (On-line Public Access Catalog), OCLC (Online computer Library Centre) 的联机联合目录。我国一些图书馆正在加紧建立自己的 OPAC 系统。

1.2.3 索引

1.2.3.1 索引的定义

索引是将报刊书籍中某些重要的或有意义的信息, 如书名、刊名、篇名、主题、人名和地名等分别择录出来, 按一定方式编排, 并注明出处, 以供检索的工具。它包括四个基本要素, 即索引源、索引款目、编排方法和出处指引系统。

索引不仅为读者提供多种文献检索途径, 并能通过检索词的使用, 反映某一文献的主题内容以及关于某一学科或课题的最新观点和发展趋势。

1.2.3.2 索引的作用

索引可以将文献所包含的若干信息分析摘录出来, 作为排检标识, 这种标识可以根据需要, 在一定的范围内进行任意深度的选择和标引, 从而满足更深层次或更多途径的查找要求。索引是在目录广泛使用的基础上, 随着对文献检索深度和广度的要求不断加大而产生的。在熟悉目录的基础上 (或者将两者配合使用), 就可以在宏观和微观两方面满足人们的信息需求。索引可以方便用户更快地达到检索目标, 是依附在目录或文摘等检索工具之上的辅助性工具, 不能独立存在。

1.2.3.3 常用索引

(1) 主题索引

主题索引是将文献中具有实质意义的词语或能揭示文献主体概念的词语抽出来, 经过规范化处理后, 再按字顺排列起来形成标识系统。通常, 各主题词下面给出副标题词, 或在各主题词下面给出篇名性的说明语、关键词性的说明语, 在说明语后列出文摘号。主题索引提供了一个按揭示内容的主要词汇进行查找文献的检索途径, 能迅速查找到专指度高的文献内容, 越来越受到文献工作者的青睐。

(2) 分类索引

分类索引是将文献内容所属的分类号按隶属关系排列起来, 并列出发分类号对应的类名。分类索引提供了一个按学科分类角度进行查找文献的检索途径, 可以迅速查到泛指度较高的文献。

(3) 著者索引

著者索引是以文献著者的姓名进行标引的索引, 按姓名的字顺排列。著者索引在著者名字后面列出该著者所著文献的文摘号, 读者根据文摘号可查阅文摘或原文。著者索引提供了一个按著者姓名进行查找文献的检索途径。

另外, 常用的索引还有著者所在单位索引 (或团体著者索引)、化学分子式索引、引用刊物索引、出版单位索引等。

1.2.4 文摘

1.2.4.1 文摘的定义

文摘是以提供文献内容梗概为目的, 简明、确切地记述文献主要内容的短文。我国国家

标准 GB 3468—1983 定义:“除题录部分外还对文献内容作实质性描述的文献条目标文摘。”文摘客观、如实地反映了原文献的内容,但又比原文文字简洁。作为一种检索工具,文摘将论文或书籍中摘录出来的主要论点、数据等按一定方式编排,供读者检索和阅读。

目录和文摘的共同点都是以论文、文章、报告等为主要的报道对象,其区别主要是以报道文献的外表特征为主还是以报道内容特征为主,在形式上看有无摘录或评价原文的文字材料。文摘提供文献的外表特征和内容特征,目录仅提供文献的外表特征。

1.2.4.2 文摘的作用

文摘的作用概括起来有以下四点。

(1) 帮助读者快速判断文献内容

通过目录查找文献仅能查阅文献题目,通过文摘查找文献不仅能查阅文献题目,还能查阅文献内容。文献题目虽然在一定程度上表达了文献内容,但与文摘相比相差甚远,利用文摘可以减少误检和漏检。

(2) 节省阅读时间

由于文摘是对原文献主要内容的描述,一条文摘实际上就是一篇高度浓缩、信息完整的文献,因此,读者在不需阅读原文献的情况下,可真实地了解该文献的内容,明确其基本要点,并可直接引用,从而节约阅读时间。

(3) 减少语言障碍

利用文摘,用户能获得因语言障碍无法得到的科学文献。在目前出版的科学文献中,有 50% 以上的文献是采用科学家没有掌握的语言出版的,这妨碍了国际间学术信息的传递、交流和利用。用本国语言或通用语言翻译的文摘,可以帮助读者克服语言障碍,了解国外有关领域的发展水平和趋势。

(4) 替代原文

由于文摘所摘录的是经过筛选的某一学科或某一专业领域最新最有学术价值的文献,并把相关论题集中一处,因而通过一组文摘,能获得该学科专业领域的学术概况和最新进展。在难以获得原文或不需获得原文的情况下,阅读文摘就显得非常重要了。

1.3 检索语言

检索语言是根据文献检索需求创造的一种人工语言,是用于表达一系列概括文献内容的概念及其相互关系的概念标识系统。

1.3.1 分类语言

分类语言是以学科体系为基础,用号码作为概念标识,按分类编排的检索语言。体系分类法是一种直接体现知识分类的等级制概念标识系统,是通过对概括文献信息内容及某些外表特征的概念进行逻辑分类(划分与概括)和系统排列而构成的。体系分类法的主要特点是按学科、专业集中文献,并从知识分类角度揭示各类文献在内容上的区别和联系,提供从学科分类进行文献信息检索的途径。

(1) 文献分类

文献分类是按文献所载信息的科学属性,对文献进行类别的划分,根据文献信息内容之