

高等职业院校信息检索教程

Gaodeng Zhiye Yuanxiao Xinxì Jiansuo Jiaocheng

信息检索 与实践

Information Retrieval and Practice

贺霞 凌征强 司马敬敏 主编

- 整合最新应用
- 重视实践教学
- 强化信息分析



人民邮电出版社
POSTS & TELECOM PRESS

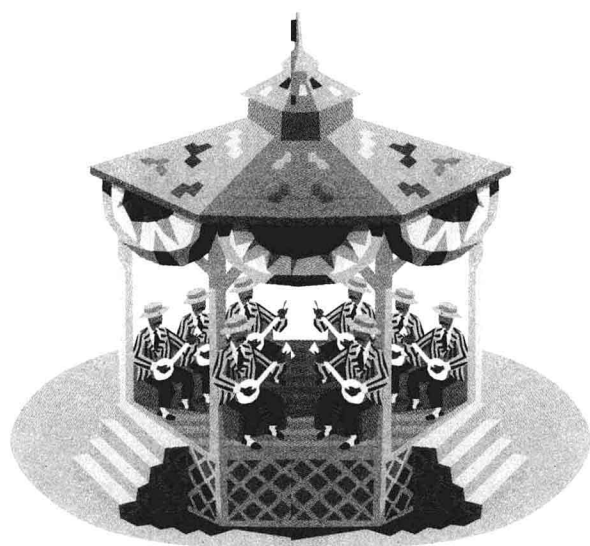
高等职业院校信息检索教程

Gaodeng Zhiye Yuanxiao Xinxi Jiansuo Jiaocheng

信息检索 与实践

Information Retrieval and Practice

贺霞 凌征强 司马敬敏 主编



人民邮电出版社

北京

图书在版编目(CIP)数据

信息检索与实践 / 贺霞, 凌征强, 司马敬敏主编

— 北京: 人民邮电出版社, 2010.3 (2011.7 重印)

ISBN 978-7-115-22309-8

I. ①信… II. ①贺… ②凌… ③司… III. ①情报检索 IV. ①G252.7

中国版本图书馆CIP数据核字(2010)第023266号

内 容 提 要

“信息检索”课程是一门公共专业基础技能课,从1992年国家教委颁布《高等学校文献检索课教学基本要求》开始,各高等院校广泛开设“信息检索”课程。本书针对高等职业院校“信息检索”课程的需求而编写,主要内容包括信息检索基础、计算机检索与图书馆有效使用、网络数据库检索、Internet 信息检索、特种文献检索以及信息检索综合利用。与现有其他教材相比,本书增加了现代图书馆如何有效利用及注重信息分析能力培养的内容,并加大了实践教学内容的比重,在每一章后面都附上习题,供学生课后思考、实践。

本书理论体系完整,实践指导性强,适合作为高等职业院校信息检索类课程的教材,也可作为学生信息检索实践的参考书。

信息检索与实践

- ◆ 主 编 贺 霞 凌征强 司马敬敏
责任编辑 贾 楠
- ◆ 人民邮电出版社出版发行 北京市崇文区夕照寺街14号
邮编 100061 电子邮件 315@ptpress.com.cn
网址 <http://www.ptpress.com.cn>
北京鑫正大印刷有限公司印刷
- ◆ 开本: 787×1092 1/16
印张: 16.25 2010年3月第1版
字数: 427 2011年7月北京第3次印刷

ISBN 978-7-115-22309-8

定价: 32.00 元

读者服务热线: (010)67170985 印装质量热线: (010)67129223

反盗版热线: (010)67171154

高等职业院校信息检索教程

信息检索与实践

编审委员会

主 编：贺 霞 凌征强 司马敬敏

副主编：井春兰 金 声 夏年军 赵俊颜

郭育凯 康思本 张慧湘

编 委：江中君 吕 刚 胡其兵

陈建文 杜 强 杨茂昌

吴蜀红 卫丽君 黄 春

叶洁娴

信息检索课程是一门公共专业基础技能课,从1992年国家教委颁布《高等学校文献检索课教学基本要求》至今,在全国范围内开设到现在只有十余年的历史,因此该课程还是一门年轻的课程,其教材的研究和编写始终处于不断改进与探索中,当前已取得了一定成就,也出现不少优秀教材。但琳琅满目的教材中也存在一些问题,如教材内容体系不合理、理论基础知识内容与实践内容缺乏逻辑联系、手工检索部分内容所占比例偏大、教材内容陈旧落伍、更新换代缓慢、不能及时反映形势等。

对高职院校来说,目前大多数检索课教材并不适应其课程的需要,即使优秀的教材也是如此,理由有三:(1)高职院校是近年才得到大力发展,而以前的教材大多是针对本科教学计划而编写的,并不适合高等职业院校;(2)现有教材通常都是以文献作为对象编写的,而无论是过去、现在还是未来,用户检索的最终目标都是可用的信息,而很少是整篇的文献,特别是对高职院校的学生,他们不像本科院校学生那样关注文献的系统性,他们更关注的是信息和数据本身,而不考虑在何种文献上获取,因此,教材的编写以“信息检索”为对象更能体现其学科特点,也更符合客观需要,对高职院校的学生也更有吸引力;(3)高职院校的课程教学中特别重视实践教学环节,但现有教材理论部分比重偏大,实践教学部分内容明显不足,因此,在教材中充分体现实践教学,更适用于高职院校的学生。

我们在编写本教材的过程中,相比现行教材力求实现如下几方面的改革。(1)增加现代图书馆如何有效利用的内容。在现行检索课教材中,很少能见到现代图书馆利用或与之相关的内容。据调查,现在的大学生、研究生,甚至许多高校教师仍然不能够充分利用图书馆,不能有效地从传统或虚拟的图书馆获取自己所需要的文献信息。对于高职院校的学生来说,在校就读的两至三年的时间里,最重要的文献资源就是图书馆现存的、可以直接利用的纸质文献和电子资源,因此,在教材中增加现代图书馆利用的内容显得非常必要,这也是对高职院校学生来说是最为实用的内容之一。(2)强化信息分析能力的培养。现代社会,信息获取并不困难,困难的是信息分析。从海量的信息中提取出的信息,如何判断其是否有效和怎么鉴别垃圾信息是现代信息使用者最大的困惑。现行教材中往往也强调信息分析的重要性,但是缺少培养信息分析的方法与内容。高职院校的学生年龄普遍偏小,而且学习时间较短,学生很多来自对口招生的中等职业技术学校,这些学生的基础教育相对于普高学生来说较为薄弱,相比本科院校学生来说信息利用方面的意识也较弱。高职院校的学生普遍存在希望能够在学习过程中快速找到自己需要的、与所学专业相符的信息资源而利用这些信息资源的能力却相对薄弱的特点,更加缺乏对所获信息的分析能力,所以在教材中强化信息分析能力的培养非常必要。(3)改革教材内容体系,加大实践教学内容的比重。信息检索课是理论性与实践性相互依存、相互促进的一门课程,如果只停留在对基础理论、基本知识的了解和掌握上,既达不到深化知识、

学以致用的目的，也无法起到培养学生创新精神与检索能力的目的。因此，加强实际检索能力的培养，是提高信息检索课教学质量的重要环节。高职院校非常重视学生的实践教学环节，信息检索课程的教学也应如此，本教材也有意识地在编写过程中加强实践教学内容，以提高学生实际动手能力；同时在每一章后面都附上习题，供学生课后思考和实践。

本教材共 7 章，由贺霞主持编写，同时指导完成第 3 章、第 6 章和第 7 章的编写工作，并担任教材的终审；凌征强负责全书的统稿和校对工作，指导完成第 1 章、第 2 章和第 5 章的编写工作，参与了第 4 章的编写。第 1 章由金声编写；第 2 章由井春兰编写；第 3 章由郭育凯编写；第 4 章由赵俊颜编写；第 5 章由夏年军编写；第 6 章和第 7 章由司马敬敏编写；另外还有不少同志参与了教材编写论证和资料搜集工作。

编者在本书编写过程中参阅了大量相关文献，也得到了广东高校图工委高职分委会和广东省各高职院校图书馆馆长的大力支持和无私帮助；教材的出版得到了广东世纪光华书业有限公司的资助，在此，我们一并致以衷心的感谢。

限于编者水平，疏漏或不当之处在所难免，敬请广大读者提出宝贵意见。

编 者

2009 年 12 月

目 录

第1章 信息检索基础.....1

- 1.1 信息、知识与文献.....1
 - 1.1.1 信息化与信息素养.....1
 - 1.1.2 信息的定义及其属性.....2
 - 1.1.3 知识、情报、文献及其与信息的关系.....3
- 1.2 文献信息基础知识.....5
 - 1.2.1 信息资源与文献信息.....5
 - 1.2.2 文献的属性与功用.....6
 - 1.2.3 现代文献的特征.....7
 - 1.2.4 文献的类型.....9
- 1.3 信息检索概述.....10
 - 1.3.1 信息检索的概念及意义.....11
 - 1.3.2 信息检索的类型.....11
 - 1.3.3 信息检索的方法和途径.....12
- 习题.....14

第2章 图书馆利用.....15

- 2.1 图书馆概述.....15
 - 2.1.1 图书馆发展概况.....15
 - 2.1.2 图书馆类型.....19
- 2.2 图书馆文献分类.....25
- 2.3 工具书概述.....36
 - 2.3.1 工具书的概念与特征.....36
 - 2.3.2 工具书的类型.....37
 - 2.3.3 工具书的结构.....37
 - 2.3.4 工具书的排检.....38
 - 2.3.5 检索工具书.....41
 - 2.3.6 参考工具书.....47
- 习题.....55

第3章 计算机检索基础.....57

- 3.1 计算机检索概述.....57
 - 3.1.1 计算机检索的概念和原理.....57
 - 3.1.2 计算机信息检索的系统构成.....58

- 3.1.3 计算机信息检索的发展.....59
- 3.2 计算机信息检索的基本方法与策略.....62
 - 3.2.1 计算机检索的基本方法.....62
 - 3.2.2 计算机检索策略.....63
- 3.3 搜索引擎基础知识.....65
 - 3.3.1 搜索引擎的定义和工作原理.....66
 - 3.3.2 搜索引擎的结构和工作流程.....66
 - 3.3.3 搜索引擎的基本类型.....68
 - 3.3.4 搜索引擎的检索技术.....69
- 3.4 常见搜索引擎.....70
 - 3.4.1 Google.....70
 - 3.4.2 百度.....84
 - 3.4.3 北大天网.....91
 - 3.4.4 中国搜索.....93
 - 3.4.5 其他搜索引擎.....94
 - 3.4.6 搜索引擎的使用策略.....94
- 习题.....95

第4章 常用网络数据库检索.....97

- 4.1 网络数据库概述.....97
- 4.2 常用中文期刊数据库.....98
 - 4.2.1 中文期刊数据库概论.....98
 - 4.2.2 CNKI 系列数据库.....100
 - 4.2.3 维普中文科技期刊数据库.....109
 - 4.2.4 万方知识服务平台.....114
- 4.3 常用中文图书数据库.....120
 - 4.3.1 中文图书数据库概述.....120
 - 4.3.2 超星数字图书馆.....121
 - 4.3.3 书生之家数字图书馆.....124
 - 4.3.4 方正 Apabi 电子图书.....125
- 4.4 其他信息资源数据库.....127
 - 4.4.1 常用外文数据库.....127
 - 4.4.2 国内常用的其他文献数据库.....131
- 习题.....132

第5章 Internet 信息资源检索.....134

5.1 Internet 信息资源概述.....134	6.4.1 《中国学位论文通报》.....197
5.1.1 Internet 的发展历程.....134	6.4.2 《中国科学院博士学位 论文文摘》.....197
5.1.2 我国 Internet 的发展.....136	6.4.3 《国际学位论文文摘》.....197
5.1.3 Internet 信息资源的类型.....137	6.4.4 《学位论文综合索引》.....198
5.1.4 Internet 信息资源的特点.....140	6.4.5 《美国博士学位论文》.....198
5.1.5 Internet 信息资源的组织.....141	6.4.6 万方学位论文数据库.....198
5.1.6 Internet 信息资源的评价.....144	6.4.7 中国知网的硕博学位论文 数据库系统.....201
5.2 Internet 信息资源检索的特点和 方法.....145	6.4.8 《CALIS 高校学位论文全文 数据库》.....205
5.2.1 Internet 信息资源检索的特点.....145	6.4.9 高校学位论文数据库.....205
5.2.2 Internet 信息资源检索方法.....146	习题.....206
5.2.3 影响 Internet 信息检索的因素.....147	
5.3 门户网站.....148	第 7 章 信息检索的综合利用207
5.3.1 主要中文综合门户网站简介.....149	7.1 信息的综合利用.....207
5.3.2 中文垂直门户网站简介.....151	7.1.1 信息的累积.....207
5.4 Web 2.0 环境下的信息资源检索.....152	7.1.2 信息检索方法.....210
5.4.1 Web 2.0 概述.....152	7.1.3 综述的写作.....211
5.4.2 Web 2.0 环境下信息资源的特点.....155	7.1.4 学术论文的写作.....213
5.4.3 Web 2.0 环境下的信息检索 对策.....156	7.1.5 文摘的写作.....215
习题.....159	7.2 参考咨询服务.....216
第 6 章 特种文献160	7.2.1 参考咨询服务概述.....216
6.1 专利.....160	7.2.2 国内数字参考咨询台.....219
6.1.1 专利与专利文献.....160	7.2.3 国外数字参考咨询台.....231
6.1.2 专利的类型.....162	7.3 知识产权.....237
6.1.3 专利文献的组成.....162	7.3.1 知识产权与著作权.....238
6.1.4 专利的分类.....163	7.3.2 著作权保护范围.....238
6.1.5 专利检索语言.....163	7.3.3 著作权的内容.....239
6.1.6 专利检索途径.....166	7.3.4 著作权的保护期限.....240
6.1.7 专利检索工具.....167	7.3.5 著作权的合理利用.....241
6.2 标准.....178	7.3.6 著作权的法定许可利用.....242
6.2.1 标准的定义与分类.....178	7.3.7 强制许可制度.....243
6.2.2 标准文献.....179	7.3.8 信息网络传播权.....243
6.2.3 标准的检索工具.....182	7.3.9 计算机软件著作权保护.....244
6.3 科技报告.....191	习题.....245
6.3.1 科技报告概述.....191	附录 1 国内行业标准组织代号表247
6.3.2 美国四大科技报告与检索.....192	附录 2 省、区辖市行政代码表249
6.3.3 中文科技报告检索.....196	参考文献250
6.4 学位论文.....196	

第 1 章

信息检索基础

本章介绍信息检索基础知识，包括信息、知识与文献的基本概念与属性，文献信息基础知识以及信息检索概述，为学习后续章节的内容奠定基础。

1.1 信息、知识与文献

本节介绍信息、知识与文献的基本概念以及其相互之间的关系。

1.1.1 信息化与信息素养

在人类社会发展的历史长河中，信息活动从来没有间断过，信息一直在人类的生产、生活中扮演着重要角色，然而，信息技术及信息产业的发展、人类对于信息的需求和依赖、信息对人类社会生活的影响远远没有达到今天的程度。1946 年，世界上第一台通用电子计算机 ENIAC 在美国诞生。1957 年，前苏联成功发射了人类历史上第一颗人造地球卫星。计算机与通信技术的结合，催生了社会信息化时代的到来。1963 年，日本学者梅棹忠夫在《信息产业论》一文中第一次提到“信息化”这个概念。1967 年，日本政府的一个科学、技术、经济研究小组在研究经济发展问题时，比照“工业化”，正式提出“信息化”，并将其界定为：向信息产业高度发达且在产业结构中占优势地位的社会——信息社会前进的动态过程，它反映了由可触摸的物质产品起主导作用向难以触摸的信息产品起主导作用的根本性转变。

20 世纪 80 年代以来，由于第三次信息技术革命的推动和知识经济的蓬勃兴起，人类社会逐渐转变为信息社会——一个生活质量、社会变化和经济活动越来越多地依赖于信息及其开发利用的社会。在信息社会时代，国家通过信息技术开发、信息产业的发展、信息人才的配置，最大限度地使用信息资源以满足社会的信息需求，提升人的素质、提高工作效率和生活品质，实现全人类的共同进步和发展。随着全球经济一体化进程的进一步加快，信息化水平已经成为衡量一个国家或地区的现代化程度、经济增长能力、国际竞争力和综合实力的重要标志。

信息社会，或者称之为社会的信息化，具有两个突出的特征。一是信息数量呈几何级数增长。据专家估计，20 世纪 40 年代以来产生和积累的信息量已经大大超过了在此之前人类有史以来的所有信息量之和。19 世纪以来人类知识信息量每 50 年增长 1 倍，20 世纪中叶每 10 年增长 1 倍，而 20 世纪 70 年代以后每 5 年就增长 1 倍。二是人们对信息的需求和依赖日益增强。在信息技术广泛应用于社会生活各个领域的大趋势下，我们随时随地都在自觉或不自觉地接受、传递、储存和利用来自各方面的信息，人们越来越强烈地意识到信息在工作和生活中的重要性。除此之外，在社会的信息化程度越来越高的进程中，诸如信息超载、信息泛滥、信息失衡、信息污染、信息

障碍等全球性信息环境问题也变得愈来愈突出,干扰着人们对信息的开发利用。因此,具备一定的信息素养,掌握一定的信息技术,以便快捷地从浩如烟海又纷繁复杂的信息来源中搜寻并利用有价值的信息,正在成为各行业从业人员的迫切需求。

信息素养是指个人通过生产、生活实践以及接受信息教育而形成的获取、利用信息的意识和综合能力。信息素养是现代文化素养的一部分,是人们终身学习和知识创新的基础技能之一。信息素养由信息意识、信息能力和信息道德3个方面构成,其具体表现为对信息源内容的了解程度,通过信息解决问题的基本意愿,信息获取方法的掌握程度,知道在何时、通过何种信息源、如何解决相关问题,具备良好的信息道德意识,善于合理、合法地利用信息的意识等。信息素养的核心是获取和加工信息的能力。

在信息化的社会里,人的信息素养的高低,主要是信息素养教育的结果。信息素养教育的根本目的就是培养人的信息意识和信息能力,其包括以下内容。

- (1) 敏锐的信息意识。
- (2) 快速获取信息的能力。
- (3) 正确评价和鉴别信息的能力。
- (4) 善于将信息技术和获取的信息用于学习、工作和生活的能力。
- (5) 利用所获得的信息开发信息的能力。

其中,快捷的信息获取能力主要是通过掌握信息检索的知识和技能来实现的,因此,学习信息检索的知识和技能就成为信息素养教育的重要内容之一。

1.1.2 信息的定义及其属性

1. 信息的定义

“信息”一词,英语为 information,有知识、资料、数据、情报、消息等多种含义。汉语中的“信息”一词最早见于唐代诗人李中《暮春怀故人》中的诗句“梦断美人沈信息,目穿长路倚楼台”,显然,这里的“信息”有音信,消息的意思。然而,这远非是现代意义的信息概念。半个世纪以来,人们对信息的利用和研究越来越广泛、深入,对信息的认识理解也就越来越多样化。不同学科、不同时期、从不同角度对信息有不同解释。其中,最有代表性是美国人香农和维纳的理论。

信息论的创始人香农(C.E.Shannon)在《通信的数学模型》中提出了“狭义信息论”,将信息定义为“用于减少随机不确定性的东西”,也是说信宿(信息接收者)未接到某种消息前不知道信源(信息发出者)发出什么消息,只有在收到消息后才能知道消息的内容,用于消除信源的不确定性的东西就是信息。

控制论的创始人维纳(Norbert Wiener)认为:“信息是人们在适应外部世界并且使这种适应反作用于外部世界的过程中同外部世界进行交换的内容的名称。”同时又进一步指出:“信息这个名称的内容就是我们对外界进行调节,并使我们的调节为外界所了解时与外界所交换来的东西。”还说:“信息既不是物质,又不是能量,信息就是信息。”

我国《辞海(第六版 彩图本)》(2009年10月出版)中关于信息的解释是:“①音讯;消息。②通信系统传输和处理的对象,泛指消息和信号的具体内容和意义。信息是指对消息接受者来说预先不知道的报导”。^①

《新华词典(大字本)》(2001年修订版)对信息的表述是:“①音信;消息。②信息论中指用符号传送的报道,报道的内容是接受符号者预先不知道的。③事物的运动状态和关于事物运动状态的陈述”。^②可

^① 《辞海(第六版 彩图本)》(2009年10月出版)第2556页。

^② 《新华词典(大字本)》(2001年修订版)第1097页。

见,对于究竟什么是信息,目前学术界并没有一个统一的、能为各界普遍认同的定义。相对得到大家认可的观点是:信息是指应用文字、数据或信号等形式通过一定的处理和传递,来表现各种相互联系的客观事物在运动中所具有的特征的总称。不同事物有不同的存在方式和运动规律,从而构成了各种事物的不同特征,发出不同的信息。总之,信息是对客观世界中各种事物的变化和特征的反映,是客观事物之间相互作用和联系的表征,是客观事物经过感知或认识后的再现。通俗地讲,信息是通过信号带来的消息。

2. 信息的属性

信息广泛存在于自然界、人类社会及人类的思维领域,可以说无处不在,无时不有。信息具有以下特征。

(1) 客观真实性。信息不是事物和事件本身,但信息产生于物质,是事物存在形式和运动状态的反映,因此无论你是否感觉到,信息同客观事物一样也是客观存在的。

(2) 可存储性。既然信息是事物存在形式和运动状态的反映,就必然依附一定的事物而存在。信息不能孤立存在,而是存储于所依附的各种物质载体之上。

(3) 可传输性。信息作为事物的一种属性或者某种特性的表现,可以借助某种物质或方式在事物间传递和输送,譬如,借助光、电、声、文字、符号等进行传送。

(4) 可转换性。信息可以被加工、整理、细分、综合等使之浓缩或扩充并转换各种形态。例如,将文字信息转换成图形信息,将声音信息转换成文字信息,将光信息转换成电信息等。

(5) 共享性。一定的物质和能量被专人占有使用时往往不能同时被其他人占有和使用,而且物质、能量被使用后,会因消耗而减少甚至不复存在。然而信息则不同,信息可以同时被多人占有和使用,并且在使用后,不仅不会因此消耗减少,还可能会增加新的信息。

1.1.3 知识、情报、文献及其与信息的关系

1. 知识

从本质上说,知识属于认识的范畴,有着丰富的内涵和广阔的外延。《辞海(第六版 彩图本)》(2009年10月出版)对知识的解释是:“人类认识的成果或结晶。依反映对象的深刻性,可分为生活常识和科学知识;依反映层次的系统性,可分为经验知识和理论知识。经验知识是知识的初级形态,系统的科学理论是知识的高级形态。按具体的来源,知识虽可区分为直接知识和间接知识,但是从总体上说,人的一切知识(才能也属于知识范畴)都是后天在社会实践中形成的,是对现实的能动反映。社会实践是一切知识的基础和检验知识的标准。知识(精神性的东西)借助于一定的语言形式,或物化为某种劳动产品的形式,可以交流和传递给下一代,成为人类共同的精神财富。”^①知识与信息的含义密切相关。人类为了进行知识的传递和交流,必须使知识具有能为人类感觉器官所感知的形式,如文字、语言、符号、代码、电磁波、图像、实物等,这种能为感官所感知的形式就是信息。知识的形成与传播离不开信息。在生产、生活、科研等活动中,人脑通过对相关事物发出的信息的接收、选择、甄别,得到对事物一般特征的认识,形成概念,再通过对相关概念的分析、判断、综合、推理,得到对事物本质的认识,这些认识被人脑系统化,就构成知识。因此,在同信息关系问题上,可以认为“知识是人类通过信息对自然界、人类社会以及思维方式运动规律的认识和掌握的结果,是人的大脑通过思维重新组合的、系统化的信息集合”。人类不仅要通过信息感知世界、认识世界和改造世界,还要根据所获得的信息组成新的知识。

^① 《辞海(第六版 彩图本)》(2009年10月出版)第2934页。

知识的特性主要归纳为 3 个方面：知识系统性、知识复杂性和知识内隐性。知识系统性与独立性相对应，知识的系统性体现了被转移知识与背景知识的关联性。独立的知识只需描述该领域知识本身即可明白，而系统的知识需通过其他相关领域知识辅助才能理解。也就是说当某项知识所牵涉的领域愈多，则此项知识的系统性愈高。知识复杂性指在知识的学习过程中，是否可以借助文字、图表、语言等载体传递知识本体。较简单的知识只需以少许信息即可将其意义描述清楚，而较复杂的知识则必须通过大量的信息才能将其意义解释清楚。知识内隐性或外显性反映知识的显性化或隐性化程度，是事物的一体两系，二者相互对立、此消彼长，即知识的内隐性程度越高，则外显性程度越低。隐性知识是个人、团队、组织长期积累的经验，难以用文字和语言完整地表达出来，其有效的传播方式通常是在实践中摸索总结或师徒传授。而显性知识则——可以通过文字或文件加以陈述或编撰，因此显性知识比较容易通过文件或数据的形式分享。

依据不同的分类标准，可将知识分为不同的类型，国际经济合作与发展组织在《以知识为基础的经济》的报告中将知识分为四大类：第一类，“知道是什么（Know-what）”：即关于事实方面的知识；第二类，“知道为什么（Know-why）”：即关于自然原理和规律方面的知识；第三类，“知道怎样做（Know-how）”：是关于做某种事情的技艺和能力方面的知识；第四类，“知道谁有知识（Know-who）”：是关于到哪里去寻求知识的信息，侧重于创造思想、方法、手段、过程、特点等的了解。

2. 情报

情报同信息、知识的含义相近。最初，情报是作为军事用语出现的，是“关于敌情的报告”。用于军事领域。如今，情报概念的应用领域越来越广。除了军事领域外，政治、经济、科技以至教育等领域都有情报的应用。目前，虽然学术界对于什么是情报尚未有公认的定义，但一般认为：情报是关于特定时间、特定对象有利用价值的特定信息或知识。情报来源于信息和知识，是“激活了的、活化的知识”。情报是信息或知识的一部分，但不是所有的信息和知识都是情报。同信息和知识相比，情报有如下 3 个突出的特征。第一，情报具有很强的目的性、特定性，也就是说，只有对特定对象有利用价值的信息或知识才能成为情报。如果收集到的信息或知识不能给特定对象带来实际的效用，这个信息或知识就不能称之为情报，只是信息或知识。第二，情报还有很强的时效性。除了对特定对象有特定用途外，针对特定对象的需求，反映特定时间的信息或知识才是情报。在特定的时限内，某项信息是有价值的情报，但超过了这个时限，信息就失去了曾拥有的价值，也就是不再是情报了。第三，情报是处于传递状态的信息或知识，处于静止状态的信息、知识不能算作情报。文献中蕴藏的大量对特定对象有特定用途的知识信息，但这些知识信息若无人查阅，未经传递，就只能是知识信息，而不是情报。

3. 文献

与信息密切相关的另一个概念是文献。“文献”一词古今含义不同。在古代，“文献”一词专指有历史价值的图书资料。如今，“文献”的外延和内涵都已发生变化，以现在的应用看来，文献是“记录知识的一切载体”，即凡是储存了信息和知识的各种载体都称之为文献。需要强调的是文献是以文字、图形、符号、音频、视频等技术手段记录人类知识的物质载体。文献有 3 个基本要素：一是文献含有的知识信息；二是负载知识信息的物质载体；三是记录手段。由于文献的物质载体和记录手段不同出现了各种不同的文献形态。自商周以来，我国文献的形态几经变迁。就文献载体而言，先后出现甲骨、青铜、竹木、丝帛为载体，又转为纸张。就文献的记录手段而言，先后经历了刻、铸、书写到印刷几种方式变化。同是以纸张为载体，以印刷为记录手段的文献又可根据出版形式的不同分为常规文献（如图书、期刊、报纸等）和特种文献（如科技报告、专利、标准等）。随着社会发展和文献生产技术的日新月异，文献形态发生了巨大的变化，除了传统的纸质文献呈现多样化发展趋势外，非纸质的文献，譬如声像文献、缩微文献和软磁盘、光盘等机读文献相继问世，开创了文献发展史上的新纪元。

4. 信息、知识、情报、文献之间的关系

将信息、知识、情报、文献之间的关系描述如下。信息是事物发出的、体现事物存在和运动状态的信号和消息。信息经过人脑的加工使之系统化而成为知识。知识与信息之间是一种互动关系,知识创造需要信息,而相关信息的开发也需要知识的应用。反映特定时间对特定对象有利用价值的特定信息或知识构成情报。知识和情报来自信息,是信息的一部分,是经过人脑加工后系统化的信息;而信息是知识与情报的来源,是组成知识与情报的原材料。信息、知识、情报记录在一定的载体上就形成文献。文献承载着知识、情报和信息;知识、情报、信息通过文献载体得以记录、存储和传播。但文献上所记录的知识、信息不全是情报。信息、知识、情报也不全是以文献形式记录。所以说,知识、情报、文献同信息之间虽有十分密切的联系,但又有明显的区别。

信息、知识、情报以及文献之间的逻辑关系如图 1-1 所示。

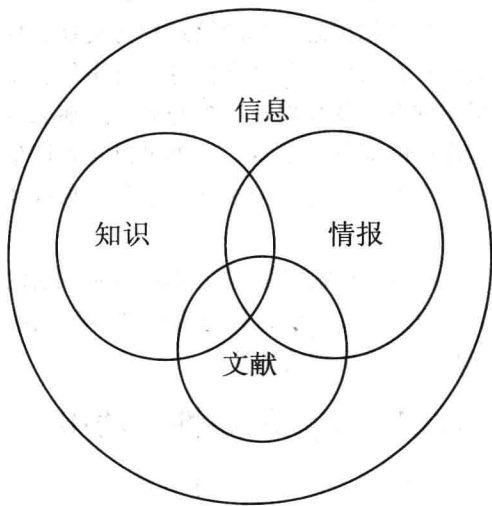


图 1-1 信息、知识、情报以及文献之间的逻辑关系

1.2 文献信息基础知识

本节介绍文献信息基础知识,包括信息资源概念、文献属性与功用、现代文献特征、文献类型等内容。

1.2.1 信息资源与文献信息

信息有 3 个基本要素:信息源、信道、信宿。信息源即信息的来源,是以信号形式发出信息的主体。信息信道是传输信号的通道和媒介。信宿是信息传送的对象或称信息的接收者。把各种来源的信息筛选、组织、加工后,存储在一定的载体上以满足人类对信息的需求。将这些收集加工后的信息集合称为信息资源,即信息资源是经过人类筛选、组织、加工后,存储在一定载体上并且能够满足人们需求的信息集合。

在现代社会中,信息资源与材料、能源一起构成经济发展的三大支柱,是社会发展的重大战略资源。人类的社会生产活动依赖于资源的投入,如同物资生产需要投入物资资源一样,人类的知识生产也需要投入以知识为基本内容的信息资源。另外,信息资源本身的生产即信息的收集、加工和存储也大量消耗人们的知识劳动,凝结着信息资源生产者创造的知识价值。所以,知识是信息资源的主要构成要素,知识性是信息资源的根本属性,信息资源与物资资源和能源资源的区别即在于信息资源是一种知识性资源。此外,信息资源还具有其他一些特殊属性,这些属性是由信息资源的内涵——信息的特性所决定的,如生产的无限性、无损耗性、实效性和共享性等。

信息资源的种类很多,总体上可以分为两大类:一类是文献信息资源,另一类是非文献信息资源。凡以文献形式体现的信息资源称为文献信息资源。而非文献信息资源主要包括以实物为载体的实物信息资源和以人的大脑为载体、通过口头传播的口头信息资源等。非文献信息资源虽然有

生动直观、选择性和针对性强、获取速度快的优点,但稳定性差、不便组织管理。显然,相对于非文献信息资源,文献信息资源因其信息量大、相对稳定、容易收集保存、便于检索利用而成为现代人获取信息的最主要来源。从各种文献信息资源中获取信息,检索和利用文献信息也就成为人们满足信息需求最基本的方式之一。

来自文献信息资源的信息称为文献信息。文献信息一旦进入人的思维活动中,便会产生两种信息:有用信息与冗余信息。有用信息是人类在接收、整理、提炼和利用信息的过程中使其知识结构发生变化以适应自身需要的那部分信息;而冗余信息则是对接收者原有的知识结构没有起作用的那部分信息。人类可以对有用信息进行再加工、再处理,以形成新的再生信息,将这种再生的新信息再记录到一定的文献载体上,就产生具有更新的知识结构形态和价值的文献信息。

文献信息检索资源是以文献为特定载体,以特定的检索方法才能获取和利用的信息资源,文献信息检索是文献信息利用的基础。要掌握文献信息检索的知识和技能,必须先了解作为文献信息资源载体的文献的属性。

1.2.2 文献的属性与功用

“文献”一词,最早见于春秋战国时期的《论语·八佾》:“子曰:夏礼,吾能言之,杞不足征也。殷礼,吾能言之,宋不足征也。文献不足故也。足,则吾能征之矣。”宋代朱熹在《论语集注》中解释《八佾》这段话:“文,典籍也。献,贤也。”意思是所谓“文”是古代典籍,而“献”是指掌握丰富知识的贤人。也就是说,在我国古代“文献”一词指的是文字、言论资料 and 见多识广、满腹经纶的贤人。然而,随着时间的推移,文献概念中,“贤人”的含义慢慢消失了。明代以来,人们所说的文献,通常是指有参考价值的图书资料。如明代永乐年间,编纂了一部大型类书,辑录了明初以前各类图书资料,这部规模宏大的类书就叫《文献大成》也就是后来的《永乐大典》。

至于“文献”在当代的含义,有多种表述方式,或繁或简。国际标准化组织颁布的《文献情报术语国际标准》将文献定义为:“在存储、检索、利用或传递记录信息的过程中,可以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。”1983年我国颁布了《中华人民共和国国家标准文献著录总则》,对文献的定义是:“文献是记录有知识的一切载体。”所谓“一切载体”除了最常见的纸张外,既包含古代用来记载文字图案的甲骨、金石、竹帛、羊皮等,也包含现代科学技术条件下的胶片、磁带、光盘乃至网络等。目前,这一定义已为数学者所接受,并被广泛引用。

如前所述,当代意义上的“文献”包含如下3个基本要素。

一是包含知识信息。凡是文献必以文字、图像、符号、视频、音频等作为记录手段对知识信息加以记录,这是文献的本质属性。知识是文献的实质内容,没有记录下任何知识或信息内容的纸张、胶卷、光盘等不能称之为文献。离开了知识和信息的内容,文献便不复存在。

二是物质载体。只存在于人们头脑中的知识是不能称作“文献”的。知识和信息必须记录在各种载体上。载体是文献得以存在的外在形式,文献所表达的知识信息必须借助一定的符号,依附一定的物质载体,才能长时间保存和传递。离开了物质载体就无法保存和传递知识信息,文献也就不存在了。

三是记录手段。知识信息只有通过一定的手段和方式才能记录在载体上,如通过书写、铸刻、印刷、录制、摄制、刻录等手段和方式将文字、数字、图像、声音等信息符号记录在各种载体上,

形成文献。

文献的产生,是人类进入文明时代的标志。恩格斯在《家庭、私有制和国家的起源》一书中说到:“从铁矿的冶炼开始,并由于文字的发明及其应用于文献记录而过渡到文明时代。”文献在人类社会生活中扮演着重要角色,发挥着无可替代的作用。文献是记录和传播知识信息的重要手段,是人类认识和改造客观世界的主要资源,是人类进行科学交流,获取情报的重要工具,是科学研究的结晶,也是进一步进行研究的基础。人类不断生产文献,又不断使用文献。文献的生产和使用相互促进,生生不息,推动着人类精神文明和物质文明的进步,可以说,没有文献,就没有人类文明的进步。

1.2.3 现代文献的特征

1. 数量大、增长迅速

由于科学技术的飞速发展,人类所拥有的知识财富也越来越多,反映在记录知识的载体方面,则为“文献数量激增”。如今,全世界每年出版图书达70万种,发行期刊为13万种,每年还有新创办的期刊达1000种左右;专利文献每年的登记量超过40万件;产品资料有50至60万件;会议文献的出版量每年高达100万篇。发达国家的科技资料以每两年翻一番的速度递增,近75%的科技资料是在近40年的时间里创造的。如此“浩如烟海”的文献量,使人类社会进入了“文献爆炸”的阶段。

随着科学技术在深度和广度上的不断发展,科学技术文献的数量与类型也在急剧增加。据不完全统计,在科学技术领域,全世界每年出版的研究报告70余万篇,专利说明书100万件以上,会议论文10多万篇,技术标准75万件,产品样本50万件,学位论文10万件,政府出版物20多万篇以上,技术档案百万件。科技文献的出版发行不仅数量庞大,而且增长的速度也令人惊叹。据统计,非科技内容的文献每30~50年才增长1倍,而且科技文献平均7~8年增长1倍,特别是某些尖端科学领域和新兴学科的文献,其增长的速度更快,如原子能科学、环境科学和计算机科学的文献每2~3年就翻一番。文献数量的激增表明文献信息资源的丰富,但同时也产生了“文献污染”,给人们选择、利用文献,获取所需信息造成了不便。

2. 文种多、形式复杂

世界各国用于记录文献的语言文字符号是多种多样的。据统计,世界各国出版的科技期刊连续出版物所采用的文种就有70~80种之多,其中较为集中的文种如英文,占50%左右,德文、俄文各占10%,法文占7%,日文占3%,西班牙文占2%。中文和其他稀有文种占8%左右。随着科学技术的发展和新型材料的问世,新的记录手段不断出现,新型文献载体不断涌现。纸张型文献已失去了一统天下的局面,多种载体的文献相互依存、相互补充、共同发展已成为趋势。新型的非纸张文献如缩微型、机读型、视听型等,加快了信息传递与检索速度,实现了资源共享。

同时随着发表范围和文献交流传递的加速,各国文献所用语种不断增多。据国际连续出版物数据系统(International Serials Data System, ISDS)报道,世界上连续出版物使用的语种多达141种,常用的就有10余种,其中英文约占42%。

3. 内容广泛、相互渗透、重复交叉

现代文献的内容十分广泛,涉及社会各个领域。文献内容的广泛分散主要表现在两个方面:一是同一专业文献,分散刊登在许多专业刊物;二是许多学科与专业刊物发表的文献涉及多个学科领域。

现代科学约有 2 500 多门学科, 它们之间相互渗透、相互包容。原有的学科越分越细, 有的还打破了文科、理科的界限; 同时学科之间又相互交叉, 派生出新的学科, 如边缘学科、交叉学科, 出现了彼此相关的学科群。这些学科发展的特征势必相应地出现在文献内容上。各学科相互渗透体现在以下几个方面。

(1) 各学科期刊论文的相互引用。

(2) 文献内容重复表现在: ①许多在学术会议上宣读的论文先被编入会议预印本, 而后在期刊上发表, 后又被会议文集收录, 有的还被编入图书或科技报告中; ②有的期刊论文不仅被录制到期刊光盘里, 同时又被专业论文集收录, 还可能被制作成缩微胶卷或缩微胶片。

(3) 某一专业的科学论文发表在本专业刊物上的只占一半, 另外一半则发表在其他相关专业的刊物或综合性刊物上。

此外, 文献的重复交叉现象也十分严重。一是科研选题重复。二是同一篇科技文献用多种形式发表。三是同一件技术发明可以用多种文字发表, 据统计, 世界各国每年公布专利说明书的重复率为 65% 以上。四是世界各国相互翻译出版相同内容的文献。据联合教科文组织《信使》杂志报道, 1976 年有 73 个国家相互翻译出版图书达 4 万多种, 约占当时世界图书出版种数的 10%。五是重版、改版文献数量日益增加, 据统计, 每年全世界重版、改版书占出版总量的 20%~30%。

综合和细化是学科发展的趋势, 随着现代科学技术的发展, 学科之间的界限日趋淡化, 学科之间的相互联系、交叉渗透逐渐增强, 这使文献分布呈现出集中又分散的现象。例如, 论文的发表, 某一专业的大部分文章发表在少量的专业性期刊中, 而另外一部分发表于相关专业甚至不相关专业的期刊中。

4. 更新快、时效性强

因为科学的飞速发展, 新的学科不断出现, 原有的知识和学科受其影响, 不同程度地在内容上逐渐老化。因而, 许多文献逐渐失去了原有的价值, 这一点在自然科学的文献上表现得特别明显; 社会科学中的哲学、历史等, 其知识老化缓慢, 但经济类的多数文献知识老化速度也在加快。此外, 文献的语种繁多, 文献的质量良莠不齐, 也是目前文献发展中不容忽视的现状。如此现状, 不可避免地给文献的利用, 即快速、准确地获取知识信息带来了困难。

据前苏联《发明问题》杂志统计, 各类文献的平均寿命为: 图书著作 10~20 年, 期刊论文为 3~5 年, 科技报告为 10 年, 学位论文为 5~7 年, 技术标准 5 年, 产品样本为 3~5 年。西方国家认为, 科技文献中 80%~90% 的内容使用寿命为 5~7 年。

知识和一切生物一样, 是有生命周期的。通常文献的寿命是用“半衰期”来描述文献老化情况。所谓文献半衰期是指某学科领域目前正在使用的全部文献中, 较新的一半所出版的年限。国外有人统计各类学科文献的半衰期为: 地理学 16.1 年, 地质学 11.8 年, 数学 10.5 年, 机械工程 5.2 年, 社会科学 5 年, 冶金学 3.9 年。

随着科学技术的发展, 新知识、新理论、新技术、新产品层出不穷, 加速了知识的新陈代谢, 使文献老化加速, 文献的半衰期越来越短。

5. 文献的载体形式多样化

在“文献爆炸”的今天, 以纸张印刷方式出版的传统文献虽仍占据主要地位, 但因科技现代化而出现的各种各样的新型载体却使文献呈现多样化的状态 (1.2.4 小节有详细描述)。这些采用新材料、新技术制成的文献, 有的加大了信息的储存密度, 有的加快了信息、知识的检索和传递速度, 还有的声像兼备, 增加了信息的传播效果。

6. 文献的来源途径多

由于文献类型的复杂多样,文献的来源也有多种渠道。一般有:①由出版部门正式出版,通过邮局和新华书店公开发行的;②利用出版社名义出版,但不通过邮局和新华书店公开发行的;③国外原版文献通过正常途径或非正常途径进入国内的;④不通过出版部门正式出版,由各学术机构、研究单位和厂矿企业的组织者自行制作的;⑤利用现代化的手段直接通过网络、联机系统获取的;⑥通过单位之间、个人之间等渠道相互交换的。

1.2.4 文献的类型

文献的种类多种多样。

1. 按载体和记录手段不同分类

按载体和记录手段不同将文献分为以下6种类型。

(1) 书写型文献。书写型文献是指手工书写的文献,目前主要指一些古版文献和未经印刷的手稿以及技术档案之类的资料。

(2) 印刷型文献。印刷型文献是一种传统的文献形式,它以纸为主要载体,以石印、油印、胶印、铅印等为主要记录手段,目前仍然为文献的主要形式,如纸质的图书、期刊、报纸等。印刷型文献的优点是便于阅读、传递和广泛流传,缺点是信息存储密度低、体积大、分量重、所占收藏空间大,难以实现机械化、自动化管理。

(3) 缩微型文献。缩微型文献是指以感光材料为载体,以微缩摄影为记录手段的一种现代文献。常见的缩微型文献的载体有缩微胶卷和胶片。利用摄影技术将文献影像缩小记录在胶卷或胶片上的缩微型文献,其主要优点是体积小、容量大、信息存储密度高、便于保存、保存期长,但必须借助于专门的设备——缩微阅读机或缩微复印机才能进行阅读,不像印刷型文献那样可以随时随地阅读。

(4) 机读型文献。机读型文献是指计算机可读的文献,是以磁性材料为介质,以穿孔、打字或光学字符识别装置为记录手段,并通过编码或程序以及程序设计进行计算机处理而产生的一种文献形式。机读型文献的物质形式主要有磁带、软盘、硬盘、光盘等。机读文献最大的特点是信息存储的密度极高、存取与检索的速度快、原有记录可以删除或更新、可重复使用。

(5) 多媒体型文献。多媒体型文献又称视听型或声像型文献,是以唱片、录音带、录像带、幻灯片、电影片、视听光盘等形式出现的记录声音和图像的文献。多媒体型文献是一种把文字、语言、图片、动画、音乐等多种媒体信息综合在一起,通过专用设备播放的虚拟文献。这种文献不仅体积小、存储密度高,而且阅读时直观真切,在帮助人们认识某些复杂或罕见的自然现象,探索物质结构和运动机制等方面具有独特的作用。

(6) 电子网络文献。电子网络文献是把知识信息转化为电磁波,通过卫星通信、光通信等手段传递到用户的电子终端上的文献形式,是一种知识信息共享的新型现代化文献类别。网络文献的传播消除了时间和空间的限制,为实现区域乃至全球范围内的信息资源共享提供了可能,是目前发展最快、应用日益广泛的文献形式之一。

2. 按出版形式不同分类

按出版形式不同将文献分为以下4种类型。

(1) 图书。图书是指系统论述或介绍某一学科或领域知识的出版物。图书文献的特点是有相对独立的内容体系、相当的篇幅和完整的装帧形式。图书涉及的领域广、品种繁多、出版量大、信息含量高,是人们获取知识信息的最重要、最常用文献类型之一,如专著、教科书、工具书、