

21世纪高等院校计算机教材系列

信息资源组织与检索

● 郝风素 李 莉 编著
曹彩英 杨亚晶

21 世纪高等院校计算机教材系列

信息资源组织与检索

郝风素 李 莉 曹彩英 杨亚晶 编著



机械工业出版社

本书介绍了信息检索的知识与各种方法,在实践的基础上强调通俗易懂的信息利用理论,加强了网络信息检索及网络数据库的使用等反映信息检索系统最新进展的内容。在编排体系上,本书针对典型的信息检索系统,将传统的印刷型工具和网络数据库结合在一起进行介绍,给读者以直观、完整、鲜明的印象,并使不同信息环境下的读者都能有所裨益。每章均配有思考题,以指导读者深入地进行学习。

本书可作为高等院校本科生和研究生信息及其相关专业的教材,也可作为广大信息用户进行信息检索的指南性读物,是广大科技人员和图书情报人员的学习参考书。

图书在版编目(CIP)数据

信息资源组织与检索/郝风素等编著. —北京:机械

工业出版社,2005.2

(21世纪高等院校计算机教材系列)

ISBN 7-111-16050-9

I. 信... II. 郝... III. ①信息管理—高等学校—教材②情报检索—高等学校—教材 IV. ①G203
②G252.7

中国版本图书馆 CIP 数据核字(2005)第 005159 号

机械工业出版社(北京市百万庄大街 22 号 邮政编码 100037)

策 划:胡毓坚

责任编辑:戴琳 李馨馨

责任印制:施红

北京金明盛印刷有限公司印刷·新华书店北京发行所发行

2005 年 2 月第 1 版·第 1 次印刷

787mm×1092mm $\frac{1}{32}$ ·12.75 印张·309 千字

0 001—5000 册

定价:19.00 元

凡购本图书,如有缺页、倒页、脱页,由本社发行部调换

本社购书热线电话:(010)68993821、88379646

68326294、68320718

封面无防伪标均为盗版

前 言

信息检索是一门关于信息资源存储、整序和查找理论、方法的学问,其研究内容包括检索理论、检索语言、数据库、信息组织方法、检索系统、检索策略、检索服务等。本书是一本介绍信息检索原理和方法及最常用、最重要的检索工具(包括数据库、搜索引擎等)的教科书。主要包括信息资源组织方式与检索的原理方法,各种印刷型、电子型检索工具的特点、作用和使用方法,网络信息资源的组织与有效利用方法,专利、标准等特种信息资源的利用。

本书每章后附有习题,书后附有可供继续阅读的参考书目,以帮助读者的深入学习。

全书内容注重:

1. 全面性。近年来一些教材,论述面较为单一,或仅有网络检索,或侧重传统印刷型文献检索,或主要讲科技文献检索,或强调社科文献检索。本书则整合各种载体、各个学科、各种检索方式、各种检索方法的知识,将人为分散的检索知识连成一个有机整体。

2. 新颖性。本书注重介绍近几年新出现的较为可靠的检索方法、检索工具和最新动态,如数字图书馆、智能搜索引擎等。同时在全书结构安排上,也力求有所创新,适当增加了思维逻辑、信息评价、信息重组和使用等内容,注重解决实际问题的能力,从而使读者在更深的层次上理解和掌握信息检索知识。

3. 准确性。本书力求讲述准确,精选和考证有关资料。如在介绍检索工具或网址时,尽可能根据最新版的第一手资料,或亲自上网检索后才下笔。

4. 实用性。本书编写宗旨在于帮助读者树立一个合理的信息来源评价体系,掌握有效获取信息的手段,提高信息意识和信息能力。所以在内容编排方面避免进入“信息查询工具”或者“数据库使用说明”的误区,在实践的基础上强调通俗易懂的信息分析利用理论。

全书共6章:其中第1章、第2章2.1、2.2、2.4、2.5节由郝风素编写,第2章2.3节、第3章3.3节、第4章4.4节、第5章、第6章6.2、6.3、6.4节由李莉编写,第4章4.3节、第6章6.1节由曹彩英编写,第3章3.1、3.2节、第4章4.1、4.2节由杨亚晶编写。统稿工作由郝风素完成。

参加本书编写其他辅助工作的还有王小凤、杨咏梅、王冰、赵殿文等。本书的顺利出版,得到河北科技大学教务处和图书馆领导的大力支持和帮助,在此表示衷心感谢。

由于时间仓促,书中难免存在不妥之处,请读者提出宝贵意见。

作 者

目 录

前言

第1章 信息资源概述	1
1.1 信息的内涵	1
1.1.1 信息的概念	1
1.1.2 信息的特征	2
1.1.3 信息与知识、情报、文献的关系	3
1.2 信息的表现形式及类别	4
1.2.1 文献信息资源	5
1.2.2 网络信息资源	10
1.3 思考题	12
第2章 信息的组织方式与检索原理	13
2.1 信息组织方式与信息检索语言	13
2.1.1 信息检索语言简介	14
2.1.2 信息检索语言类型	14
2.2 信息检索工具与数据库	19
2.2.1 数据库的类型	19
2.2.2 数据库结构	20
2.3 数字图书馆与信息资源组织	21
2.3.1 数字图书馆的功能	21
2.3.2 数字图书馆的三种主流模式	22
2.3.3 数字图书馆发展历程	23
2.3.4 数字图书馆的信息组织方式:元数据	23
2.3.5 我国主要数字图书馆项目介绍	26
2.4 信息检索类型	29
2.5 信息检索途径与方法	29
2.5.1 信息检索途径	29
2.5.2 信息检索技术与信息检索方法	31
2.6 思考题	34
第3章 信息检索系统	35
3.1 事实、数据检索系统	35
3.1.1 事实、数据检索系统简介	35
3.1.2 事实、数据检索系统的类型	35
3.1.3 检索工具举要	39
3.2 书目信息检索系统	42
3.2.1 书目信息检索系统简介	42
3.2.2 目录性检索工具	42

3.2.3 题录性检索工具	46
3.2.4 文摘性检索工具	50
3.3 全文检索系统概述	55
3.3.1 全文检索系统简介	55
3.3.2 全文数据库含义与特征	56
3.3.3 全文检索系统的检索方式	56
3.3.4 提高检索效果的措施	56
3.3.5 全文检索系统示例:《中国期刊全文数据库(CJFD)》	57
3.3.6 全文检索系统示例:电子图书系统	61
3.4 思考题	64
第4章 国际大型检索系统	65
4.1 EI	65
4.1.1 印刷版《工程索引》	65
4.1.2 网络版数据库	68
4.1.3 EI 工程信息网	73
4.2 SCI 和 ISTP	75
4.2.1 SCI	75
4.2.2 ISTP Index to Scientific & Technical Proceedings	85
4.3 CA	87
4.3.1 概况	87
4.3.2 《化学文摘》分类及简介	88
4.3.3 《化学文摘》的内容组成及著录格式	90
4.3.4 CA 的索引及使用方法	94
4.3.5 CA 的检索途径与方法	105
4.3.6 CA 数据库的网络检索	106
4.4 Dialog 国际联机检索系统	108
4.4.1 Dialog 数据库的学科范围	108
4.4.2 Dialog 的一些著名数据库列举	109
4.4.3 Dialog 系统的 Internet Web 界面的检索方法	109
4.5 思考题	119
第5章 网络信息的组织与检索	120
5.1 网络信息的组织与管理	120
5.1.1 因特网知识	120
5.1.2 网络资源的组织方式	121
5.1.3 信息过滤机制	122
5.2 网络信息资源的检索	123
5.2.1 网络信息检索的特点	123
5.2.2 网络信息检索的方法与策略	123
5.2.3 网络信息资源的评价	125
5.3 搜索引擎	126
5.3.1 搜索引擎原理	126
5.3.2 搜索引擎检索系统的主要类型与组织查询策略	130

5.3.3 几种主要的搜索引擎的组织查询技术	133
5.4 重要学术信息的网络获取	143
5.4.1 网络资源指南检索	143
5.4.2 期刊论文	143
5.4.3 学位论文	145
5.4.4 会议信息	146
5.4.5 科技报告	147
5.4.6 人物、机构、电子邮件	147
5.4.7 预印本服务器	148
5.4.8 大学资料库(电子归档)	149
5.4.9 Blog 信息	149
5.4.10 其他免费学术资源	149
5.5 因特网上的其他信息服务	151
5.5.1 FTP 文件传输协议	151
5.5.2 菜单式查询系统 Gopher	152
5.5.3 Usenet	152
5.5.4 Telnet	153
5.6 思考题	154
第6章 特种信息的检索与利用	155
6.1 专利文献及其检索	155
6.1.1 专利知识概述	155
6.1.2 专利文献	159
6.1.3 国际专利分类法	162
6.1.4 专利及其文献的编号	164
6.1.5 专利文献的检索	166
6.2 标准信息的检索与利用	171
6.2.1 标准信息的分布	172
6.2.2 标准的类型	173
6.2.3 标准信息的组织管理	174
6.2.4 标准的作用	176
6.2.5 标准信息的查找	177
6.2.6 标准的检索工具	178
6.2.7 标准认证	179
6.3 竞争情报	181
6.3.1 情报、竞争情报、竞争情报系统	181
6.3.2 竞争情报的含义	182
6.3.3 竞争情报的作用	182
6.3.4 竞争情报的研究内容	183
6.3.5 竞争情报的工作步骤	184
6.3.6 竞争情报的信息来源	184
6.3.7 竞争情报的搜集方法	185
6.3.8 竞争情报的分析方法	186

6.4 电子商务信息资源的利用	187
6.4.1 电子商务的含义	187
6.4.2 电子商务信息	188
6.4.3 电子商务信息的工作模式	189
6.4.4 电子商务信息的搜集与研究方法	190
6.5 思考题	191
参考文献	192

第1章 信息资源概述

随着科学技术的迅猛发展,人类正从工业社会步入信息社会。信息已同物质、能量并列成为现代社会的三大支柱。信息作为一种资源在社会生产和人类生活中起着愈来愈重要的作用。随着信息的爆炸性增长和人们对信息需求的与日俱增,认识信息、获取信息和分析评价信息的知识和能力已成为现代人的一项必备素质。本书编写的目的即是帮助读者掌握快速、有效获取信息的技能,提高其信息意识和能力。在此,有必要先介绍信息的内涵、特征及其表现形式和类别。

1.1 信息的内涵

1.1.1 信息的概念

对信息的认识和利用古已有之。结绳记事是早期人类存储信息的方式,烽火狼烟是古人传递信息的手段,而人们耳熟能详的名言“知己知彼、百战不殆”则诠释了古人对信息重要性的初步认识。

然而,只有到了现代社会,信息才如此深入、普遍地走入人们的生活,并且对人类社会、自然界产生了广泛深刻的影响。由日常生活到科学研究,信息无处不在。人们从来没有像今天这样重视信息、研究信息。信息的本质不断被揭示,并被引入哲学、管理学、传播学、图书情报学、信息论、系统论、控制论和通信、计算机等领域中。

信息的广泛应用,导致人们对信息的认识和定义上的差别,不同学者从不同学科角度去感知信息、解释信息、界定信息,对信息定义得出了不同的结论。哲学家认为,“信息是物质存在方式和运动状态所蕴含的间接存在物的标志,是对客观世界中各种事物的变化和特征的反映”;数学家认为信息是使概率分布发生变动的东西;经济学家认为信息是提供决策的有效数据;图书情报学家则认为,信息应是文献、资料、情报、知识、数据以及消息、新闻的总称。而信息论的创始人,美国科学家香农(C.E. Shannon)对信息的定义是“事物运动状态或存在方式的不确定性的描述”。诸种看法都在一定程度上反映了信息的本质,但囿于其各自角度的不同,而各有其局限性。

目前,哲学界和科学界较为公认的一种看法是,物质、能量和信息是物质世界的三大支柱,是科学历史上最重要的三个基本概念。世界是物质的,而信息是物质的一种普遍属性。从这种观点出发,我们可以得出广义的信息概念,即信息是事物运动的状态和方式。广义的信息也叫本体论信息,从这个意义上讲,信息是一种客观存在,与我们主观上是否感知它没有关系。需要指出的是,定义中所说的“事物”泛指一切物质客体和精神现象,“运动”则泛指一切意义的变化。

的确,宇宙万物,无时不在运动。只要有运动,就会产生各种各样事物运动的状态和方式,其运动状态和方式就给人们提供了各种各样的信息。可见,信息是普遍存在的。信息是物质

的属性,但不是物质自身。事物运动的状态和方式一旦体现出来,就可以脱离原来的事物而相对独立地载负于别的事物上,而被提取、变换、传递、存储、加工和处理。因此信息不等于它的源事物,也不等于它的载体。信息虽不等于物质本身,但也不可能脱离物质而独立存在,必须以物质为载体,以能量为动力。这三者是相辅相成,缺一不可的。

在本体论信息的基础上,如果引入认识主体的一些约束条件,就形成了狭义的信息概念,即信息是事物运动和状态方式的反映。

狭义的信息也叫认识论信息,在这个意义上,信息是认识主体对事物运动状态和方式的反映。同一个事物的运动状态和方式,对于不同的认识主体来说是不同的,有的从中获得了大量的信息,有的获得了较少的信息,有的一无所获。在认识论层次上,那些认识主体感觉不出来的,或感觉到而不能理解的,都不能称其为信息。

1.1.2 信息的特征

信息具有以下几个特征:

1. 普遍性

信息广泛存在于自然界、人类社会及思维领域中。只要有事物存在,只要有事物的运动,就会有信息存在。因此信息是普遍存在的。

2. 无限性

在整个宇宙时空中,无论是在有限的空间,或是在无限的时间长河中,事物都是无限丰富的,因而它们所产生的信息也是无限的。

3. 寄载性

信息的产生、存储以及传递,必须依附于一定的物质载体。信息自身不能独立存在和交流。同时,信息对物质的寄载性也具有相对性,表现在它可以由不同的载体和不同的方法来载荷。同种意义的信息可以负载在不同的物质载体和能量形式上,不同信息又可依附于同一物质而存在,换句话说,信息可以在各种物质和能量形式之间进行变换。信息与载体,如同形式与内容的关系一样,既不能将二者混为一谈,也不能将二者割裂开来。

4. 传递性

信息可以在时间和空间上从一点传递到另一点。信息总是处在一定的传递过程中,并且在传递中发挥价值。时间上的传递称为存储;空间中的传递称为通信。信息的传递因现代通信技术的出现,特别是互联网的大规模使用,而达到了前所未有的程度。

信息的寄载性和传递性十分重要,使得信息可以用一种形式储存起来,用另一种形式再发送出去,从而对人类知识的积累、传播和交流发挥巨大的作用。

5. 可知性与可贮性

信息是具体的,可为人们所感知、感受并利用,同时又可被收集、加工、整理、归纳、综合、记忆。信息不是虚无缥缈的东西,也不是可以随意想象和创造的事物。其真实性和具体性使得信息具有可知性与可贮性。

6. 动态性

信息是事物运动的状态和方式,事物是不断发展变化的,则信息也必然在不断的变化中。无论就信息整体而言还是具体的某一条信息而言,其内容都会随着时间的变化而不断更新。脱离了信息源的信息,因不能反映变化中的母体的新的运动状态和方式,其效用就会降低,以

致失去价值。因此,信息是具有动态性即时效性的。

7. 共享性

信息可以脱离源事物相对独立地存在并寄载于其他载体,还可以无限制地进行复制、传播或分配给众多的用户,为大家所共享。信息的这一特性与物质和能量有着本质的区别,物质和能量的利用表现为占有和消耗,当物质和能量一定时,利用者表现为明显的竞争关系,即“你多我就少”,一部分人利用多了,其他人就只得少利用或不利用。同一内容的信息可以在同一时间由两个或两个以上的使用者使用,而且并不因为信息的多人、多次重复使用而减少,反而会因信息的交流而呈现出内容的扩增。

信息除了以上的基本特征外,还具有其他一些特征,如客观性、可度量性、层次性、累积性等,这些都构成了信息的复杂性。

1.1.3 信息与知识、情报、文献的关系

信息与知识、文献、情报之间有着极其密切的关系,互有交叉重复但又彼此不同,信息的概念前面已有介绍,下面将对其他三个概念及四者之间的关系进行介绍。

1. 知识

知识是人类长期实践经验的结晶,是人类通过信息对自然界、人类社会以及思维方式与运动规律的认识,是人的大脑通过思维重新组合的系统化的信息的集合。

知识包含两方面的内容:一方面是人们认识世界的结果;另一方面是人们改造世界的方法。而认识论层次的信息即是主体所感知的事物运动的状态和方式。认识世界的结果必然是主体所感知的事物运动的状态和方式,改造世界的方法必然是主体再生以及表述出来的事物运动的状态和方式,由此可知,知识是认识论层次的信息。但是信息不一定是知识,信息具有知识的秉性,但它又比知识的内涵更广泛。

可见,信息包含了知识,知识是信息被认识的部分。知识可分为主观知识和客观知识。信息经过人脑接收、选择、处理而形成并存在于人脑中的知识称为主观知识。主观知识借助语言符号,通过各种物质载体记录下来,就变成可以传递的客观知识,即文献。

2. 文献

国际标准化组织《文献情报术语国际标准》(ISO/DIS 5217)对文献的描述是:“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 1983 年颁布的《中华人民共和国国家标准·文献著录总则》对文献的定义为:“文献是记录有知识的一切载体”。我们可以理解为,文献是将人类知识用文字、符号、图形及声频视频信号等记录方式在一定介质(如甲骨、竹帛、纸张、感光材料、磁性材料等)上所生成的物质成果。

文献由三个基本要素组成,即知识、载体和记录。知识决定文献的内容,载体决定文献的形态,记录则是构成文献的手段。

文献是传递知识的有效工具。文献是知识的一部分,知识的另一部分存在于人脑中或其他实物上。当文献中记录的知识,一旦传递到用户并为其所利用时,文献中的知识就转化为情报。当然,文献中的知识不可能全部转化为情报,转化的只是特定的一小部分,这部分知识是用户原来不知道的新知识,而且是能够解决实际问题的知识。

3. 情报

情报是一个在国际上被广泛使用但没有确切定义的概念。早期人们只将情报和军事联系在一起,认为情报是战时关于敌情的报告,20 世纪中叶,情报的概念扩展到科技领域,出现了科技情报,并因此而形成了一个新的行业,即科技情报业。随着科技情报事业的发展,科技情报概念逐渐成为一个主流的情报概念。20 世纪后期,为减少在国际交流中产生的误会,拓宽科技情报事业,国家将科技情报更名为科技信息。目前,情报的概念逐渐被信息替代。但必须指出,信息学与情报学各自有不同的学科范畴,二者不能混同。

一般来讲情报是指为一定目的搜集和传递的有特定效用的知识。知识性是情报的根本属性。但知识必须在特定时间内传递给特定的用户,并能为用户采用且有效益的才能成为情报。传递性、效用性是情报的基本属性。可见,文献中记录的知识并不都是情报。而情报也不都是文献中记录的知识,因为情报还能以语言、信号等非文献形式存在。当然,知识转化为情报,还是主要通过文献这一传递方式。所以文献和情报二者关系十分密切,而且有所交叉。

综上所述,我们可以将信息、知识、文献和情报的概念及其关系作一归纳:信息是事物运动的状态和方式及其反映;知识是人类通过信息对自然界、人类社会以及思维方式与运动规律的认识,是人的大脑通过思维重新组合的系统化的信息,是信息的一部分;情报是发挥效用,有影响的那部分信息和知识;文献是记录有知识的载体,也是信息和情报的载体。关于信息、知识、文献、情报之间的关系如图 1-1 所示。

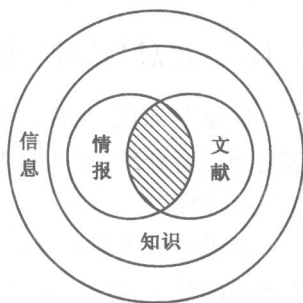


图 1-1 信息、知识、文献、情报之间的关系

1.2 信息的表现形式及类别

信息作为一种复杂的研究对象,可以按不同的标准划分出不同类型。例如:按照信息的来源分类,有自然信息、社会信息、知识信息;按照信息的性质分类,有语法信息、语义信息、语用信息;按照信息的应用分类,有工业信息、农业信息、军事信息、政治信息、科技信息、文化信息、经济信息、管理信息等。

信息作为一种资源,还可以按照开发程度分为潜在信息资源和现实信息资源两大类型。潜在信息资源是以人的大脑为载体的信息资源,现实信息资源又可分为体载信息资源、实物信息资源、文献信息资源和网络信息资源。

由于现实的信息资源是我们研究、开发和利用的主体,故对其分述如下:

- 体载信息资源:是通过人体各种表达方式,包括口头语言(交谈、聊天、授课、讨论等)、身体语言(手势、表情、姿势等)等方式来传递的信息资源,它又可分为口语信息资源和体语信息资源。
- 实物信息资源:是通过实物(如文物、产品、模型、雕塑)等来存储和表现的信息资源。
- 文献信息资源:是指以文字、图形、符号、声频、视频等方式记录在一定载体上的信息资源。它包括图书、连续出版物、小册子、特种文献、手稿、乐谱、录音资料、录像资料、计算机文档或电子资源等。文献信息资源是当前数量最大、利用率最高的信息资源。

- 网络信息资源:是以电子数据的形式将文字、图像、声音、动画等各种形式的信息存放在光、电、磁等非印刷形式的载体中,并通过网络通信、计算机或信息终端等方式再现出来的信息资源。由于网络信息资源内容形式极其丰富,且具有传递迅速、内容更新快以及交互性和开放性等特点,目前已成为人们最常使用的主要信息源之一。

由于文献信息资源和网络信息资源是本书主要介绍的信息源,下面我们就对这两种类型的信息资源分别进行介绍。

1.2.1 文献信息资源

文献是记录有知识的一切载体,文献信息资源是以文字、图形、符号、声频、视频等方式记录在各种载体上的知识和信息资源。文献积累和保存了人类认识世界和改造世界的知识和经验,并在传递和交流知识信息的过程中发挥了巨大作用。文献信息资源的类型多种多样,不同形式的文献又各自呈现出不同的功用和特点。基于文献是信息的重要载体,而我们在进行信息检索活动时往往以文献信息为主要信息源,在此着重介绍一下文献信息资源的类型划分,这对于我们学会利用各种文献信息来解决实际问题具有非常实际的意义。

1. 按信息的载体形式划分

(1) 印刷型文献

是一种具有悠久的历史,至今仍在发挥着巨大作用的、主导型的文献形式。它以纸张为存储介质,以印刷为记录手段,虽然存储密度小,占据空间大,但它便于阅读,携带方便,保存时间长。曾有人预言,20世纪末人类将进入“无纸化社会”,知识的传递将完全由计算机来完成。事实证明这只是一种不实的猜测,计算机技术的广泛应用并未减少纸张的消耗,印刷型图书的出版反而呈现出更加蓬勃的势头,加上人类几千年阅读史的浸润,预计在相当长的时间内,印刷型文献仍将是文献信息的主要载体形式。

(2) 缩微型文献

以感光材料为存储介质,以摄影为记录手段,将文献影像固化在缩微平片或缩微胶卷上形成的文献形式,其缩小倍率目前可达1/22500。由于它具有体积小、密度高、易于保存和传递但读取不便的特点,常用来保存一些具有一定保存价值的资料。如中国国家图书馆的缩微文献阅览室,主要收藏国外博士论文、美、日政府出版物及解密资料,建国前后出版的中文报纸等。从使用角度讲,随着存储密度更高、读取更方便的机读文献的发展,缩微型文献将逐渐会被机读型文献取代。

(3) 声像型文献

以感光材料和磁性材料为介质,使用声、光、磁、电等技术将声像信息记录下来的文献形式,如唱片、录音带、电影片、电视片、录像带、幻灯片、游戏卡等。其特点是直观、形象,具有独特的信息再现功能,在传递信息方面,能起到印刷型文献无法起到的作用。缺点是读取信息需要专门设备。随着计算机技术、数字压缩技术、网络技术、多媒体技术的发展,传统的声像型文献正向机读型文献转变,部分已被机读型文献融合。

(4) 电子型文献

原称机读文献,也称电子出版物。是通过编码和程序设计将信息转换成机读语言,存储到计算机外存储设备上而形成的文献。它具有存储密度高、存取速度快,数据可以修改、删除、更新并能实现自动检索等特点。目前电子型文献异军突起,发展速度极快,逐渐取代了其他介质

形式。随着互联网的普及,信息实现了远程快速传递和检索,电子型文献更是达到了无时、无处不在的状态。

2. 按信息获取难易程度划分

(1) 白色文献

信息已公开、较易获得的文献。通常指已正式出版并在社会上公开流通的文献,如图书、报纸、期刊等。这类文献向社会所有成员公开,人人皆可利用。

(2) 灰色文献

指信息内容部分已公开、部分未公开的文献。主要指非公开发行的内部文献或限制流通的文献,如内部刊物、内部技术报告、内部教材和会议资料等。这类文献出版量小,发行渠道复杂,流通范围有限、不易获得。

(3) 黑色文献

信息完全未公开,极难获得的文献。主要是指处于保密状态或不愿公布其内容的文献,如未解密的政府文件、内部档案、个人日记和私人信件等。这类文献除作者和特定人员外,一般社会成员极难获得和利用。

3. 按信息内容的加工深度划分

按照信息内容的加工深度可以划分出不同的文献级次,一般分为四种:

(1) 零次文献

指未经公开交流的最原始的文献,如私人笔记、论文手稿、书信、新闻稿、工程图纸、试验记录、会议记录、技术档案等。这是一种零星的、分散的、完全未经加工的信息资源。它具有内容新颖、来源真实但不成熟、获得较困难等特点。近年来,零次文献逐步受到了人们的重视和认识。

(2) 一次文献

是以作者本人的研究成果为依据创作而成,并在社会上公开交流或发表的文献。其加工程度很浅,亦称原始文献。如期刊论文、研究报告、预印本、会议论文、专利说明书、学位论文、技术标准等。一次文献是文献检索和利用的主要对象,一切以文献形态存储、传递的信息都取自于一次文献,因此,它也是获取可靠数据的依据。其特点是内容新颖丰富,论述详细、系统,有较高参考价值,但数量庞大,分散、不易直接查找利用。

(3) 二次文献

是对大量分散无序的一次文献进行收集、加工、整理而产生的报道性、提供查找线索的文献,如目录、题录、文摘、索引等。它以不同的深度揭示一次文献,是查找一次文献的工具。二次文献具有浓缩型、汇集性、有序性等特点。

(4) 三次文献

是利用二次文献,并在其基础上对大量一次文献进行综合、归纳和分析而形成的述评性或综述性文献,也叫参考工具书,如手册、百科全书、年鉴、辞典、综述、专题述评等。三次文献源于一次、二次文献,又高于一次、二次文献,是一种再生性文献。

在文献信息的层次结构演变中(见图 1-2),从零次文献、一次文献到二次文献、三次文献,每个环节都不断融入了著者及文献工作者的创造性劳动,使文献信息得到鉴别、提纯,不断满足人们的各种需求。文献信息经过加工、整理、浓缩,从一次文献到三次文献的变化,是文献信息由博而约、由分散到集中、由无序到有序化的过程;文献信息内容随层次的变化逐步老化,但其可检性、易检性及可获得性在不断递增;文献信息的这一层次变化,使人们获取信息变得有

章可循、有径可问。

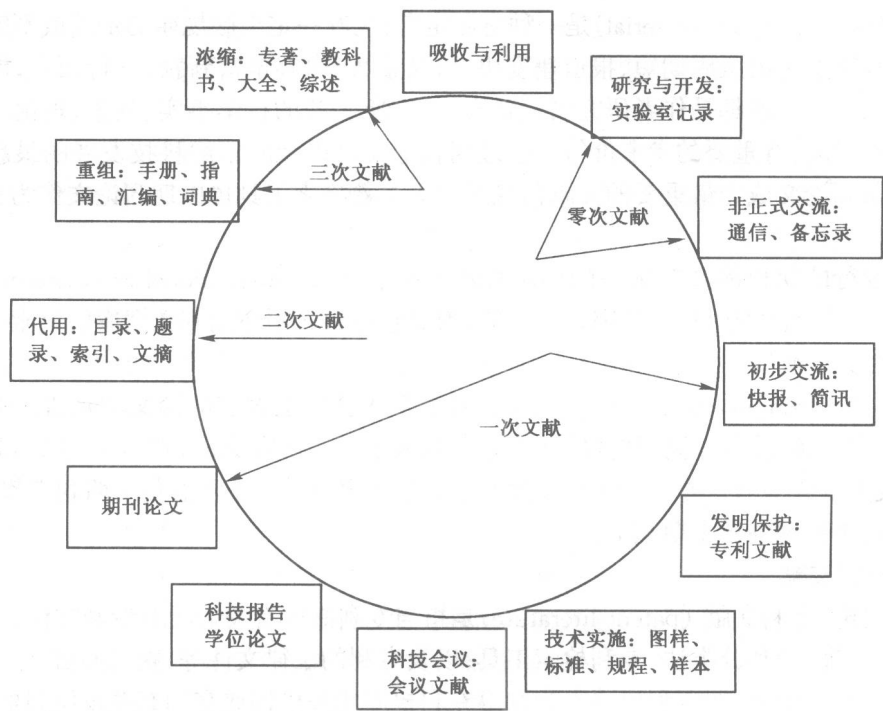


图 1-2 文献信息的层次结构

4. 按信息的出版形式划分

(1) 图书

图书(book)是正式出版且具有一定篇幅的非连续性出版物,其内容往往是作者对已有的研究成果、生产技术、实践经验或某一知识体系进行的概括和总结。按照国际标准 ISO5127 规定:图书为“48 页以上并构成一个数目单元的文獻”。一般包括有封面、封底、书名页、前言、出版说明、目次、正文、附录、索引和版权页等内容。48 页或少于 48 页的小书被称为小册子。

图书的特点是内容系统、全面、成熟,出版形式较固定,缺点是出版周期较长,报道速度较慢,新颖性差,其大部分内容非首次发表。图书是文献中最古老、最重要的类型,也是传播知识、培养人才的重要工具。图书按其内容性质和作用又可分为普及读物、教科书、丛书、专著、论文集、参考工具书等。

每一种公开出版发行的图书,都有一个惟一的国际标准书号,简称 ISBN(International Standard Book Number)。它是国际标准化组织于 1972 年公布的一项国际通用的出版物统一管理标识。ISBN 号由 10 位数字组成,共分四段:组号代表出版者的国家、地理区域、语种等,我国的组号为“7”;出版者号代表组内所属的一个具体出版者(出版社,出版公司等);书名号代表某出版社所出版的一种具体出版物的书名;校验号是 ISBN 号的最后一位数字,用以检查 ISBN 号转录过程中有无差错。

例如:《化学及相关学科信息源》一书的 ISBN 号是 7-302-06383-4。其中,7 代表中国;302 代表清华大学出版社;06383 是书名号;4 是检验码。

由于 ISBN 号是惟一的,因此,它常被用作书目数据库的检索字段。

(2) 期刊

期刊(periodical journal serial)是一种有固定名称、有一定出版规律的定期或不定期的连续出版物。其特点是出版周期短、报道速度快、信息量大、内容丰富新颖、发行面广,能及时传递信息。期刊论文一般都是作者研究的最新成果,它所刊载的科学事实、数据、理论、技术方法、构思和猜想,都具有重要的参考价值。通过期刊论文可以及时了解科技发展的最新水平和动态,因此,期刊论文成为最重要的文献信息源。大多数检索工具也以期刊论文作为报道的主要对象。

公开发行的期刊都具有统一的标准刊号:ISSN(International Standard Serials Number)。因每一种期刊的ISSN号都是统一的,因而它也常被用作期刊数据库的检索字段。

(3) 会议文献

会议文献(conference literature)是指在各种学术会议上发表的论文和报告。此类文献学术性强,内容新颖、质量较高,往往能代表某一领域最新的研究成果和发展动态,有较高的参考价值,是极为重要的信息源。会议论文通常以会议录及学会、协会主办的期刊专辑形式出版,有的分散转于相关的学术期刊上。

(4) 专利文献

广义而论,专利文献(patent literature)是指与专利制度有关的所有专利文件,包括专利说明书、专利公报、专利分类法、专利检索工具,以及专利的法律文件等;狭义而言,专利文献指各种级别的专利说明书,即专利申请人为了取得专利权向相关机构递交的有关发明目的、构成和效果的技术文件。它既是法律性文件,阐明受保护的技术内容,又是技术性文件,对该发明创造作详细说明。它具有统一编号,数量大,内容丰富新颖实用,报道迅速,是一种重要的科技信息来源。

(5) 科技报告

科技报告(sci-tech report)原是研究机构向上级主管部门或经费提供单位(甲方)呈报的关于研究工作进展、研究成果的总结报告。随着科技的发展,科技报告的数量越来越多,终于发展成为科技文献的一大类型,也是一种重要的信息源。其特点是:

- 1) 每篇报告单独成册,统一编号,由主管机构连续出版。
- 2) 内容新颖专深、详尽、成熟可靠。许多报告不仅包括有关原理、技术方案、实验数据和研究结论,还涉及到研究过程中的失败事例及原因分析。
- 3) 出版周期短、报道速度快,在一定程度上能反映一个研究机构或某一学科乃至一个国家的科研水平。
- 4) 保密性强、相当一部分科技报告不公开发行,因而不易获得。

目前,全世界每年约发行70万件科技报告,其中仅美国政府及其民间机构发表的就在50万件以上。最为著名的科技报告是美国政府的四大报告,即美国商务部系统的PB(Publishing Board)报告;美国国防系统的AD(Accessioned Documents)报告;美国宇航系统的NASA(National Aeronautics and Space Administration)报告;美国能源系统的DOE(Department of Energy)报告。

(6) 标准文献

标准文献(standard literature)包括标准及与标准相关的出版物。标准是经过公认的权威机构批准的以特定的文件形式出现的标准化工作成果,通常又称为“技术标准”。按内容区分,

技术标准是标准文献中的重要分支,其特点是对标准化对象描述详细、完整、内容可靠、实用,有法律约束力,且时效性强,适用范围明确。利用标准文献可了解有关方面的技术政策、生产水平和标准化水平,对引进、研制产品及设备,提高产品质量和生产水平,进行科学管理等有重要的参考价值。

标准文献按内容划分有基础标准、产品标准、方法标准、安全卫生标准等;按成熟程度划分有法定标准、推荐标准、试行标准;按使用范围划分有国际标准、区域标准、国家标准、行业标准、企业标准等。国际标准化机构中最重要、影响最大的是1947年成立的国际标准化组织(ISO)和1906年成立的国际电工委员会(IEC),它们制定或批准的标准具有广泛的国际影响。

标准文献都有标准号,它通常由国别(组织)代码+顺序号+年代组成,如ISO3297—1986。我国的国家标准分为强制性的国标(GB)和推荐性的国标(GB/T);行业标准代码以主管部门名称的汉语拼音声母表示;企业标准编号:Q/省、市简称+企业名代码+年份。

(7) 学位论文

学位论文(dissertation)是学位制度的产物,是指高等学校或研究机构的学生为取得某种学位,在导师的指导下撰写并提交的学术论文,包括学士论文、硕士论文、博士论文。

学位论文一般为非正式出版物,通常只在学位授予单位的图书馆及授予国的国家图书馆或信息中心保存有副本,因此学位论文的收集与利用不如其他类型的文献方便。例如,我国的国家图书馆和中国科技信息研究所文献馆分别收藏博士、硕士论文。在美国,由国际大学缩微品公司负责收藏和报道全美的博士、硕士论文,并收集报道其他国家的学位论文。在英国,由英国国家图书馆(不列颠图书馆)负责收藏和报道本国的学位论文。在日本,国立大学的学位论文由日本国立图书馆统一收藏,私立大学的学位论文则收藏在学位授予单位的图书馆。

学位论文的选题一般来自导师所承接的研究课题,通常为某一方向的子课题。因机构、导师、研究课题性质及攻读学位学生素质和能力等多方面因素影响,学位论文的学术水平差别很大。一般来说,博士论文具有一定独创性,所涉及的内容比较专深,对科研、教学有较大参考价值。

(8) 政府出版物

政府出版物(government publication)又称官方出版物,是指各国政府部门及其所属机构出版的文献。它可分为两大类。

1) 行政性文献:包括政府法令、方针政策、规章制度、决议、指示、统计资料等,主要涉及政治、法律、经济等方面。

2) 科技文献:主要是政府部门的研究报告、标准、专利文献、科技政策文件、公开后的科技档案等,有些研究报告在未列入政府出版物之前已经出版过,故它与其他类型的文献有重复。

政府出版物对了解各国的方针政策、经济状况及科技水平,有较高的参考价值,一般不公开出售。一些西方国家对政府出版物非常重视,较大的公共图书馆、研究型图书馆均设立政府出版物的单独借阅部门。我国图书馆通常将政府出版物按各自的出版类型分散收藏于不同的部门。

(9) 产品资料

产品资料(product literature)是厂商为推销产品而印发的介绍产品情况的文献,包括产品样本、产品说明书、产品目录、厂商介绍等,它是对定型产品的性能、构造原理、规格、操作规程等的具体说明。产品资料形象直观、图文并茂,常附较多的外观照片和结构简图,因此它是技