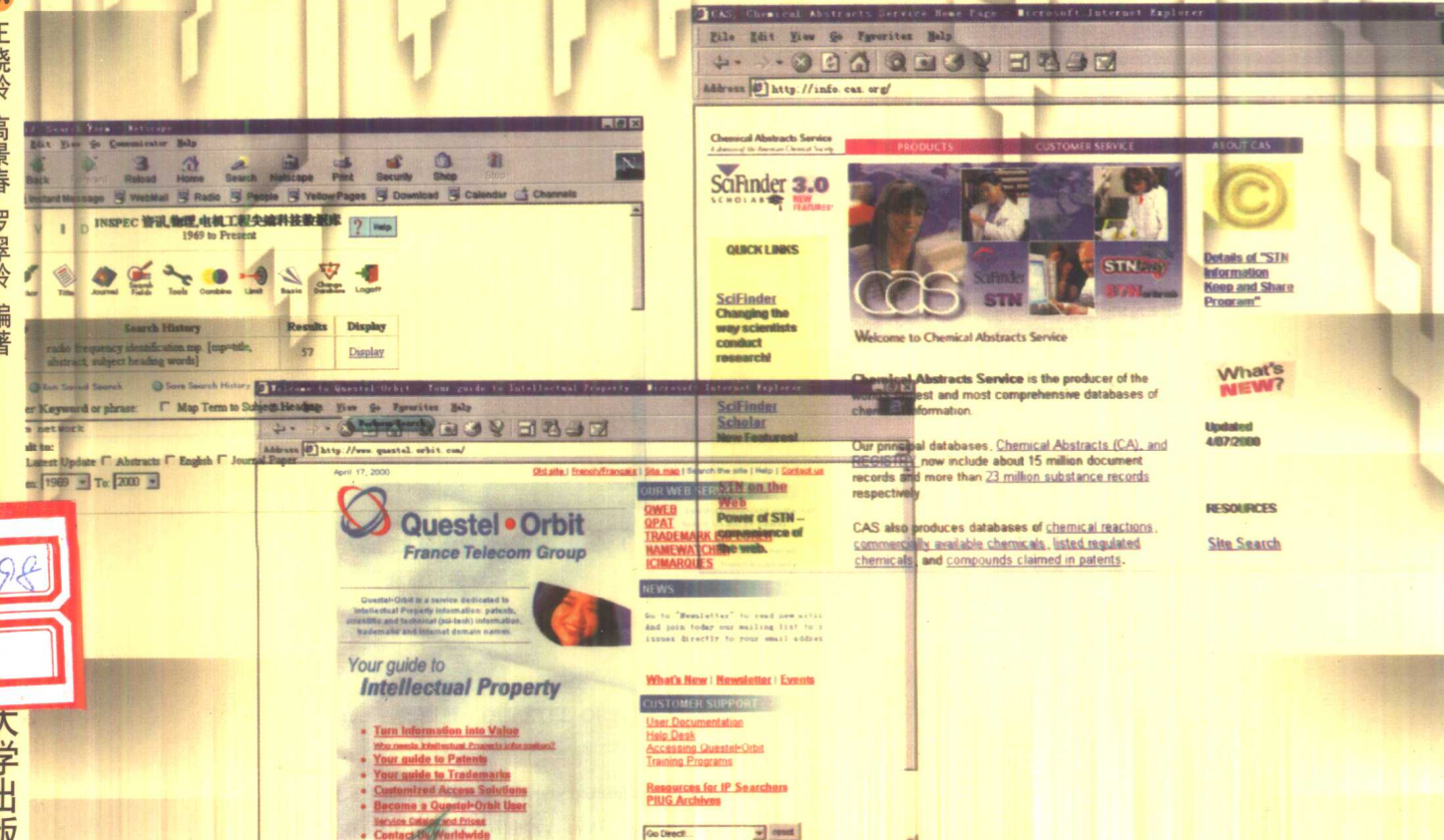


网络时代的 电子资源检索与利用

王晓玲 高景春 罗翠玲 编著



北京邮电大学出版社

网络时代的电子资源检索与利用

王晓玲 高景春 罗翠玲 编著

北京邮电大学出版社

· 北京 ·

图书在版编目(CIP)数据

网络时代的电子资源检索与利用/王晓玲,高锦春,罗翠玲编著. ---北京:北京邮电大学出版社,2000.7
ISBN 7-5635-0427-3

I. 网… II. ①王…②高…③罗… III. 计算机网络-文献检索与利用 IV. G354.2

中国版本图书馆 CIP 数据核字(2000)第 37742 号

-
- 书 名: 网络时代的电子资源检索与利用
Wangluo Shidai de Dianzi Ziyuan yu Liyong
编 著: 王晓玲 高景春 罗翠玲
责任编辑: 孙伟玲 时友芬
出 版 者: 北京邮电大学出版社(北京市海淀区西土城路 10 号)
邮编: 100876 电话: 62282185 62283578
网址: <http://www.buptpress.com>
经 销: 各地新华书店
印 刷: 北京市源海印刷厂
印 数: 3001 - 6000 册
开 本: 787 mm×1 092 mm 1/16 印张: 10.625 字数: 271 千字
版 次: 2000 年 7 月第 1 版 2000 年 12 月第 2 次印刷
书 号: ISBN 7-5635-0427-3/G·57
定 价: 20.00 元
-

前 言

当今是一个信息产业飞速发展、信息量急剧增加的时代,每一个人的工作都或多或少地需要建立在他人工作成果与信息积累的基础上,因此,借鉴、参考他人的经验、已有的技术和科学的方法已是非常经常与非常必须的工作。如何在今天急剧膨胀的信息海洋里,尽快获取对自己有价值的信息,特别是如何借助现代电子化、网络化的手段去获取这些有价值的信息是每一位科技工作者所必须掌握的技能。

近年来,中国国家科技部、国家教育部明文规定,将国际上三大检索系统收录的文献量作为衡量一所高校、一个研究所、一名教授的学术水平、学术成就高低的重要标准。这三大检索系统是美国《工程索引》、美国《科学引文索引》、美国《科技会议录索引》。这些大型数据库检索系统都是用英文环境开发的,且是付费的检索系统,因此,对于这些系统的文献收录范围、收录方式、检索方式的了解与掌握,无疑对我们检索自己发表的文献是否被收录,或增加自己论文的收入率都有十分重要的意义。

本书正是从上述这些角度,介绍国内外著名的各种数据库检索系统收录的信息范围、用户检索界面、使用方法等,给出了国内外各著名数据库系统的 WWW 站址以及这些站点上提供了哪些信息、哪些信息是免费的,等等。

全书共分十一章,其中用七章篇幅介绍了国外数据库检索系统,包括美国《工程索引》、《科学引文索引》、《科技会议录索引》等;用一章的篇幅介绍新兴的国内数据库光盘或联机数据库系统,包括中国学术期刊全文数据库、中国科技期刊数据库、万方数据等;另外,用三章篇幅介绍了专利知识与专利文献的检索。

全书由王晓玲统一构思编排,并对大部分章节进行了审核与修改。其中第一章由王晓玲、罗翠玲编写,第二章由罗翠玲、王晓玲、郑泓编写,第三、四、六、七、八章由王晓玲编写,第五、九、十、十一章由高锦春编写。

承蒙北京大学图书馆朱强研究员对全书进行了认真的审核、修改,并提出许多宝贵意见,在此深表最诚挚的感谢!

作 者

2000 年 7 月于北京邮电大学

目 录

第一章 科技信息检索的基本知识与概念	1
第一节 科技信息检索的意义	1
第二节 信息检索发展简史	2
第三节 检索语言与检索途径	2
第四节 文献的类型、识别与著录	11
第二章 国内重要信息源与检索工具简介	16
第一节 中文科技期刊数据库光盘	16
第二节 中国学术期刊全文数据库	19
第三节 中国科学技术信息研究所与《万方数据》	30
第四节 《人大复印报刊资料》数据库光盘	32
第五节 《经济学管理学论文题录》光盘数据库	36
第六节 中国高等教育文献保障系统	40
第三章 国外重要信息源与检索工具简介	41
第一节 Dialog 数据库系统	41
第二节 STN 数据库系统	43
第三节 美国化学文摘社	44
第四节 Questel-Orbit 数据库系统	45
第五节 OCLC First Search 数据库系统	46
第六节 IEEE 和 IEL 光盘	47
第七节 IEE 及其出版物	50
第八节 IHS 信息公司	51
第九节 UnCover 数据库	53
第四章 美国《工程索引》	56
第一节 概 述	56
第二节 印刷版《工程索引》	56
第三节 《工程索引》光盘	59
第四节 《工程索引》网上检索	68
第五节 EI 信息村	72
第五章 英国《科学文摘》	75
第一节 概 述	75

第二节	印刷版《科学文摘》	76
第三节	INSPEC 光盘	84
第四节	INSPEC 联机数据库及 Web 网络版	90
第六章	美国《科学引文索引》	93
第一节	引文索引概述	93
第二节	印刷版《科学引文索引》	94
第三节	《科学引文索引》光盘	96
第四节	《科学引文索引》联机数据库	106
第五节	《科学引文索引》Web 方式检索	107
第七章	美国《科技会议录索引》	111
第一节	《科技会议录索引》概述	111
第二节	印刷版《科技会议录索引》	111
第三节	《科技会议录索引》光盘	113
第四节	《科技会议录索引》联机数据库	119
第八章	美国《政府报告通报及索引》	120
第一节	概 述	120
第二节	印刷版美国《政府报告通报及索引》	122
第三节	美国《政府报告通报及索引》光盘	125
第四节	美国《政府报告通报及索引》联机检索	129
第九章	专利与国际专利分类	131
第一节	专利基础知识	131
第二节	国际专利分类法	135
第十章	中国专利检索工具	139
第一节	概 述	139
第二节	印刷版中国专利检索工具	140
第三节	中国专利数据库光盘检索	144
第四节	中国专利联机数据库及 Web 网络版	148
第十一章	《世界专利索引》	153
第一节	概 述	153
第二节	印刷版《世界专利索引》	154
第三节	国际联机检索《世界专利索引》	161
参考文献		164

第一章 科技信息检索的基本知识与概念

本章主要内容

- ▶ 科技信息检索的意义
- ▶ 信息检索发展简史
- ▶ 检索语言与检索途径
- ▶ 文献的类型、识别与著录

第一节 科技信息检索的意义

目前,我们正进入一个飞速发展的信息时代。能否广泛、迅速、准确地获取所需要的信息直接影响到科研、教学工作的效率与成败。在较短的时间里获取最有价值的信息,是每一位现代科技工作者必须具有的基本能力。

从狭义上讲,在茫茫信息海洋中快速获取对自己有价值的信息主要取决于两个因素,一是了解掌握现有的信息源,即到什么地方去找自己需要的信息;二是用什么方法找,即如何获得信息。到什么地方去找和如何查找,正是本书要解决的问题。

从广义上讲,信息获取的速度,获取知识的全面性和准确性,还取决于对本专业知识的了解和掌握情况,取决于对信息的表述语言(常常是某一种外语,如英语、日语等)的掌握情况。

众所周知,现今是一个信息爆炸的时代,信息量随着时间的推移,急剧增加;随着计算机技术、电子技术的飞速发展,越来越多的信息存储于计算机软盘、硬盘或光盘上,并建立了相应的信息检索系统,即数据库检索系统,从而使信息的存储量更大、检索更方便。不同的信息开发商所收集的专业的专业领域常常有所不同,其检索系统、检索方法也不尽相同,本书的一个重点就是介绍国际著名的各数据库信息收集的范围、其检索系统的检索方法、记录显示格式等,例如,国际著名的 Dialog 数据库系统、STN 数据库系统等。

随着 Internet 的出现与普及,特别是 Web 技术的发展,越来越多的专业信息数据库通过浏览器提供检索服务,从而使检索方法更简单,用户只需要熟悉浏览器的使用方法,基本上就可以操作;同时,越来越多的信息也通过各出版商、各大公司、学校、信息发布机构等的主页提供给用户,因此,Web 界面下的信息检索是查找所需信息的重要环节。目前关于浏览器的工作原理与使用方法方面的书籍较多,本书将用较小的篇幅介绍浏览器环境下的信息检索技术,重点介绍一些网上著名的信息资源。

由于还有相当比例的信息是以纸本形式存在的,检索这些信息的工具也是纸本的,称检索工具书或检索刊,因此,本书仍以一小部分篇幅介绍常用的印刷版信息检索工具,以满足那些没有条件进行计算机检索的读者,同时,通过纸本信息的检索,也便于对一些检索概念、检索原理的引入与理解。

第二节 信息检索发展简史

最早的信息检索主要是信息分类,早在 2000 多年前,我国的汉代就有简单的图书分类法了,称作《七略》。随着社会的发展,信息量越来越大,简单的分类已不能完全解决快速查找有用信息的问题,特别是随着科技期刊的出现,定期发布相当数量的论文。随着期刊种类的不断增多,对于大多数人来说,已没有时间将所有期刊上的所有文献都阅读或浏览一遍,而且就一个读者而言,一本期刊中也不可能每篇文献都有阅读价值。因此出现了文献索引,这样读者可以根据自己的需要查找相关的文献。最早的文献索引,且至今仍有很强影响力的是 1891 年出现在美国工程学会联合会会刊(Journal of Association of Engineering Society)后的索引,该索引将 1884 年至 1891 年该会刊发表的所有论文按作者姓名的字顺排列,供读者查阅。以后这样的文献索引越来越多,并单独成册,定期收集某些期刊的文献,按文献的题名、作者、主题词等建立索引,并以检索类期刊的形式出版发行。印刷版检索刊在 20 世纪上中叶的信息检索中起到了巨大的作用。

计算机数据库的建立与计算机信息检索系统的出现几乎与计算机的发展是同步的。世界上最早的信息检索系统是由美国海军军械中心(NOS)于 1954 年在 IBM-701 计算机上研制开发出来的。大约 10 年后,美国国立医学图书馆建立了以医学信息为主的计算机文献检索系统,并逐步开放对公众服务,与此同时,美国洛克希德导弹和宇航公司下属的情报所也研制成功了 Dialog 检索系统。至今,该系统仍为世界上最著名的信息检索系统。

随着 Web 技术的出现,特别是 CGI, JAVA, API, ASP 等交互式网页制作技术的发展,Internet 网上信息检索于 20 世纪 90 年代兴盛起来,越来越多的信息出版商将他们的信息放到网上,提供浏览器环境下的信息检索与获取。同时,一批新兴的网站公司也正在迅速崛起,例如,美国的 Yahoo, Netscape, 中国的搜狐、网易、首都在线等,都通过浏览器免费提供大量的各种信息,由于其内容丰富,而且图文并茂、方便快捷,大有超过其他环境下检索系统的势头。但这些网站提供的信息一般娱乐性、生活性的比较多,科技含量较低。对于大多数科技人员来说,有科研价值的信息大多数必须来源于那些高收费的大型科技文献数据库系统,如 EI, Dialog, CA 等。由于这些系统是有偿服务的,而且其付费方式多数是将联机时间作为重要的收费标准,因此熟练掌握检索技术,节省联机检索时间就变得非常重要。

随着多媒体技术的发展,大型数据库技术的发展,已有越来越多的其他媒体,如声音、图像、动画等信息可以通过 Internet 进行传输、检索,获得原文件。多文种、多媒体信息的网上检索技术正日趋成熟与流行,对人类科技的发展与进步,人民生活质量的提高起到了极大的促进作用。

第三节 检索语言与检索途径

一、检索语言概述

文献各方面的特征可以用词、词组、短语或符号作为标识表示出来,如果将这些标识按一

定的顺序排列起来,就形成了一种标识系统。反之,从这些标识入手,就可以从系统中查找出待指的某一篇或某一类特征的文献,这种用来描述文献、组织文献记录,进行文献检索的标识系统,称为检索语言。检索语言是文献检索系统存储与检索信息所使用的共同语言,它是标引人员与检索人员之间沟通思想、取得一致理解的桥梁,所以检索语言实质上是标引和检索之间的约定语言,为避免漏检和误检,提高检索效率,使用检索工具时,必须了解这一约定语言。

检索语言的种类很多、划分方法不一,对于非图书情报专业的文献检索教学而言,通常分为表达文献外表特征和表达文献内容特征的两大类,如图 1-1 所示。

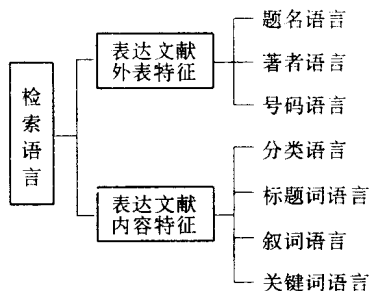


图 1-1 检索语言

在两类检索语言中,描述文献外表特征的检索语言比较直观、容易理解、使用起来比较简单、容易掌握,因此,我们重点介绍表达文献内容特征的检索语言。

二、体系分类语言及分类法

1. 体系分类语言

体系分类语言是一种直接体现知识分类的等级制概念的标识系统,是按文献内容特征进行分类的检索语言。其主要特点是按学科、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,提供从学科分类角度检索文献的途径。所谓“类”是指具有许多共同属性事物的集合。凡用来表达同一类事物的概念称为类目。每类事物,除了有共同的属性外,还有其个性。也就是说还可以用个性相近事物为标准再进行划分。例如,通信类,可以划分为通信系统、有线通信、电话、电报、传真等子类目。通信属于被划分的类,一般称其为母类或上位类(属概念),经过一次划分所形成的一系列概念称为子概念或下位类(种概念)。子类与子类之间称同位类(并列概念)。如果将子类中某一概念再进行划分,例如通信系统,又可以分成数字通信系统、扩展频谱通信系统、多址通信系统等。它们各自的属类同上,不再赘述。由此可见,体系分类法是以学科的分类为基础,概括文献的内容特征及某些外表特征,运用概念划分的方法,按知识门类的逻辑次序,从总到分、从一般到具体、从简单到复杂,进行层层划分,划分一次,产生许多个类目,逐级划分,又产生不同级别的类目,所有不同级别的类目中,层层隶属,形成一个严格有序的直线知识门类的等级制体系。

等级体系分类语言广泛用于图书、资料的分类和检索,它是图书情报界使用最普遍的一种检索语言,它的具体体现形式就是图书资料分类法。

2. 分类法

检索的基础就是分类与标引,即根据文献的某一特性,如文献题名、文献作者等信息对文献进行标引、著录,并依据这些特性,按一定逻辑规律(字顺或逻辑顺序等)对文献进行排序,即形成索引。用户根据自己已知的信息去查找索引,并根据索引提供的信息再去找原来的那篇文

献(简称原文)。计算机的数据库系统是对其可检索字段的所有信息进行分类、排序,根据用户输入的检索词,查找符合要求的信息,并将检索到的信息按一定格式显示给用户,这一系列工作均由计算机完成,因此,极大地提高了检索效率与检索速度。

由于文献作者在其论文中使用的语言,对事物的描述方法不一定规范或标准,因此,通过计算机进行的检索常常也不十分准确,不可能替代所有环节。因此,仍旧需要人们对检索语言、检索原理、检索途径等基本知识有一定的了解与掌握,不断更换检索词与检索方式以适应相应的数据库系统。

人类很早就意识到图书分类摆放与管理的重要性,早在 2000 年前,即公元前 6 年,我国汉代的刘向、刘歆就根据当时的国家藏书,完成了第一个图书分类目录,称为《七略》。以具有悠久历史文化著称的古罗马、古希腊在印本书出现前的公元 250 多年前就已经采用与手稿内容相适应的分类大纲。印刷版图书的出现使图书馆与图书分类法有了迅速的发展。比较系统、严谨的分类方法是 1545 年,由瑞士学者、被称为“西欧书目之父”的盖士纳(Conrd Gessner)编制的,称为世界书目,该分类法为以后其他众多分类法的出现奠定了基础。

目前,各国均根据自己的文献出版收藏情况,编制了相应的图书分类法,即将所有学科的图书、期刊或其他资料按其学科内容分成几大类,每一大类下分许多小类,每一小类下再分子小类,等等;最后,每一种书都可以分到某一个类目下,每一个类目都有一个类号,例如,根据目前我国应用最广的《中国图书馆图书资料分类法》,TP319 表示自动化技术、计算机技术类中的程序设计子类。完整的图书资料分类法有助于图书、期刊等在书库中的摆放及计算机数据库系统中索引的建立,使人们可以根据图书的类别或其分类号进行检索。目前,世界上有许多版本的分类法。大多数较发达国家都根据各自文献出版的情况、语言特征颁布自己的文献分类法。本书选取在世界上较有影响力的分类法,以及我国目前主要采用的分类法作一些简单的介绍。

(1) 杜威十进分类法(Dewey Decimal Classification)

杜威十进分类法简称杜威法(DDC),是由美国图书馆学家麦维尔·杜威(Melvil Dawey, 1851~1931 年)于 1873 年创建的,第一次出版发行是在 1876 年。该分类法是目前全世界使用最广泛、影响最大的图书资料分类法,已用于 130 多个国家的图书馆,并被翻译成 30 多种语言,对世界各国其他图书分类法的建立亦具有重要影响,许多分类法均是借鉴于杜威法产生的。杜威法是将所有的学科归纳成 9 大类,不能归到其中任何一类的为第 0 类,称总论,各类均用 1 到 9 的数字表示。一级大类下,以圆点相隔再分小类,称为子目。每个一级大类都再分为 10 个子类,每个子类下又分为 10 个种子类。下面为杜威法各一级类目的分类号与类目名称。关于杜威十进分类法更详细的信息参见 <http://ivory.lm.com/~mundie/DDHC/DDH.html> 或 <http://www.oclc.org/oclc/fp> 等网上站点。

- 000 总论(Generalities)
- 100 哲学(Philosophy & Psychology)
- 200 宗教(Religion)
- 300 社会科学(Social Sciences)
- 400 语言学(Language)
- 500 科学(Natural Science & Mathematics)
- 600 技术(Technology(Applied Sciences))
- 700 美术(Arts)

800 文学(Literature & Rhetoric)

900 地理与历史(Geography & History)

杜威法的各级类目均用数字表示,其中一些类目已不太适用于今天的科技发展,因此,杜威法自产生到今天一直在不断地改进与发展。例如,目前的杜威法将新兴的计算机学科放在 000 Generalities 类下的 001~006 子类中,其下又分为:

001 Knowledge

002 The book

003 Systems

004 Data processing, Computer science

005 Computer programming, programs, data

006 Special computer methods

在 000 Generalities 类下还有 010 Bibliography, 050 Magazines 等 10 个子类。

(2) 美国国会图书馆图书分类法(Library of Congress Classification)

美国国会图书馆图书分类法(<http://lcweb.loc.gov/catdir/cpsd/lcco/lcco.html>)简称国会法或 LC 分类法,是目前美国大多数图书馆采用的分类方法。由于美国的科学技术十分发达,其科学技术信息对全世界的影响也十分大,因此,了解掌握国会法对检索英文文献十分重要。

国会法是以 1891 年发表的克特展开分类法(Cutter's Expansive Classification)为基础编制的,其类目的设立与次序并没有严格地按照科学体系划分,没有细致地考虑学科之间的亲疏或从属关系,各门类的相关专家分别就各自负责的类别的实际情况逐级列类。因此,国会法一直没有一个所有类的合订本。但国会法最早是来源于一个图书馆的需要,因此,充分考虑了美国图书馆的实际藏书情况,是先有书而后有分类法,因此不会出现有类无书的情况。

目前国会法已发展成 21 大类,分别由 A 到 Z 的 21 个字母组成(除去 I, O, W, X, Y)。在组织结构上属列举式分类法,采用英文字母与数字混合标记,主要类目由 1 到 3 个英文字母组成,副类目由 1 到 4 位阿拉伯数字组成。类目标记具有弹性,可无限延展,至小数点以后,各大类均有详细索引及专用的复分表。国会法的 21 个大类的类名如下,这里 E 与 F 合为一类,均为美洲史:

A 总类;著作集

B 哲学、宗教

C 历史相关科学

D 历史;世界史

E-F 历史;美洲史

G 地理、人类学

H 社会科学

J 政治

K 法律

L 教育

M 音乐

N 美术

P 语言、文学

- Q 科学
- R 医学
- S 农业及其他
- T 工业技术
 - TC 水力工程
 - 353 水闸
 - 355 码头
 - 357 防波堤及其他
 - ⋮

- U 军事科学
- V 海军兵学
- Z 书目及图书馆科学

美国国会图书馆始建于 1800 年,最初只是为美国政府服务的地区性图书馆,目前已是世界上最有影响力、最著名的图书馆之一,其 WWW 网址是 <http://www.loc.gov> 或 <http://lcweb.loc.gov>。

(3) 人民大学图书馆图书分类法

我国近代图书分类法的发展也受美国杜威分类法的影响,在 1949 年以前便已有了根据杜威法改良的中国图书分类法。1949 年以后,社会政治、经济、文化生活发生了巨大的变化,中国出版界出版图书的内容也发生了巨大的变化,因此,图书分类法也随之修订。1952 年,中国人民大学图书馆研究编制了新的中国图书分类法初稿,该分类法称中国人民大学图书馆图书分类法,简称人大法。该分类法从 1953 年起陆续在全国各图书馆试用,后经过多次修订再版,目前,仍有许多图书馆采用此种分类方法。特别是 80 年代以前出版的图书基本都采用人大法分类。

人大法将所有图书分为 17 个类,其下再分子类。所有类目号均用阿拉伯数字表示,为展开层累制。即逐级下分展开。这里我们以 15.108.72 为例作一简单说明。类号 15.108.72 由以下五级类目组成:

- 15. 为第一级类目,表示第 15 大类:工业技术
 - 10 为第二级类目,是第 15 大类里的第 10 子类:电力工程
 - 8 为第三级类目,是第 10 子类下的第 3 子子类:机械手、机械人
 - 7 为第四级类目,是再下一级的第 7 类:计算技术、计算机
 - 2 为第五级类目,为电子数字(不连续作用)计算机

人大法的 17 个大类是:

- 1 马克思主义、列宁主义、毛泽东思想、邓小平理论
- 2 哲学
- 3 社会科学、政治
- 4 经济
- 5 军事
- 6 法律
- 7 文化、教育、科学、体育
- 8 艺术

- 9 语言、文字
- 10 文学
- 11 历史
- 12 地理
- 13 自然科学
- 14 医药卫生
- 15 工程技术
- 16 农业科学技术
- 17 综合参考

(4) 中国图书馆图书分类法

中国图书馆图书分类法简称中图法,是1971年由北京图书馆、中国科学技术情报所等单位共同编制完成的,于1974年出版,并经过多次修订与再版,目前已修订至第四版。

中图法采用五分法,即分五个基本部类,在各基本部类下再展开,共形成22个大类,除工业技术类外,其余各大类均用1个英文大写字母表示一级类目名,并以字母的顺序反映大类的顺序。根据各学科的内容需要,大类下再分二级、三级、四级、五级子类目,各子类目用阿拉伯数字表示。现以TN710.1分类号为例,说明其分类体系。TN710.1类由以下五级类目组成:

T 为第一级大类:工业技术

TN 为T大类下的子类:无线电电子学、电讯技术

7 为第三级类目:基本电子电路

10 为第四级类目:电子电路及设计

1 为第五级类目:电子管电路

中图法的类号比人大法类号更容易识别与记忆,但也存在着对当今飞速发展的电子技术考虑不足,分类不细的问题。

从1990年起中图法开始筹备第四版的修订工作,历时8年多,于1999年初逐步推向全国使用。第四版的中图法更多地考虑了随着科学技术的发展,一些大类的发展与修正。重点修订了F(经济)类、TN(无线电电子学、电信技术)子类、TP(自动化技术、计算机技术)子类。例如,在F类中增加了一些经济学的基本问题,在社会主义经济类下增加了一些新概念,如资本、剩余价值、私有制经济等;为适应通信技术的迅速发展,对TN913和TN917等子类作了较大的修改,把日新月异的通信网TN913.2及下位类移至TN915类,并扩充加细为30多个类目。第四版中图法一级类目简表及以下四级类目示例详见表1-1。

(5) 中国科学院图书馆图书分类法

中国科学院图书馆图书分类法简称科图法,是由中科院图书馆于1954年秋季编制、1958年正式出版使用的国内另一大有影响力的图书分类法。该分类法目前主要用于中国科学院图书馆及其所属各科研单位的图书馆(室)的文献分类编目。1974年2月中科院图书馆对自然科学部分进行了第二版修订;1979年对社会科学部分进行第二版修订。目前使用的是1994年12月修订的第三版。该分类法最早将所有文献分为五大部共二十五大类,所有大类号均用两位数字表示。最早的五大部是00类马克思列宁主义、毛泽东思想、10类哲学、20类社会科学、50类自然科学和90类综合性图书。以后因工程技术发展迅速,该学科类的图书剧增,因此增为一个大部,并增加了一些大类,目前科图法的六大部38大类是:

表 1-1 中图法一级类目表及以下类目录示例

基本大类	二级子类	三级子类	四级子类
A 马克思主义、列宁主义、毛泽东思想			
B 邓小平理论			
C 哲学、宗教			
D 社会科学总论			
E 政治、法律			
F 军事			
G 经济			
H 文化、科学、教育、体育			
I 语言、文字			
J 文学			
K 艺术			
L 历史、地理			
M 自然科学总论			
N 数学、物理学、化学			
O 天文学、地球科学			
P 生物科学			
Q 医药、卫生			
R 农业科学			
S 工业技术			
T 交通运输			
U 航空、航天			
V 环境科学			
X 综合性图书			
Z			
	TB 一般工业技术		
	TD 矿业工程		
	TE 石油、天然气工业		
	TF 冶金工业		
	TG 金属学与金属工艺		
	TH 机械、仪表工业		
	TJ 武器工业		
	TK 能源与动力工程		
	TL 原子能技术		
	TM 电工技术		
	TN 无线电电子学、电信技术		
	TP 自动化技术、计算机技术	TP1 自动化基础理论	
	TQ 化学工业	TP2 自动化技术及设备	
	TS 轻工业、手工业	TP3 计算技术(流控技术)	
	TU 建筑科学	TP6 射流技术	
	TV 水利工程	TP7 遥感技术	
		TP8 远动技术	
			TP30 一般性问题
			TP31 计算机软件
			TP32 一般计算机和计算机
			TP33 电子数字计算机(不连续作用电子计算机)
			TP34 电子模拟计算机(连续作用电子计算机)
			TP35 混合电子计算机
			TP36 微型计算机
			TP37 多媒体技术与多媒体计算机
			TP38 其他计算机
			TP39 计算机应用

- 00 马克思列宁主义、毛泽东思想
- 10 哲学
- 20 社会科学
 - 21 历史、历史学
 - 27 经济、经济学
 - 31 政治、社会生活
 - 34 法律、法学
 - 36 军事、军事学
 - 37 文化、科学、教育、体育
 - 41 语言、文字学
 - 42 文学
 - 48 艺术
 - 49 无神论、宗教学
- 50 自然科学(总论)
 - 51 数学
 - 52 力学
 - 53 物理学
 - 54 化学
 - 55 天文学
 - 56 地球科学(地学)
 - 58 生物科学
 - 61 医药、卫生
 - 65 农业科学
- 71 工程技术
 - 72 能源学、动力工程
 - 73 电技术、电子技术
 - 74 矿业工程
 - 75 金属工艺(物理冶金)
 - 76 冶金学
 - 77 金属工艺、金属加工
 - 78 机械工程、机器制造
 - 81 化学工业
 - 83 食品工业
 - 85 轻工业、手工业及生活供应技术
 - 86 土木建筑工程
 - 87 运输工程
- 90 综合性图书

科图法比较多地考虑了科学体系,但又强调图书分类与科学分类不完全相同,类号的分配原则是按照各类的内容及其所收图书的数量,并适当考虑科学发展的需要,因此,一级类目较多、较细且不采用十进制。各大类所占用的一级类号数也不完全相同,例如,医药、农业和工程

技术共占用了 60~89 三十个两位数类号,其中工程技术就用了 70~89 二十个两位数类号。大类后用小数点隔开,以后的数字表示更细的类目。为了使类目具有较短的号码,各细目的安排,一般不受位数的限制,也不严格要求号码代表分类的等级。为了便于记忆,小数点后的细类目尽可能容纳在九类之内。如果某类包括的范围特别广泛,需要超过九类以上的细目时,采用分配更多的类号来表示,例如,37 为“文化、科学、教育、体育”一级大类类目号,37.1 为“文化建设、文化事业”;37.2 代表“科学研究事业、科学学”;37.216 代表“情报事业、情报学”,即增加了 37.216 类号为与 37.1 和 37.2 同级的细目。

所有分类法都不是一成不变的,随着社会的发展、科技的进步而不断修正与完善。但修正的同时又必须考虑原来的框架与基础,因为分类法的修正意味着大量图书在书架上摆放位置的变化和众多计算机检索系统的更新,后台的工作量十分繁重,因此,所有的修订都应是慎重的。当然,在大多数情况下,计算机检索系统都提供其所采用的分类体系类库,用户可以调用系统已有的分类库进行检索,而不必考虑分类法的具体规定或组成。

三、主题语言

主题语言是一种描述语言,是用自然语言中的名词、名词性词组或句子描述文献所论述或研究的事物概念。这些名词或名词性词组就是主题词。

主题词的概念最早源于手工检索。因为不可能将所有文献题名、文摘中的所有词都选出来排成索引,特别是同样的含义、同样的概念,不同的作者用词可能不同,例如,自行车可以说成脚踏车或单车,再如,信息在港台等地常说成资讯,等等;另外,某些词的含义很广,用这类词标引文献,可以将许多文献都归在该类下,起不到索引的作用,例如,computer 的含义很广,可以指大型计算机、工作站,也可以指微机,而且可以与许多词搭配,用于许多不同的领域,代表不同的含义,在文献中出现的次数也会非常多,如果将其排在索引中,那么仅 computer 一词下可能就会有几万条甚至几十万条文献,没有起到索引的作用。因此,需要将那些针对性较强并代表某一方面、含义明确的词整理、规范化,并用这些词建立索引,用户通过这些词去检索自己所需要的信息。这些词就是主题词或称规范词。由这些词编辑成册即为主题词表,或称规范词表或叙词表。例如,computer 不可作为主题词,因其范围太广,一般常用 computer aided analysis, computer aided design, computer aided instruction, computer applications, computer architecture, computer control, computer graphics 等词组作为主题词进行检索,使检索范围更小,检索结果更准确。

一般主题词表是由各个学科领域最权威的专家,经过认真严格的考虑与筛选,制定出来的。但不同的信息出版商召集的专家有所不同,其侧重的方向也有所不同,因此所定义的主题词的范围与内容亦有所不同。一般大型信息出版机构都有一套自己的专家队伍,根据科技发展的情况专门于每年或每几年修订一次主题词表,如《EI 主题词表》、《科学文摘叙词表》等。我国也编有专门的《汉语主题词表》以及各专业的主题词表,如《邮电通信汉语主题词表》等。

在计算机检索中,大多数系统仍然具有主题词检索的功能,因为通过主题词检索可以缩小检索范围、提高准确度、缩减检索速度。为了便于用户检索使用,一般在系统中预先将本数据库系统所定义的全部主题词存入系统,用户可以根据系统的主题词库直接检索,不必牢记各个主题词或翻查主题词表。

很多情况下,我们并不十分了解自己要检索的文献归属于哪一个主题词下,或用一两个主题词很难代表要查找文献的全部含义,这时可以使用自由词或关键词检索,计算机系统可以对

所有的词进行检索,并对每个词在每篇文献中出现的频率进行统计,可以根据某一词出现的频率判断某篇文献与某一主题的相关程度。一般主题词的检索局限在题名、文摘、全文等字段中,而自由词或关键词的检索则不局限于字段,可以对所有字段进行自由词检索。

从数据库的实现角度来看,主题词、自由词或关键词、文献题名、文摘等是相应数据库或数据表的不同字段,各字段收录的内容完全不同,用户在选择检索范围时(如主题词检索或关键词检索),实际上是告诉计算机系统需要在哪个字段内,用用户输入的词进行检索。其检索结果与用户自己输入的检索词有关,也与数据库系统数据的录入情况有关。数据录入错误或对某一主题概念的理解有偏差,均不能得到正常、正确的检索结果。例如:很多系统不认可空格键,当希望检索“library automation”方面的文献时,不可以直接输入“library automation”,如果输入这样的词,从数据库查询语言的语法上来说并没有错误,因此系统并不报错,但系统数据库中却没有此类型的数据,或者系统对空格键有其他的规定与含义,这样虽然也可以得到一个检索结果,但可能是不正确或不准确的。

很多情况下,系统对文献题名、文献出处、主题词、作者、作者单位等字段的检索是采用前方一致的检索方式,即采用数据库查询语句(SQL)中的“select 检索词*”的方式进行,即只要用户输入的检索词与数据库系统中记录的前几位一致即为命中记录。在文摘字段,全文字段中一般多采用全文检索技术,即按检索词在该字段记录中出现的词频来判断其是否命中。但中文汉字的多重组合性,使中文检索系统的全文检索有时不能十分正确地反应用户需求,这时需要附加其他条件进行复合检索。

需要特别指出的是:不同的数据库生产商,其采用的文献分类方法可能完全不同,很多国际著名的数据库系统或检索刊物均有自己的文献分类体系,并不能彼此移用,但大多数的中文数据库系统均采用中图法进行文献分类。

第四节 文献的类型、识别与著录

一、文献的类型与识别

正确识别文献的类型对于快速查找需要的文献特别是获取原文具有十分重要的意义。可以有多种方式对文献和文献检索工具进行分类,主要有按载体形式、按加工深度、按文献的出版形式、按文种等方式。

1. 按文献的载体形态分类

可以将文献或相应的检索工具分为印刷版,例如:各种检索期刊;光盘版,如各种图书光盘、检索工具光盘等;联机版,如 Internet 网上联机数据库,大多数的网上数据库是授权使用的,即访问时需要相应的用户名与口令;缩微版,即缩微胶片,一般需要在特殊的缩微阅读机上才能阅读;磁带版,即将文献录制在计算机磁带上,再将相应的检索或阅读系统安装到计算机上,方可阅读;当然,还有一些数据库信息是存储在计算机软盘上的。

2. 按文献被加工的深度分类

可以将文献分为题录、文摘、全文。作者根据自己的科学实验或发明、发现撰写的论文称一次文献,即作者直接从科研与生产实际中总结出的科研成果。例如:大多数的期刊论文、会议论文、学位论文等均是一次文献,大多数的科技工作者所要检索的也是这类一次文献,并且希