

高等院校通识教育“十二五”规划教材

信息检索

Information Retrieval

明均仁 主编

李爱明 曹鹏 李静 杜治波 副主编



人民邮电出版社
POSTS & TELECOM PRESS

C14054972 高等院校通识教育“十二五”规划教材

G254.97-43
09

信息检索

Information Retrieval



明均仁 主编

李爱明 曹鹏 李静 杜治波 副主编



北航

C1744619

人民邮电出版社

北京

G254.97-43
09

图书在版编目 (C I P) 数据

信息检索 / 明均仁主编. — 北京 : 人民邮电出版
社, 2014. 8
高等院校通识教育“十二五”规划教材
ISBN 978-7-115-35844-8

I. ①信… II. ①明… III. ①情报检索—高等学校—
教材 IV. ①G252.7

中国版本图书馆CIP数据核字(2014)第146031号

内 容 提 要

本书是一本全面、深入介绍大学生常用检索工具及其数据库的教学用书。本书以通过各种信息资源检索工具集数据库获取不同类型的信息资源为主线,介绍了信息检索的基本知识,详细介绍了图书、期刊、会议论文、科技报告、学位论文、专利等类型的信息资源的特点、国内外主要检索工具和检索方法,并针对计算机和网络技术的发展,对网络信息检索和图书馆检索服务进行了深入的阐述,结合大学生的需求,详细介绍了学位论文写作所需的资料准备、构成和撰写等内容。本书实用性强,适用面广,具有鲜明的时代特征。

本书注重理论与应用的结合,适用于高等院校图书馆学情报学专业、信息管理专业、工商管理和其他相关专业的教学用书,同时也可用作大专院校学生的文献检索课程教材。

-
- ◆ 主 编 明均仁
 - 副 主 编 李爱明 曹 鹏 李 静 杜治波
 - 责任编辑 王小娟
 - 责任印制 张佳莹 焦志炜
 - ◆ 人民邮电出版社出版发行 北京市丰台区成寿寺路 11 号
 - 邮编 100164 电子邮件 315@ptpress.com.cn
 - 网址 <http://www.ptpress.com.cn>
 - 北京艺辉印刷有限公司印刷
 - ◆ 开本: 787×1092 1/16
 - 印张: 14 2014 年 8 月第 1 版
 - 字数: 349 千字 2014 年 8 月北京第 1 次印刷
-

定价: 38.00 元

读者服务热线: (010)81055256 印装质量热线: (010)81055316
反盗版热线: (010)81055315

前 言

我们面临的是一个崭新的信息社会,信息已成为当今社会发展的重要战略资源。信息检索作为专门研究信息存储与信息获取的学科,对于提高大学生的信息素养和信息获取能力具有重要的意义。

《信息检索》就是以增强大学生的信息意识,培养学生的信息素养为目标而编写的一本专门针对大学生的信息检索教程。本书不仅介绍了信息检索的基本知识,并按照信息资源的类别详细介绍了各类信息资源的检索和使用,包括图书信息检索、期刊信息检索、特种文献检索、网络信息检索和图书馆检索服务,最后还专门针对每个大学生必须完成的学位论文,详细论述了学文论文的资料准备、构成和撰写等内容。

目前,几乎所有高等院校都开设了类似信息检索的通识教育课程,其目的就是为了使非图书馆情报专业的学生能更好地使用各种检索系统,让其能更好地利用各种信息资源。因此,我们组织多位优秀教师编写本书,目的在于使读者初步掌握大学生经常使用的信息检索系统,并且结合自身需要完成对相关信息资源的检索。

本书共8章,由明均仁担任本书的主编,负责全书的组织设计、质量控制和统编定稿。本书编写分工如下:第1章由黄梅林编写;第2、3章由李爱明编写;第4章由曹鹏编写;第5章由余倩和杜治波编写;第6、8章由吴紫娟、李静和李艳编写;第7章由明均仁编写。

在本书编写过程中,编者借鉴了部分国内外最新的出版物、数据库和网上资料,因为篇幅所限未能在文中一一注明,在此谨向各位专家学者表示由衷的感谢。由于书中涉及国内外的信息检索系统较多,且各个检索系统不断升级更新,也由于作者水平有限,加之时间仓促,书中难免存在不妥或错误之处,竭诚恳请同行与读者批评指正,以期改进。

编 者

2014年4月

目 录

第1章 绪论	1
1.1 大学生信息素质教育	2
1.1.1 信息素质的定义	2
1.1.2 信息素质的个体特征和评判标准	2
1.1.3 信息素质教育的意义	3
1.1.4 信息素质教育的现状	4
1.2 信息与信息资源	5
1.2.1 信息、知识、情报	5
1.2.2 信息资源的类型	6
1.2.3 网络信息资源	8
第2章 信息检索基本知识	10
2.1 信息检索概论	10
2.1.1 信息检索的定义与特点	10
2.1.2 信息检索基本原理	12
2.1.3 信息检索工具	13
2.1.4 信息检索的意义和作用	15
2.1.5 大学生学习信息检索知识的意义	17
2.2 信息检索语言	18
2.2.1 检索语言的含义与作用	18
2.2.2 分类语言和分类法	20
2.2.3 主题语言和主题词表	22
2.2.4 自然语言	23
2.2.5 网络环境下信息检索语言的发展趋势	25
2.3 计算机信息检索技术	27
2.3.1 布尔逻辑检索	27
2.3.2 截词检索	28
2.3.3 进阶检索	29
2.3.4 限制检索	29
2.3.5 词位检索	29
2.4 信息检索的方法与途径	30
2.4.1 信息检索的方法	30
2.4.2 文献检索的途径	31
2.5 信息检索的基本步骤	32
2.5.1 分析研究课题	33

2.5.2 确定检索工具	33
2.5.3 选择检索方法	34
2.5.4 确定检索途径和检索策略	34
2.5.5 实施检索并调整检索策略	34
2.5.6 索取原始文献	35
2.6 信息检索效果评价	36
2.6.1 信息检索效果的定义	36
2.6.2 信息检索效果的评价指标	36
2.6.3 影响查全率和查准率的因素	36
2.6.4 提高检索效果的措施	37
第3章 图书信息检索	38
3.1 图书馆书目检索系统	38
3.1.1 图书分类与排架方法	38
3.1.2 图书馆馆藏书目检索	40
3.1.3 CALIS 联合目录数据库	42
3.1.4 中国国家图书馆联机公共目录馆藏查询	43
3.2 网上书店	45
3.2.1 当当网	45
3.2.2 卓越亚马逊	48
3.3 电子图书	49
3.3.1 电子图书概述	49
3.3.2 方正 Apabi 电子图书	51
3.3.3 NetLibrary 电子图书	57
3.4 数字图书馆	60
3.4.1 书生之家数字图书馆	61
3.4.2 超星数字图书馆	64
第4章 期刊信息检索	66
4.1 中国期刊网 CNKI 全文数据库	66
4.1.1 CNKI 及数据库简介	66
4.1.2 中国期刊全文数据库检索	67
4.1.3 CAJ 阅读器常用功能介绍	72
4.2 维普科技期刊全文数据库 (VIP)	74
4.2.1 维普数据库简介	74
4.2.2 维普数据库检索	74
4.2.3 维普 (VIP) 阅读器	80

4.3 万方数据资源系统	84	6.3 学术搜索引擎	168
4.3.1 万方数据库简介	85	6.3.1 读秀学术搜索	168
4.3.2 万方数据库检索	85	6.3.2 Google 学术搜索	170
4.4 SPRINGER-LINK	89	6.3.3 免费网络学术资源搜索	171
4.4.1 SPRINGER-LINK 数据库简介	89	6.4 专题型、专业型网络检索	174
4.4.2 SPRINGER-LINK 数据库检索	89	6.4.1 白页与黄页搜索	174
4.5 ELSVIER	92	6.4.2 地图搜索	175
4.5.1 ELSVIER 数据库介绍	92	6.4.3 域名搜索	175
4.5.2 ELSVIER 数据库检索	94	6.4.4 FTP 搜索	177
4.6 EBSCO	99	6.4.5 新闻搜索	178
4.6.1 EBSCO 数据库介绍	99	6.5 常用搜索引擎	179
4.6.2 EBSCO 数据库检索	101	6.5.1 百度	179
第5章 特种文献检索	108	6.5.2 谷歌	183
5.1 学位论文	108	6.6 学科信息资源	185
5.1.1 学位论文概述	108	6.6.1 学科信息资源的类型	185
5.1.2 学位论文数据库的特点与检索方法	108	6.6.2 学科信息的查找途径与方法	185
5.2 会议文献	118	6.7 大学生常用信息网站	186
5.2.1 会议文献概述	118	6.7.1 英语	186
5.2.2 会议文献数据库的特点与检索方法	118	6.7.2 考研	188
5.3 科技报告	125	6.7.3 就业	188
5.3.1 科技报告概述	125	6.8 搜索网上信息的方法与技巧	189
5.3.2 科技报告的类型	126	6.8.1 关键词的选择	189
5.3.3 科技报告的网络检索	126	6.8.2 基本检索语法	190
5.4 专利文献的检索	128	6.8.3 一些特殊的搜索功能	191
5.4.1 专利文献概述	129	6.8.4 给用户的建议	192
5.4.2 国际专利分类法简介	129	第7章 图书馆检索服务	194
5.4.3 专利文献的网络检索	130	7.1 参考咨询服务	194
5.5 标准文献的检索	152	7.2 定题服务	196
5.5.1 标准文献概述	152	7.3 馆际互借与文献传递服务	197
5.5.2 标准文献的特点和作用	153	7.4 科技查新服务	200
5.5.3 标准文献的分类体系和代号	154	第8章 学位论文写作	202
5.5.4 标准文献的检索	155	8.1 学位论文资料准备	202
第6章 网络信息检索	167	8.2 学位论文结构	202
6.1 网络信息的特点	167	8.3 学位论文撰写	205
6.2 搜索引擎的分类	167	8.4 参考文献标准	209
		8.5 论文写作学术规范	212
		参考文献	216

第1章 绪 论

知识是人类进步的阶梯，而书籍是知识传承的载体，藏书丰富的高校图书馆就是大学生们身边一座取之不尽的知识“金矿”，是走出课堂后仍陪伴身边的最博学的老师。

著名作家、哲学家赵鑫珊在他的散文自传《我是北大留级生》中坦言，他为了能在北大图书馆继续汲取知识，故意挂科留级一年，如饥似渴地大量阅读国际大师们的原著，打通文理科学界限，为自己种下了哲学、艺术、科学的种子，为后来瞩目的成就奠定了坚实的基础。当代女作家迟子建也饱含深情地回忆，她因高考失利只考上一所名不见经传的大兴安岭示范专科学校，可是却拥有充足的时间泡在图书馆里看书，觉得图书和大自然对她的帮助很大，并建议人们多读经典书，“不读经典就像你每天早上只吃维生素不吃早饭，这样会造成营养的失衡。文化的传承一定要有一个过程，不读经典作品，你就不知道它的滋养有多深厚。”

如果说，课堂上老师把你领进知识的大门，那么图书馆就是个人修行的最好去处，身处图书馆书山学海之中，那种刻苦钻研、自强不息的精神定会得到极大的激发，而独立思考、勇于探索的能力也会得到很好的培养。这也为日后能培养终身学习的习惯，适应社会发展变化、不断接受新生事物、勇于创新打下良好的基础。

有人曾告戒即将走入大学的学子们：如果说，在大学里有一门课不能逃，那就是图书馆；如果说，在课堂之外有个能让你成长的地方，那就是图书馆；如果说，在学校里一个最有价值的地方是哪里，那就是图书馆；因此，倘若在大学的几年时间里你很少去图书馆，就等于浪费了一大笔财富。即使你所在大学没有像北大、清华那样的名牌大学的图书馆，但是里面的信息量肯定足够让你吸收了。有空就去图书馆吧，那里绝对是一块“风水宝地”。

其实，每一个不愿荒度光阴、对知识如饥似渴的大学生都会迫不及待地走入图书馆这座知识的宝殿。然而，令人遗憾的是，面对种类繁多的图书、期刊、报纸资料，身处一排排森然林立的书柜，新入校的大学生往往感到茫然不知所措。更可惜的是，一些快毕业的大学生竟然对馆内所藏数字图书及网络资源还一无所知，当他们撰写毕业论文时才急匆匆地跑去查询。这些现象说明大学生们对图书馆资源缺乏深度了解，也反映了信息时代大学生对信息缺乏敏锐感，其信息处理和利用能力亟待提高。

为了帮助大学生能更好地利用好图书馆资源，各高校专门开设信息检索课程，旨在提高大学生的信息素养，使他们了解并掌握各种文献资源及其检索方法，提高信息获取与利用能力。本教材首先对电子图书资源、期刊信息资源、特种文献资源、网络信息资源等资源的检索方法做全面而具体的介绍；然后，介绍图书馆可以提供的各种检索服务；最后，针对大学生在撰写学位论文时可能遇到的问题进行技术指导。



1.1 大学生信息素质教育

网络时代,文献信息服务方式和知识载体的多元化趋势日益明显,知识的传播手段和传播渠道有了重大的改变,表现出了前所未有的多样性和复杂性。虽然,人们可以轻而易举地通过互联网获取大量的信息和知识,然而,要在庞大而杂乱的信息海洋里迅速准确而全面地筛选所需的信息却很难。作为一个文献信息储存和传递中心的图书馆,就在信息整合、信息检索、信息咨询等方面默默地开展自己的工作,为读者提供快捷方便的服务。所以,图书馆仍然是大学生查询信息知识的首选之地。

尽管图书馆提供了丰富的信息资源,但大学生读者也需培养和提高信息素质,才能更好地获得自己所需要的信息。

1.1.1 信息素质的定义

信息素质一词,最早是由美国信息产业协会主席 Paulski 在 1974 年提交的一份报告中提出的。当时,他将信息素质定义为“经培训后能够在工作中运用信息的人即认为具备了信息素质,他们在掌握了信息工具的使用及熟悉信息源的基础上,能够解决实际问题。”经过 5 年的研究,信息产业协会于 1979 年更新了该定义,认为具备信息素质的人即“掌握了自用信息工具的知识与技能,并能够将其应用于解决实际问题的人。”

1989 年美国图书馆协会(American Library Association, ALA)将信息素质的概念明确为“具备信息素质的人,是能够敏锐地洞察信息需求,并能够进行相应的信息检索、评估和有效利用所需信息的人。”

我国学者马海群指出:信息素质的广义为在信息化社会中个体成员所具有的各种信息品质,包括信息智慧(涉及信息知识与技能)、信息道德、信息意识、信息觉悟、信息观念、信息潜能、信息心理等。

美国学者 H.B.Rader 的定义:信息素质是在解决问题和作决策时具有有效地找到和评价信息的能力。包括在现代信息环境中求生存、谋发展的能力,迅速适应外部环境变化的能力,解决问题时能把合适的信息找出来的能力,还包括熟练使用计算机方面的能力。他认为,信息素质是未来社会生存的必备条件,具备较强信息素质的人将是在信息社会中能够终身学习的人。

简单地说,信息素质就是人们获取、评价和使用知识信息资源的能力,是人们跨入信息社会,在生理素质、心理素质和社会文化素质等基本品质的基础上,发展并优化出来的一种新品质,是信息时代对人类的一种高要求,是现代人才在信息环境下获取、加工和利用知识信息必需的技能。信息素质的内涵包括 3 个方面:信息意识、信息能力、信息道德。信息意识是指人们对信息的敏感程度。信息能力是指发现、评价、利用和交流信息的能力。信息道德是指在信息活动中应遵循的道德规范。

1.1.2 信息素质的个体特征和评判标准

信息素质是现代人所必备的一种能力、品质或素养,作为一个具备信息素质的个体应具有的特征如下。

美国全国图书馆协会和教育传播与技术协会于1998年制定了九大信息素质标准。下面从信息素质、独立学习和社会责任这三方面加以表述。

信息素质的标准包括:具备信息素质的学生能够有效并高效地存取信息,能够全面地、评判性地评估信息,能够准确地、创造性地利用信息。

独立学习的标准包括:具有独立学习能力的学生具备较高的信息素质,能探求与个人兴趣有关的信息,能够理解文献及对信息进行其他形式的创见性的表达,能够在信息查询与知识创新中力争最优。

社会责任的标准包括:对学习社区和社会有积极贡献的学生具备信息素质,并能认识信息对民主社会的重要性,能够基于伦理道德规范利用信息及信息技术,在探求和创新信息的过程中具有团队精神。

除了上面介绍的由一些国家机构或协会制定的标准和指南之外,一些大学在实施信息素质教育教学的过程中,也制定了切合自身特色的信息素质标准。例如:在加州大学(CSU)信息素质教育计划中,确定信息素质共包括10种核心能力:①确定研究主题;②确定适合于研究课题的信息需求;③检索与研究主题相关的信息;④利用信息技术获取信息;⑤评估信息;⑥组织和分析信息;⑦利用信息技术进行交流;⑧明确有关信息和信息技术的伦理、法律和社会政治问题;⑨评判性地利用、评价和对待从大众媒介中获取的信息;⑩对研究过程和结果进行评价。

1.1.3 信息素质教育的意义

1. 信息素质教育是社会经济发展的需要

在信息社会,经济主体由制造业转向以高新科技为核心的第三产业,即信息和知识产业,劳动力主体也不再是机械的操作者,而是信息的生产者和传播者。信息技术在资料生产、科研教育、医疗保健、企业和政府管理及家庭中广泛应用,在对经济和社会发展产生了巨大而深刻的影响的同时,也从根本上改变了人们的生活方式、行为方式和价值观念。高效查询、存储、利用信息成了人们日常工作和学习中必不可少的技能。只有接受信息素质教育,人们才能适应信息时代的需求,找到理想的工作,自由自在地享受信息技术带给人们的各种便捷服务。

2. 信息素质教育是信息时代促进高等院校进行的一场教育改革

高等院校是顺应时代培养符合社会需要的合格人才,信息社会需要具备高信息素质的人才,高等院校就必须在教育观念、教学计划、教学课程安排等方面不断地进行革新,对在校大学生进行信息素质教育,培养其良好的信息意识、较强的信息能力和较高的信息道德,尤其是要使他们牢固树立终身学习的观念,不断地更新自我,完善自我,适应知识经济时代发展的需要。

目前,高等院校在信息素质培养上存在一定的误区,如将信息检索课程或计算机课程完全等同于信息素质教育。事实上,信息素质教育的内涵不仅包括利用信息的能力,还包括强烈的信息意识和信息道德意识等。高等院校信息素质培养不能光靠图书馆一个部门,可以结合各高等院校的实际情况大胆尝试各种教学形式。例如,采用联机信息素质指导(Online Information Literacy Instruction)替代课堂教学,作为课堂教学的辅助。还可以将信息素质教学活动置于实现其他课程教学目标的背景下进行,在辅助完成其他课程教学任务的同时,也实现了培养学生信息素质能力的目标。总之,信息素质教学目标的实现需要图书馆员、教师、



学生及管理层的密切合作,才能更好地开展信息素质教育。

3. 信息素质教育是当代大学生必须接受的一项通识教育

大学生首先应该懂得,信息素质是现代人类社会一种新的基本品质,掌握信息技术是现代人类立足社会应该通晓的一项普通技能。在大学期间要自觉地接受信息素质教育,努力提高个人信息素质。大学生要在课堂外多多利用图书馆,善于利用图书馆,把图书馆当成课堂外的另一个不可缺少的教师。刚走入大学的新生要迅速熟悉和了解图书馆各种馆藏资源(纸质文献、数字图书、网络资源等),要结合专业课的学习主动上图书馆查询相关信息资料,以获得及时辅导,并拓展其知识结构,培养一种自我探究式的学习能力。在将来走向社会后,大学生应该将这种学习方式延续到今后的生活和工作中,适应日新月异的信息技术的变革和不断发展的时代步伐。

1.1.4 信息素质教育的现状

自20世纪50年代第一台电子计算机诞生以来,信息技术得到迅猛发展,随着20世纪80年代末互联网的出现,一个全新的网络经济由此产生,人类社会开始从工业时代跨入了信息时代。

可以说,互联网正在快速地改变着全球经济,也改变着人类的文明,改变着我们每一个人的生活方式和工作方式。信息取代物质、资本而成为社会的战略资源,要想在社会竞争中立于不败之地,就必须最先获取丰富的信息资源,因此,信息竞争中急需大批具备信息素质的人才。高等院校就是为信息社会培养信息利用的合格人才的地方,信息素质教育应当成为高等院校的重要任务。

美国是一个危机意识十分强烈的国家,尤其在教育领域显得尤为突出。早在20世纪70年代末80年代初,由于多项研究成果表明,美国学生在数理学科和其他学科领域的测验成绩令人遗憾,教育优化全国委员会发表了一篇《处于危险中的国家》的报告,后来在里根总统的倡议下,该委员又撰写了一份报告,阐明了学业水准下降的现状与提高教学质量的必要性,并在全美的电视、报纸等新闻媒体广为报道,引起了公众对教育问题前所未有的关注,促使许多州的立法机关和学校采取措施以改进教学工作,并制定了一系列信息素质相关标准。从信息素质教育评估标准可以看出,当时美国社会已经在为未来信息社会竞争培养人才做准备,具有高瞻远瞩的战略意义。

20世纪80年代初,邓小平提出“计算机普及要从娃娃抓起”,这句意义深远的话推动了我国计算机普及事业和信息科技事业的长足发展,也引起了教育界对信息素质教育的高度重视。

2000年,我国教育部印发了《中小学信息技术课程指导纲要(试行)》的通知,明确中小学信息技术课程的主要任务是:培养学生对信息技术的兴趣和意识,让学生了解和掌握信息技术基本知识和技能,了解信息技术的发展及其应用对人类日常生活和科学技术的深刻影响;通过信息技术课程使学生具有获取信息、传输信息、处理信息和应用信息的能力,教育学生正确认识和理解与信息技术相关的文化、伦理和社会等问题,负责任地使用信息技术;培养学生良好的信息素养,把信息技术作为支持终身学习和合作学习的手段,为适应信息社会的学习、工作和生活打下必要的基础;并要求信息技术课程的设置要考虑学生心智发展水平和不同年龄阶段的知识经验和情感需求,小学、初中和高中阶段的教学内容安排要有各自明确的目标,要体现出各阶段的侧重点,要注意培养学生利用信息技术对其他课程进行学习和探



讨的能力;努力创造条件,积极利用信息技术开展各类学科教学,注重培养学生的创新精神和实践能力。

教育部一直倡导高等院校培养大学生良好信息素养以适应信息时代的需要。虽然文献检索课不能取代信息素质培养,但它也是信息素质培养的一个重要部分。20世纪80年代初,教育部就颁发了《中华人民共和国高等学校图书馆工作条例》,将文献检索课规定为高等院校图书馆工作任务之一。1992年5月,国家教委又印发了《文献检索课教学基本要求》,对文献检索课的课程性质、教学目的要求、课程组织计划、教学检查评估做了更细致而全面的规定。1999年6月,中共中央、国务院在《关于深化教育改革全面推进素质教育的决定》中指出:在“一次教育”已经过时,“终生教育”成为时代要求的形势下,现代高等教育的一个重要目标就是要培养大学生的信息素质。2002年,教育部高等院校图书情报工作指导委员会主办的全国高等院校信息素质教育学术研讨会在黑龙江大学召开。首次将文献检索课程教学学术研讨会改名为“信息素质教育学术研讨会”,表明文献检索课教学已进入了新的阶段,发生了质的变化,图书馆用户教育又向前迈进了一大步。近几年,教育部大力提倡包括信息素质教育在内的素质教育,高教司已在“新世纪高等教育教学改革工程”中设立了一个“网络条件下的文献信息用户教育研究”项目,并委托清华大学图书馆主持研究。

尽管国家教育部门十分重视大学生信息素质的培养,但是,我国高等院校对信息检索课程普遍不予重视。大多数高等院校把信息检索课仅作为选修课,甚至由于学生选修的人比较少而干脆取消这门课。由于我国教育还是以应试教育为主,信息素质教育大都只停留于口号中,所以导致大学生的信息素质的层次普遍较低。60%的学生不了解信息系统,停留在纸本文献上;70%以上的大学生在撰写毕业论文、毕业设计时不会利用图书馆藏书,甚至工作后也不会获取与利用信息。

目前,国内还缺乏像美国高等院校那样的信息素质评估标准,更没有一个评估和考量的机构。总之,在信息素质教育上,我国高等院校同国外比还显得过于形式化、表面化,仍存在很大的差距。

1.2 信息与信息资源

1.2.1 信息、知识、情报

信息、知识、情报等,通常被人们混为一谈,但学术界却一直对这些概念进行探讨,认为它们之间有着区别与联系。

什么是信息,不同领域、不同学科之间有着不同的见解。有的认为信息是物质的范畴与属性,有的认为信息就是事件,有的把信息看成结构,有的把它看做知识的代表等。但总体可概括为以下两类基本观点。

一是信息的哲学观。从哲学角度来看,信息的概念存在两个基本层面,即本体论信息和认识论信息。所谓本体论信息是指一切事物的存在方式和运动状态的自我表述。不同的事物有着各自不同的存在方式与运动状态、运动规律,这就构成了各种事物的特征,也就是这些事物各自发出的不同信息。这强调的是信息的客观存在性。而所谓认识论信息是指主体对事物的存在方式和运动状态的具体描述。人们是通过识别信息来认识世界和改造世界的。这强

调的是信息的主观认识性。信息既是客观存在的，又是人类主观认识的产物。

二是信息的发生观。由信息发生情况不同，可以将信息分为3种类型：自然信息、生物信息和社会信息。自然信息是指山川、河海、日月、星辰等自然界的變化所提供的信息。生物信息是指生物之间或生物与人之间相互联系使用的语言或信号，如气味语言、声音语言、色彩语言、运动语言等。社会信息是指人与人之间所传递的信息，它包括各种消息、知识、情报等。人与人传递信息的工具通常有语言、文字、图像、电子数据等。

信息具有一种基本特征，即独立性。同一信息可以用文字记录，也可以用语言表达，可以用磁带录音存储，也可以用图像显示，其意义保持不变。

信息普遍存在于自然界、人类社会及人们的认识和思维过程中，因而具有客观性、普遍性、无限性、相对性、抽象性、依附性、动态性、共享性、传递性等特点。在现代社会中，人们每天都要处理和传递各种不同形式的信息，随着信息量的增大、信息价值的增长、信息交换频率提高，信息已成为现代社会一大资源。从形式上看，信息的量是可测量的；从内容上看，信息是可评估和选择的；从价值上看，信息是不会因使用而耗损或消亡，但会随着时空的变化而失去意义。

什么是知识？同“信息”一样也没有一个统一的定义，但大致存在以下两种观点。

一是知识的认识观。即认为知识就是人类对自然和社会运动形态和规律的认识和掌握，是通过感性认识向理性认识飞跃，从现象到本质，从而掌握事物的本质的客观规律。

二是知识的信息观。认为知识就是人脑接受外界的信息，意识和思维把收到的信息与原已存储（记忆）的信息联系起来，进行加工组合，从而建立起一个一个个的信息系统，这就是一个一个个的知识单元，由这些知识单元进而构成关于客观世界的各种事物的概念、规律和理论。由此可见，信息是产生知识的原材料，知识是信息加工的产物。

知识作为信息系统，仍然具有信息的基本特征，能够传递、接受、存储、处理、输出。

什么是情报？有人认为情报是提供给特定对象思考与行动的经过筛选和整理的知识。而更多的人认为情报就是某对象所需的特定信息，例如，科技部门所需要的信息称为科技信息或科技情报，军事部门所需的信息就是军事情报，图书馆向用户提供的信息就是图书情报。

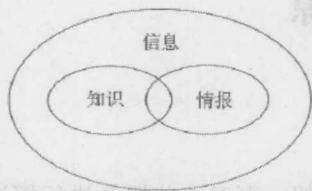


图 1.1 信息、知识和情报三者之间的关系图

信息、知识和情报三者之间既相互关联又相互区别。信息包含知识和情报，知识、情报都是特殊的信息。其中，知识是人类认识事物运动规律的系统化信息，而情报是向特定对象传递的特定知识或信息。知识和情报之间是一种交叉关系，提供给特定对象的情报有的是知识，也有的只是某种信息。因此得出关于三者之间的关系图（见图 1.1）。

1.2.2 信息资源的类型

马费成的《信息资源管理》认为：“所谓资源，是指在自然界和人类社会中可以用以创造物质财富和精神财富的原始的达到一定量的客观存在形态。如土地资源、矿产资源、森林资源、海洋资源、石油资源、人力资源等。一般可分为自然资源和社会资源两大类。”随着信息时代的到来，信息已经成为一种越来越重要的社会资源，因为，人们通过对信息进行加工、存储并利用，可以带来极为广泛的经济效益和社会效益，成为人类发展所必需的资源，与物质资源、能源资源一起，成为现代社会经济发展的三大支柱。

什么是信息资源?以“资源”的定义推之得出,可以为人类创造物质财富和精神财富的一切信息就是信息资源。其实,对信息资源的理解,应该分为两个层面:从狭义上讲,信息资源是由信源发出,通过人脑和智能机器加工处理的一切信息;从广义上讲,信息资源是指人类社会信息活动中以能创造物质财富和精神财富的信息为核心及参与加工、存储和利用这些信息的一切设备、技术、人才的集合,也就是与信息活动相关的各种资源的总称。

信息资源的范围十分广泛,种类繁多,其分类方法也多种多样。

1. 从信息资源所描述的对象来划分

从信息资源所描述的对象来划分,信息资源可分为自然信息资源、机器信息资源和社会信息资源等。其中,自然信息资源是指来自自然界的的信息资源。机器信息资源是指反映和描述机器(体系)本身运动状态及变化特征的信息资源。社会信息资源是指人类生产与生活中不断产生和交流的信息的资源。

2. 从信息资源的载体和存储方式划分

从信息资源的载体和存储方式划分,信息资源可分为天然信息资源、智力信息资源、实物信息资源、文献信息资源、网络信息资源等。其中,天然信息资源是指以自然物质为载体未经人类发掘的自然信息资源。智力信息资源是指以人脑为载体经过人脑加工的信息资源,包括人所掌握的技能、经验等。实物信息资源是指以人类生产出的产品、样品、模型等实物为载体的信息资源,人们可以通过这些加工出来的实物所传递的信息进行学习和仿制。文献型信息资源是以文献为载体的信息资源,文献信息资源依据其记录方式和载体材料又可分为刻写型、印刷型、缩微型、机读型、声像型等五大类。网络信息资源是指从计算机技术、通信技术、多媒体技术相互融合而形成的网络上可查找到的资源。

文献信息资源根据不同的划分方式又可分为以下几类。

① 按照记录方式和载体材料为依据划分书写型文献信息资源、印刷型文献信息资源、缩微型文献信息资源、机读型文献信息资源、声像型文献信息资源。

书写型文献信息资源:一般以纸张为载体,记录方式为人工抄写,包括手稿、信件、日记、原始档案等。

印刷型文献信息资源:主要以纸张为载体,记录方式主要是印刷技术,包括油印、铅印、胶印、木版印刷、复印、激光打印等。

缩微型文献信息资源:以感光材料为载体,记录方式主要是光学记录技术,主要类型有缩微胶卷、缩微平片、缩微卡片等。

机读型文献信息资源:以磁介质、光介质材料为载体,记录方式为磁录、光录技术,主要类型有磁带、磁盘、软盘、光盘等。

声像型文献信息资源:以感光材料和磁性材料为载体,记录方式为光录技术和磁录技术,主要类型有唱片、录音录像带、电影胶卷、胶片、幻灯片等。

② 按照不同加工深度为依据划分为零次文献、一次文献、二次文献、三次文献。

零次文献:主要指尚未经过系统整理形成一次文献的零散资料,如未正式发表的书信、手稿、讨论稿、实验的原始数据、工程草图、人们在某些专业会议上口头交流的经验或某些论点等。

一次文献:主要指作者以本人的研究成果为基本素材而创作或撰写的文献,如图书专著、期刊论文、专利说明书等。

二次文献:主要指文献工作者对一次文献进行加工、提炼和压缩之后得到的产物,是为



了便于管理利用一次文献而编辑、出版和累积起来的工具性文献,一般包括目录、题录、文摘、索引等。

三次文献:主要指对有关的一次文献和二次文献进行广泛深入的分析研究之后综合概括而成的产物,具体包括述评、综述、文献指南等。

③ 按照不同的著录形式为依据划分图书、期刊、报纸、研究报告、会议文献、专利文献。本书将按该种文献划分类别来讲述。

图书:论述或介绍某一学科或领域知识的出版物。它有两种类型:一是阅读类图书,二是参考类图书。

期刊:印刷或非印刷形式的出版物,具有统一的题名,定期或不定期以连续分册形式出版,有卷期或年月标志,并且计划无限期地连续出版。

报纸:有固定名称、刊期、开本,以新闻报道为主要内容,每周至少出版一期的散页连续出版物。

研究报告:一种特殊类型的信息源,一般用来反映科学研究和开发调查工作的成果或进展情况。研究报告也具有向上级或其他相关人员陈述工作成果的作用。

会议文献:各类学术交流会议的资料和出版物。它包括会议前和会议中,与会者提交的会议论文预印本、与会者在会议上宣读的论文、会议期间散发的资料和会议后经过编辑加工的正式出版物。

专利文献:由国家专利主管机关出版的,对各种发明创造进行详细描述,并表明在一定时期内发明所有人拥有独占权的法律文件的总称。

此外,还可从信息资源的内容划分为政治、经济、文化等方面的信息资源;从信息资源的来源划分为来自组织内部的信息资源和来自组织外部的信息资源。信息资源的划分标准和方式并不是唯一的固定的,之所以要从不同角度去划分,是为了更好地认识信息资源的特征,以便对信息资源进行管理、开发和利用。

1.2.3 网络信息资源

随着计算机和网络技术的高速发展,我们的信息渠道已不再局限于书籍、报刊,而更多地来自于网络。例如,出版社和数据库商提供的电子期刊、电子图书、各种数据库资源;企业和商业公司提供的搜索引擎、网络新闻、网络书籍、手机信息、网络音像等;政府部门发布的大量网上权威信息;教育和科研部门提供的开放获取活动、机构知识库建设提供的权威研究信息,还有各种机构网站和个人网站在互联网上提供的信息。随着不断的开发和传递,网络信息将越来越庞大,成为信息的海洋,不仅网络信息呈爆炸式发展,网络信息的用户也日趋增长。截至2013年6月,中国互联网络信息中心(CNNIC)公布《第32次中国互联网络发展状况统计报告》披露,中国的网民数已达5.91亿,普及率超过全球平均水平。毫无疑问,网络已成为信息传递和存储的重要载体。丰富的网络资源和便捷而超越时空的网络传递方式使人们更加依赖它,并把它当成获取信息的一大主要渠道和交流沟通的重要平台。因此,网络信息资源是网络时代人们生活中必不可少的信息资源。

网络信息资源是一切互联网中的可利用的电子化信息资源的总和。它不同于传统的文献信息资源,相比之下,网络信息资源呈现出其不同的特点。

① 信息数量庞大。每天政府、机构、企业、个人都在通过互联网发送各种信息,网络信息量在成几何级数增长,是其他任何信息资源所无法比拟的。

② 信息内容丰富。网络信息几乎涵盖各个领域,其形式也各种各样,如文字、声音、图像等,无奇不有。

③ 信息以数字化形式记录、存储和传递。这是网络信息资源的主要特征。网络信息通过磁介质、光介质存储和传送于计算机网络中,可以被反复查询和传递,并且在传递中没有损耗。

④ 信息传播速度快,共享程度高,交互性强。由于信息被转化为光信号通过光纤传播,所以信息通过网络传播的速度极快。网络用户可共享同一份信息,这使得信息能得到最广泛最便捷的利用。在网络上用户既是信息资源的享用者,又是信息资源的提供者,网络成为信息互动交流的平台。

⑤ 网上信息复杂多变。网络信息不仅在不断地增长,而且内容也不断被更新和淘汰,这使得网络信息资源复杂多变,极不稳定,处于一种动态中。

⑥ 网络信息良莠不齐。网络信息浩如烟海,鱼目混珠在所难免,要辨别信息质量高低、信息质量的真伪还需要人们的甄别和相关部门的认定。

与传统的信息资源相比,网络信息资源类型也更加丰富,常见的划分方法如下。

① 从信息内容来看,可以分为新闻信息、学术信息、教育信息、休闲娱乐信息、文学艺术信息、医疗保健信息等类型。

② 从表现形式来看,可包括文字、图片、音频、视频和多媒体等多种形式。

③ 按所对应的非网络信息的类型划分,可包括图书馆藏目录、电子书刊、参考工具书、数据库和其他类型的信息。

④ 按人类信息交流的方式,可分为非正式出版信息、半正式出版物、正式出版物。

⑤ 按信息存取方式,可分为邮件型、电话型、揭示板型、广播型、图书馆型、书目型。

⑥ 按网络信息资源的层次,可分为指示信息、信息单元、文献、信息资源、信息系统等。

第2章 信息检索基本知识

2.1 信息检索概论

2.1.1 信息检索的定义与特点

1. 信息检索的定义

信息检索的概念有狭义和广义之分。狭义的信息检索 (Information Retrieval) 是指依据一定的方法, 从已经组织好的大量有关文献集合中, 查找并获取特定的相关文献的过程。这里的文献集合, 不是通常所指的文献本身, 而是关于文献的信息或文献的线索。如果真正要获取文献中所记录的信息, 那么还要依据检索所取得的文献线索索取原文。广义的信息检索, 包括信息的存储和检索两个过程 (Storage and Retrieval)。信息存储包括以下 3 个步骤。

第一步: 信息的选择与收集。它是指检索系统根据本系统的服务目的, 确定信息收集、处理的原则, 对分布在各处的离散的信息进行收集加工。

第二步: 信息的标引。标引是信息加工人员对收集到的信息内容特征进行分析之后, 对每条信息加上系统能够识别的检索标识的过程。

第三步: 形成大量有序可检的信息集合。工作人员将标引后的信息条目录入, 并将其按照一定的顺序排列起来, 形成有序的信息集合——数据库, 从而为信息检索奠定基础。

信息的检索过程则是信息存储的逆过程。信息用户根据自己的需求对主题和概念进行认真分析后, 将自己的信息需求转化为检索表达式。该检索表达式与系统标识的比较匹配过程就是检索的过程。

简言之, 信息检索 (Information Retrieval) 是指将信息按一定的方式组织起来, 并根据信息用户的需要找出有关的信息的过程和技术。狭义的信息检索就是信息检索过程的后半部分, 即从信息集合中找出所需要的信息的过程, 也就是我们常说的信息查询 (Information Search 或 Information Seek)。对于信息用户来说, 信息检索仅指信息的查找过程。本书所涉及的信息检索概念也仅限于信息查找的概念。

2. 信息检索的特点

随着科学技术的发展, 尤其是计算机的应用, 信息检索从手工检索发展到计算机检索。

手工检索 (Manual Retrieval, 简称手检) 是一种传统的检索方法, 即以手工翻检的方式, 利用工具书 (包括图书、期刊、目录卡片等) 来检索信息的一种检索手段。检索过程是人的手工操作完成的, 其匹配是人脑的思考、比较和选择, 其最常见、最基本的方法是追溯法、工具法、混合交替法。

手工检索的方法比较简单、灵活,容易掌握。但是,手工检索费时、费力,特别是进行专题检索和回溯性检索时,需要翻检大量的检索工具反复查询,花费大量的人力和时间,而且很容易造成误检和漏检。

计算机检索(Computer-based Retrieval,机检)指人们在计算机或计算机检索网络的终端机上,使用特定的检索指令、检索词和检索策略,从计算机检索系统的数据库中检索出需要的信息,继而再由终端设备显示或打印的过程。它包括图书馆的公共书目(OPAC——Online Public Access atalogue)、各种联机数据库、Internet上的信息资源等。在检索过程中,人是整个检索方案的设计者和操纵者。

计算机检索是在计算机技术、通信技术和网络技术迅猛发展的基础上建立起来的,在信息服务领域具有划时代的意义。它产生于20世纪50年代,发展于20世纪80年代中期,20世纪90年代后随着国际互联网技术的发展而进入了一个崭新的时期,经历了脱机检索、联机检索、光盘检索、Web信息资源检索,其发展历程如下。

(1) 脱机检索(20世纪50年代末—20世纪60年代中期)

- 这时是计算机检索的原始时期。
- 只能进行简单的检索。
- 为满足专业检索人员定期批量处理用户的情报要求。
- 用户不能立刻获得检索结果。

(2) 联机检索(20世纪60年代末—20世纪70年代初)

1963—1964年,美国洛克希德导弹与宇航公司的情报实验室建立了“人—机”对话的联机情报检索系统(DIALOG的前身),此后在20世纪60年代末到20世纪70年代初联机检索系统得以快速发展。国际著名的DIALOG系统、ORBIT系统、MEDLINE系统都是从这个时期发展起来的。

- 用户可随时浏览检索结果。
- 由于这个阶段的计算机网络主要是通过电话线连接,因而联机检索受到地区的限制。

(3) 国际联机检索(20世纪70年代中期—)

卫星通信技术的出现,使得联机检索系统打破了地域限制。而数据库生产的迅速发展及微机大量的涌现,更使得国际联机检索蓬勃发展。

- 联机检索系统进入发展的黄金时期,实现了人类情报资源的共享。

(4) 单机光盘检索(20世纪80年代—)

CD-ROM技术促使计算机检索成本迅速下降(一张光盘可存储600MB机读数据、成本价格便宜,而一张DVD光盘的容量最少可达4.7G)。

(5) 光盘网络检索(20世纪90年代—)

- 光盘网络是一种计算机网络,如图书馆局域网。
- 实现多用户光盘资源共享。

(6) Web信息资源检索(20世纪90年代末—)

进入20世纪90年代后,随着网络技术的发展,尤其是互联网的迅猛发展,使计算机检索进入一个崭新的时期。检索方法更简单,检索结果更全面。

计算机检索的特点如下。

① 检索速度快。手工检索需要数日甚至数周的课题,计算机检索只需要数小时甚至几分钟就可以完成,大大地提高了文献信息的检索速度,节约了读者的检索时间,提高了检索效率。