

信息检索

主 编 高俊宽

参 编 (按章节顺序排名)

邹桂香 高 峰 严大香

丁家友 谢 欢

世界图书出版公司

图书在版编目(CIP)数据

信息检索/高俊宽主编. —上海:上海世界图书出版公司,2017.7

ISBN 978-7-5192-3064-7

I. ①信… II. ①高… III. ①情报检索 IV. ①G252.7

中国版本图书馆 CIP 数据核字(2017)第 126960 号

-
- | | |
|------|---|
| 书 名 | 信息检索
Xinxi Jiansuo |
| 主 编 | 高俊宽 |
| 责任编辑 | 吴柯茜 |
| 封面设计 | 车皓楠 |
| 出版发行 | 上海世界图书出版公司 |
| 地 址 | 上海市广中路 88 号 9-10 楼 |
| 邮 编 | 200083 |
| 网 址 | http://www.wpcsh.com |
| 经 销 | 新华书店 |
| 印 刷 | 上海梓格出版科技有限公司 |
| 开 本 | 787 mm×1092 mm 1/16 |
| 印 张 | 27 |
| 字 数 | 390 千字 |
| 版 次 | 2017 年 7 月第 1 版 2017 年 7 月第 1 次印刷 |
| 书 号 | ISBN 978-7-5192-3064-7/G·495 |
| 定 价 | 78.00 元 |
-

版权所有 翻印必究

如发现印装质量问题,请与印刷厂联系

(质检科电话:021-61421930)

前 言

信息时代,信息能力是大学生必备的能力之一。无论是过去、现在还是未来,如何实现信息的有效查询、获取和利用一直是人类思考的问题。进入大学,随着知识视野的拓宽,专业能力的增加,相应的自学能力也要随之提高和加强。大学生不仅要掌握专业知识,还要掌握相关的学科知识。信息检索课程就是帮助大学生提高自学能力,增强信息素养,实现信息有效查询、获取和利用的一门课程。通过信息检索课程的学习,对信息检索有全面的认识和理解,系统掌握各种类型信息资源的检索技巧和方法,为以后的深造和就业奠定良好的基础。本教材不仅适合信息资源管理、图书馆学、档案学等专业本科学生上课使用,也可满足其他专业本科或研究生查询获取信息和学术研究的参考。

信息技术的发展,信息资源的增加,相应的信息检索技术和信息资源类型也随之发生变化,为了更好地满足网络时代快速有效查询、获取和利用信息的需要,多位工作在教学第一线和具有丰富信息检索实践的老师,联合编写了本教材。与以往和现在的教材相比,本书具有内容丰富、编排合理、操作性强、与时俱进等特点。本书是多位老师合作的成果,具体分工如下:

高俊宽(苏州大学社会学院),负责全书大纲、通稿和第一章、第三章的编写工作。

邹桂香(苏州大学社会学院),负责第二章、第四章的编写工作。

高峰(河南理工大学图书馆),负责第五章、第六章的编写工作。

严大香(苏州大学图书馆),负责第七章、第八章的编写工作。

丁家友(苏州大学社会学院),负责第九章的编写工作。

谢欢(南京大学信息管理学院),负责第十章的编写工作。

本书的出版得到“江苏省‘十二五’高等学校重点专业建设:图书情报与档案管理类”项目的大力资助,在教材编写过程中,院系领导和同事给予了大力支持和帮助,同时参考了大量与信息检索有关的学术论文和论著。姜海涛老师为本书的顺利出版起到了关键作用,在此一并表示感谢。

高俊宽

2016年12月28日

目 录

第一章 绪论	1
第一节 信息及其相关知识	1
一、信息及其相关概念	1
二、信息的类型	4
第二节 信息素质	8
一、信息素质概述	9
二、信息素质的内容	9
第三节 图书馆与互联网	12
一、图书馆	12
二、互联网	14
第二章 信息检索基础	15
第一节 信息检索概述	15
一、信息检索原理	15
二、信息检索的类型	16
三、信息检索发展历史	18
四、信息检索的发展趋势	21
五、信息检索的意义	24
第二节 信息检索语言	26
一、信息检索语言概述	26

	二、分类检索语言	29
	三、主题检索语言	32
	四、自然语言	35
	五、信息检索语言发展趋势	37
第三节	信息检索技术	39
	一、布尔逻辑检索	39
	二、截词检索	41
	三、位置检索	43
	四、限制检索	44
	五、加权检索	45
	六、其他信息检索技术	47
第四节	信息检索评价	49
	一、信息检索数据库的评价指标	50
	二、信息检索结果的评价指标	51
	三、信息检索功能的评价指标	51
	四、用户负担的评价指标	52
第三章	中文图书信息检索	54
第一节	图书基本知识	54
	一、图书概述	54
	二、近代以来图书的形态变革与内容变化	57
	三、电子图书	62
第二节	古籍图书信息的检索	63
	一、经史图书信息的检索	66
	二、类书信息的检索	74
	三、政书信息的检索	80
	四、方志信息的检索	82
	五、检索古籍图书信息的其他主要工具	92

第三节	近代以来图书信息的检索	96
一、	利用书目(检索系统)检索图书信息	96
二、	利用数字图书馆检索图书信息	105
三、	利用网络检索图书信息	112
第四章	中文期刊论文信息检索	117
第一节	期刊基础知识	117
一、	期刊概述	117
二、	核心期刊	119
第二节	综合性期刊论文数据库的检索	121
一、	中国学术期刊(网络版)	121
二、	维普中文期刊服务平台	125
三、	万方数据知识服务平台学术期刊	130
四、	全国报刊索引	133
五、	OA 期刊论文信息的检索	135
第三节	专业性期刊论文数据库的检索	138
一、	人大复印报刊资料全文数据库	138
二、	中文社会科学引文索引(CSSCI)	144
三、	国家哲学社会科学学术期刊数据库	147
第五章	学位论文和会议信息检索	150
第一节	学位论文信息检索	150
一、	学位论文概述	150
二、	国内学位论文信息检索	151
三、	国外学位论文信息检索	167
第二节	会议文献信息检索	174
一、	会议文献概述	174
二、	国内会议文献检索	177
三、	国外会议文献检索	183

第六章 专利和标准信息检索	192
第一节 专利信息的检索	192
一、专利和专利文献	192
二、国际专利分类法(IPC)	196
三、中国专利信息检索	200
四、国外专利信息检索	215
第二节 标准信息检索	237
一、标准与标准文献	238
二、国内标准文献检索	245
三、国外标准文献检索	252
第七章 外文信息的检索	259
第一节 外文综合性信息	259
一、著名外文综合性参考数据库	259
二、常用外文综合性电子期刊及全文数据库	272
第二节 外文自然科学类专业信息	285
一、外文理工类专业信息	285
二、外文生物医学类专业信息	294
第三节 外文人文社科类专业信息	302
一、JSTOR 电子期刊	302
二、EMERALD 系列电子期刊	304
三、Westlaw International 法律资料库	306
四、ERIC 教育学文献数据库	306
五、PsycINFO [®] 心理学文摘数据库	307
六、中国高校人文社会科学文献中心(CASHL)	308
第八章 数值、事实型信息资源的检索	310
第一节 数值、事实型信息资源概述	310
一、数值和事实型数据库含义	310

	二、数值和事实型数据库的主要类型	311
第二节	人物信息和企业产品信息的检索	313
	一、概述	313
	二、万方数据系统中的企业产品和人物信息的检索	314
	三、Gale 机构和人物信息检索	318
	四、Dialog 检索系统中的企业产品信息检索	320
第三节	年鉴的检索和利用	322
	一、年鉴概述	322
	二、Apabi 中国年鉴资源全文数据库	324
	三、CNKI 中国年鉴网络出版总库	327
第四节	百科全书的检索和利用	330
	一、百科全书概述	330
	二、中国大百科全书	333
	三、大英百科全书	334
	四、维基百科	336
第五节	字、词典的检索和利用	338
	一、字、词典(辞典)概述	338
	二、CNKI 中国工具书出版总库	338
	三、网上在线电子字、词典	340
第九章	网络信息检索	344
第一节	网络信息基础知识	344
	一、网络基础知识概述	344
	二、网络信息资源	350
	三、网络信息检索	353
第二节	搜索引擎	362
	一、搜索引擎概述	363
	二、搜索引擎技术与方法	367

三、代表性搜索引擎介绍	371
四、搜索引擎发展趋势	376
第三节 多媒体信息检索	379
一、多媒体信息检索概述	380
二、多媒体信息检索类型	383
第十章 学术规范与论文写作	392
第一节 学术规范	393
一、学术规范概述	393
二、学术规范的作用	394
三、学术规范的内容	396
第二节 论文写作	402
一、学术论文概述	402
二、学术论文的写作程序	404
三、学术论文的组成及撰写	408
参考文献	417

第一章 绪论

信息资源日益成为重要的生产要素和社会财富,信息掌握的多寡、信息能力的强弱成为衡量国家竞争力的重要标志。当前,我国信息资源开发利用不足与无序滥用的现象严重,因此有必要加强顶层设计和系统规划,完善制度体系,全面提升信息采集、处理、传输、利用、安全能力,构筑国家信息优势。

第一节 信息及其相关知识

目前人类正在进入信息时代,但人们对信息的认识各持己见,众说纷纭,各自站在不同的角度,强调不同的方面。据统计,目前对信息的定义不下于一百种,至今没有统一的、规范的、权威的定义。在信息管理领域,除了信息之外,还有与信息密切相关的一些概念,如知识、情报、文献、信息资源等概念。

一、信息及其相关概念

据考证,“信息”一词在我国最早出现于唐代。唐代诗人李中在《暮春怀故人》诗中有这样的诗句:“梦断美人沉信息,目穿长路倚楼台。”其中“信息”一词也就是消息的意思。在西方,信息概念最早出现于通信领域。1928年,哈特利在《贝尔系统电话杂志》发表《信息的传输》一文,首次提出消息是信息的载体,是包含在各种具体消息(语言、文字、图像、信号等)中的抽象量,说明了消息和信息在概念上的差异,实现了信息概念的突破。

(一) 信息

1948年,美国工程师香农在《贝尔系统电话杂志》发表了《通讯的数学

理论》，在文章中从通信的角度提出“信息是人们对事物了解的不定性的减少或消除，也可以理解为信息就是两次不定性之差”。这一定义是从通信的可靠性和有效性界定的，属于狭义的信息概念。同年，维纳出版了专著《控制论》，从控制论的角度认为“信息就是我们在适应外部世界和控制外部世界的过程中，同外部世界进行交换内容的名称”，信息就是认识论，“信息就是信息，不是物质也不是能量”。

继香农和维纳之后，很多研究者从各自不同的研究领域提出了对信息的不同认识。哲学家认为信息是人类认识世界的依据。数学家认为信息是一种概率。物理学家认为信息是“熵”。通信学家认为信息是“不定度”的描述。图书信息领域的专家认为信息是可以以各种形式进行传播、记录、出版及发行的观念、事实及论著。我国《辞海》对信息的释义是：“音讯、消息；通讯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。”美国《韦氏字典》对信息的释义为：“接受由他人传来的知识，这些知识是由他们通过调研和学习获得的；从特定的事件或状况下，如消息、新闻、通知中获得的知识；用来传播的事实和数据。”

无论是从哲学的角度还是从社会科学的角度，或是自然科学的角度，有一点是共同的，就是首先承认信息是事物本身固有的，也就是承认信息的客观存在性，信息的存在是不以人的感知、认识和意志为转移的。我们可以对信息的含义得到以下基本认识：(1) 信息可以改变事物的不确定性；(2) 信息是交流的内容；(3) 信息是事物关系的体现；(4) 信息是与精神和物质并列的；(5) 信息是物质的属性。

信息普遍存在于自然界，人类社会和思维领域。对信息的理解，应当分为两个层次：一个层次是抽象的理论意义上的信息概念，另一个层次是具体的实用意义上的信息概念。抽象理论意义上的信息含义又分为没有任何约束条件的本体论意义的信息含义和受主体约束的认识论意义的信息含义。从本体论层次上看，信息是主体所感知或所表述的事物运动状态及其变化方式，是反映出来的客观事物的属性。而具体的实用信息概念是从不同的利用角度以及不同的学科角度确定信息特定的内涵和外延，相当于狭

义的信息概念。具体来说,从本体论层次上看,信息是一种客观存在的现象,是事物的运动状态及其变化方式的自我显示;从认识论层次上看,信息是主体所感知或表述的事物运动状态及其变化方式,是反映客观事物的属性。

(二) 知识

知识是人类对信息、对客观事物规律的认识,它是人们在社会实践中积累起来的经验,是通过实践对客观事物及其运动过程和规律的认识。人们对事物由表及里、由现象到本质、由感性到理性的认识深化,便形成了知识。

知识是信息被认识了的部分,可以分为主观知识和客观知识。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。所以,知识的产生来源于信息,通过信息传递,并对信息进行加工的结果。从这可以看出,知识是信息的一部分。知识是信息内容的组成部分。

经济合作与发展组织(OECD)在1996年的年度报告《以知识为基础的经济》中将知识分为四大类:(1)知道是什么的知识(Know-what),主要是叙述事实方面的知识;(2)知道为什么的知识(Know-why),主要是自然原理和规律方面的知识;(3)知道怎么做的知识(Know-how),主要是指对某些事物的技能和能力;(4)知道是谁的知识(Know-who),涉及谁知道和谁知道如何做某些事的知识。

(三) 文献

文献一词最早见于《论语·八佾》:“夏礼吾能言之,杞不足徵也;殷礼吾能言之,宋不足徵也。文献不足故也。”朱熹集注:“文,典籍也;献,贤也。”

文献是记录、积累、传播和继承知识的最有效手段,是人类社会活动中获取情报的最基本、最主要的来源,也是交流传播情报的最基本手段。正因为如此,人们把文献称为情报工作的物质基础。现在通常理解为图书、期刊等各种出版物的总和。在国内国外,都常常可以看到有人把“文献”与“信息”,“文献学”与“信息学”等同起来,虽然这种等同未必适宜,但却反映了文献在信息活动和科学中的极为重要的地位。我国《国家标准文献著录总则》

对文献有如下定义：“文献指记录有知识的一切载体。”即用文字、图形、符号、声频、视频等技术手段记录人类知识的一种载体，或理解为固化在一定物质载体上的知识，也可以理解为古今一切社会史料的总称。可见，文献的三个基本要素是：知识、载体和记录。

文献积累着知识，保存着人类文化遗产，它汇集人类创造的精神财富，是知识的宝库。它教育人们正确认识世界，开阔视野，启迪思路，了解过去，预测未来。它是人们接受终生教育，更新知识，提高创新能力的重要智力资源。文献在时间上的传递，实现世代人之间的科学、文化的继承和发展；文献在空间上的传播，促进同代人之间的信息、文化、技术的交流和沟通，推进知识、信息资源共享，加速知识的创造，推动社会的发展。文献是一种智力资源和创造力。

二、信息的类型

按照不同的划分标准，信息可以划分为不同的类型。

（一）按出版形式划分

1. 图书

图书是作者围绕一个中心主题进行研究和探讨，比较成熟的定型而不定期出版的出版物。它一般有封面、书名页、目次、正文，并装订成册。联合国教科文组织为实现国际图书统计的标准化，曾于1961年作出规定，即48页以上的出版物称之为“书”。图书一般包括下面几种类型：专著、教科书、丛书、工具书、论文集等。图书往往给人们以系统性、完整性和连续性的知识和信息。

识别图书的主要依据有书名、著者、出版项（出版地、出版社、出版时间）、总页数、国际标准书号（International Standard Book Number, ISBN）等。

2. 连续出版物

连续出版物是一种具有统一名称、固定版式、统一开本、连续编号，汇集

多位著者的多篇著述,定期不定期编辑发行的出版物。连续出版物具体又分为期刊和报纸。

期刊,又称“杂志”。一般指具有固定题名,定期或不定期出版的连续出版物。期刊上刊载的论文大多数是原始信息,包含有许多新成果和新动向,常见的出版周期有周刊、半月刊、月刊、双月刊、季刊和年刊等。其特点是出版周期短,报道信息速度快,内容新颖、发行及影响面广。

报纸是以刊载新闻和时事评论为主的,定期连续向公众发行的连续出版物,是重要的舆论工具。报纸类型多样(综合性和专业性,日报和周报等),内容丰富。

识别期刊的主要依据有期刊名称,期刊出版的年、卷、期,国际标准刊号(International Standard Serial Number, ISSN)等。

3. 学位论文

学位论文是大学生或研究生为取得学位资格而提交的学术论文。学位论文按学位分为学士论文、硕士论文和博士论文。学位论文探讨的问题往往比较专深,学术性强,往往有独到的见解。

识别学位论文的主要依据有:学位名称、导师姓名、学位授予机构等。

4. 会议信息

会议信息主要是指在国内外各类会议上宣读或交流的论文、报告或其他有关资料。此类文献代表某学科领域的最新成果,反映该学科领域的发展趋势。

识别会议文献的主要依据有:会议名称、会址、会期、主办单位、会议录的出版单位等。

5. 专利信息

专利信息是实行专利制度的国家和地区及国际性专利组织在审批专利过程中产生的官方文件及其出版物的总称。主要有专利申请书、专利说明书和专利公报等。专利信息的种类有发明专利信息、实用新型专利信息和外观设计专利信息。

6. 科技报告

科技报告是各学术团体、科研机构、大学研究所等单位或个人向上级或

委托单位撰写的关于某个科学研究成果的研究报告及其研究过程中的记录。其特点是内容详尽专深、专业性强、内容新颖,有具体的篇名、机构名称和统一的连续编号(即报告号),一般单独成册。

识别科技报告的主要依据有:报告名称、报告号、研究机构、完成时间等。美国政府的四大报告单位 PB(政府部门)、AD(军事系统)、NASA(国家航空航天局)、DOE(能源部)出版的报告都比较权威和著名。

7. 标准信息

标准是为了在一定范围内获得最佳秩序,经协商一致制定并由公认机构批准,共同使用和重复使用的一种规范性文件(GB/T 20001—2000)。标准信息是标准化活动的产物。以文件形式表现的标准化工作成果,就是标准信息。

8. 政府出版物

政府出版物是由国际组织(如联合国、世界贸易组织等)和各国中央政府及地方政府及其所属机构发表、出版的各种文献信息资料。通过这类文献可了解一个国家的科学技术、经济政策、法令、规章制度等。

9. 产品样本

产品样本是国内外生产厂商或经销商为推销产品而印发的企业出版物,用来介绍产品的品种、特点、性能、结构、原理、用途和维修方法、价格等。查阅、分析产品样本,有助于了解产品的水平、现状和发展动向,获得有关设计、制造、使用中所需的数据和方法,对于产品的选购、设计、制造、使用等都有较大的参考价值。

10. 档案文献

档案文献是指国家机关、社会组织以及个人在社会活动和技术活动中,直接形成的具有保存价值和利用价值的文字、图片、声像等不同形式的历史记录。

(二) 按加工程度划分

(1) 零次信息:指未经出版发行的或未进入社会交流的最原始的信息。如私人笔记、底稿、手稿、个人通信、新闻稿、工程图纸、考察记录、实验记录、