

21

世纪高等院校教材

信息检索教程

张俊慧 主编



科学出版社

www.sciencep.com

21 世纪高等院校教材

信息检索教程

主编 张俊慧
参编 谭 英 杨淑琼 周建芳
尚 月 陈力行

科学出版社

北 京

内 容 简 介

本书是四川省高等院校图书馆、情报与文献学规划项目“面向信息时代的文献检索课教学改革研究”的研究成果。本书系统地介绍了信息检索的基础知识,对图书信息、期刊信息、特种文献信息、数据与事实信息及网络信息的基本检索方法与技术作了详细阐述。为了提高读者网络环境下的信息资源利用能力,加强实用性,本书还将几种常用的信息检索技巧及方法集中在“综合检索案例”中加以阐述。

本书可作为高等院校本科生、研究生的信息检索课教材,也可作为教学、科研及各类读者的参考书。

图书在版编目(CIP)数据

信息检索教程 / 张俊慧主编. —北京:科学出版社, 2010. 7

21 世纪高等院校教材

ISBN 978-7-03-027652-0

I. ①信… II. ①张… 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(2010)第 092341 号

责任编辑:王伟娟 张 宁 / 责任校对:包志虹

责任印制:张克忠 / 封面设计:耕者设计工作室

科学出版社出版

北京东黄城根北街 16 号

邮政编码: 100717

<http://www.sciencep.com>

骏杰印刷厂印刷

科学出版社发行 各地新华书店经销

*

2010 年 7 月第 一 版 开本: B5(720×1000)

2010 年 7 月第一次印刷 印张: 16 1/2

印数: 1—12 000 字数: 330 000

定价: 25.00 元

(如有印装质量问题,我社负责调换)

前 言

随着现代信息技术的快速发展,信息的数量急剧增长,信息的存储和传播方式发生了巨大的变革,信息的分散性、无序性也给人们利用信息增加了难度,影响了人们获取信息的质量与效率。于是,在网络环境下有效地检索和利用信息成为科研工作者,特别是高等院校学生必备的知识 and 技能。目前,大部分高校都将信息检索列为公共基础课程,目的是培养学生的信息素质,使学生在增强信息意识的基础上,熟悉检索工具和系统,掌握信息检索的方法和技巧,以提高学生的自学能力、研究问题和解决问题的能力。

本书是在总结多年的信息检索教学实践经验的基础上,根据当代信息检索与利用的发展现状与趋势,参考大量相关文献编写而成的,特别是参考了以前被选作教材的徐天秀研究馆员主编的《信息检索》一书中的有关内容。本书全面阐述了信息检索的定义、基本原理、检索语言、检索技术、检索策略;介绍了各类型文献信息的检索方法与技巧以及各种常用数据库;同时介绍了如何搜集、分析和利用信息,进行学术论文写作的基本知识。为加强本书的实用性和针对性,提高学生的检索技能,在最后一章“综合检索案例”中汇集了十余个不同类型不同检索需求的案例以供参考。

本书共9章,编写分工如下:第1章第1节由张俊慧编写、第1章第2节由陈力行编写,第2章第1、2、3、5、6节由张俊慧编写、第2章第4节由周建芳编写;第3章由陈力行编写;第4章由尚月编写;第5章由谭英编写;第6章由杨淑琼编写;第7章由周建芳编写;第8章由张俊慧编写;第9章由周建芳编写。全书的筹划、内容大纲和统稿等由张俊慧负责。

本书在编写过程中,得到了许多同仁特别是四川师范大学图书馆徐天秀研究馆员的热情指导和大力帮助,在此表示诚挚的谢意。同时还参阅了大量的相关文献,在此向书中提及或未曾提及的参考文献的作者表示感谢。

另外,虽然本书的编者均是长期从事信息检索课教学的专业人员,但由于时间仓促,加之编者水平有限,不足之处,敬请读者批评指正。

编 者

2010年5月

目 录

前言

| | |
|------------------------------|-----|
| 第 1 章 绪论 | 1 |
| 1.1 信息 | 1 |
| 1.2 信息素质 | 8 |
| 第 2 章 信息检索的基础知识 | 11 |
| 2.1 信息检索概述..... | 11 |
| 2.2 信息检索语言..... | 17 |
| 2.3 信息检索系统与检索工具..... | 27 |
| 2.4 信息检索技术..... | 32 |
| 2.5 信息检索策略..... | 38 |
| 2.6 信息检索效果评价..... | 41 |
| 第 3 章 图书信息检索 | 44 |
| 3.1 图书概述..... | 44 |
| 3.2 古籍图书的检索..... | 45 |
| 3.3 近代图书的检索..... | 55 |
| 3.4 现代图书的检索..... | 56 |
| 第 4 章 期刊信息检索 | 69 |
| 4.1 期刊概述..... | 69 |
| 4.2 中文期刊数据库..... | 71 |
| 4.3 外文期刊数据库..... | 85 |
| 4.4 开放存取期刊..... | 99 |
| 第 5 章 特种文献信息检索 | 102 |
| 5.1 学位论文 | 102 |
| 5.2 科技报告 | 113 |
| 5.3 专利文献 | 120 |
| 5.4 标准文献 | 128 |
| 5.5 会议文献 | 136 |
| 5.6 政府出版物 | 143 |
| 第 6 章 数据与事实信息检索 | 149 |
| 6.1 字典、词典..... | 149 |

| | | |
|------------|----------------------|------------|
| 6.2 | 百科全书 | 154 |
| 6.3 | 年鉴 | 160 |
| 6.4 | 名录 | 163 |
| 6.5 | 常用的事实数据库 | 169 |
| 第7章 | 网络信息检索 | 177 |
| 7.1 | 网络信息资源 | 177 |
| 7.2 | 网络信息检索 | 180 |
| 7.3 | 搜索引擎 | 183 |
| 7.4 | 网络导航 | 209 |
| 7.5 | 网络信息资源下载方法与技巧 | 214 |
| 第8章 | 学术论文的写作 | 217 |
| 8.1 | 学术论文概述 | 217 |
| 8.2 | 学术论文的基本格式和写作要求 | 219 |
| 8.3 | 学术论文的写作程序 | 223 |
| 第9章 | 综合检索案例 | 233 |
| | 参考文献 | 253 |

第 1 章 绪 论

1.1 信 息

1.1.1 信息、知识、文献

1. 信息

当今时代,信息(information)与我们的生活息息相关。我们在日常生活中每天都在与信息打交道,首先是通过感觉器官接受外界各种各样的信息,然后又通过语言、手势、文字、图画等多种手段向外传递、交流信息。“信息”这个词最早见于我国唐代诗人李中《碧云集·暮春怀故人》一诗中的“梦断美人沉信息,目穿长路倚楼台”,其中的“信息”即为消息、音信之意。随着科学技术的飞速发展,信息已成为当今使用频率最高、最时髦的词语之一,如信息资源、信息技术、信息产业、信息系统、信息经济、信息时代、信息社会等不胜枚举。但究竟“信息”是什么?学界至今没有一个统一的定论。

由于人们研究信息的角度与目的不同,所以提出的信息的定义也多种多样,概括起来有以下几个方面。

信息论的创立者香农(C. E. Shannon)着眼于对信息作用的描述,在其经典论著中给出定义:“信息是用来消除不确定性的东西。”

控制论的创立者维纳(N. Wiener)则着眼于信息的应用,他对信息的定义是“信息是我们用于适应外部世界,并且在使这种适应为外部世界所感知的过程中,同外部世界进行交换的内容的名称”。

《辞海》(1999年版)对信息的解释是: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,通常须通过处理和分析来提取。

《汉语大辞典》对信息的定义是:现代科学指事物发出的消息、指令、数据、符号所包含的内容。

我国多数学者认为,信息普遍存在于自然界与人类社会,是客观事物各种表现的反映。物体在运动过程中,伴随着信息的产生,信息借助于物质载体记录和媒体传播。

总之,信息是事物属性的再现。信息不是事物本身,而是事物发出的消息、指令、数据等所包含的内容。

现代社会被称为信息化的社会,信息与材料、能源一起被视为社会经济发展的三大支柱,信息已成为促进社会经济发展的重要战略资源。

2. 知识

知识(knowledge)是人类社会实践的总结,是人的主观世界对客观世界的概括和如实反映。知识是人类对自然界、人类社会以及思维方式与运动规律的认识,再通过大脑的重新组合的系统化的信息集合。因此人类不仅通过信息感知世界、认识世界和改造世界,而且根据所获得的信息组织成多种多样的知识。由此可见,人类生活环境中普遍存在的信息是知识的来源,信息在被接受、选择、处理之后就转化成了知识,知识是一种特定的信息。

知识存在的形式基本上有两种:一是主观知识,指存在于大脑记忆中的知识;二是客观知识,指用文字、图形、符号、视频、声频等技术手段记录在一定载体上的知识,通常人们把这类知识称为文献信息。

3. 文献

“文献”(document)一词最早见于《论语·八佾》中,一直沿用至今。《中华人民共和国国家标准文献著录总则》(GB3792.1—83)对文献的解释是:“文献是记录有知识的一切载体。”即用文字、图像、符号、声频、视频等手段记录人类知识的各种载体。所谓载体不仅包括印刷类出版物,还包括声像制品和电子出版物。因此文献的范围非常广泛,古代刻在龟甲和兽骨上的甲骨文是文献,在石碑上刻有文字和图形的碑文是文献,青铜器上的铭文是文献,竹简和帛书是文献,纸质出版物是文献,今天的磁带、磁盘、光盘、缩微胶片等都是文献。

文献具有两个基本的要素:一是有一定的知识内容;二是一种物质载体。文献是记录知识的物质形式,也是用于传递知识的工具。正是文献的这些特点使人类的知识得以保存和传播,人类的科学技术和文化得到继承和发展。

4. 信息、知识、文献三者之间的关系

信息、知识与文献三者之间既有联系又有区别。信息的内涵范围很广,物质的运动产生了信息;各种信息经过人们的加工处理、系统化之后,信息就转化成知识。知识是从信息积累中找出的有规律性的东西,知识是信息的组成部分。特定的知识应用于实践,解决实践中存在的问题,创造物质财富或精神财富,这样知识便转化为生产力,产生新的信息,这就形成一个循环过程。而文献是记载知识的载体,它是知识存在的基本形式。因此,知识和文献是信息的组成部分。

1.1.2 信息的属性

信息的主要属性有客观性、时效性、传递性、共享性。

1. 客观性

信息是客观存在的,是现实世界中各种事物运动与状态的反映。信息的产生来源于物质,信息产生后又必须依附于物质,因此信息包含在物质之中。

2. 时效性

信息的时效性是指信息从发出到接受的时间间隔及其效率。随着时间的变化,信息的功能作用会发生变化,其新颖性也会逐渐消失。在现代社会中,信息的时效性决定信息的价值,因为信息是活跃的、不断变化的,及时地获取和运用有效的信息将获得信息的最佳价值,如时效性强的天气预报、经济信息、交易信息等应及时传递,以便开发利用。

3. 传递性

信息的传递性是指信息从信息源出发,通过信息载体的传递才能被接收并进行处理和利用,也就是说信息可以在时间、空间上进行传播和交流。信息在时间上的传播,使我们了解了历史,继承前人研究的成果;信息在空间上的传播,使人类智慧和文明得以交流。特别是随着现代科学技术的发展,信息的传递变得越来越方便快捷。

4. 共享性

信息的共享性是指同一内容的信息在同一时间或不同时间能被多个用户共同使用。信息的提供者不会因为提供了信息而使原有的信息量减少,用户分享的信息量也不会因为分享人数或分享次数的多少受到影响。因为信息在传递、扩散、使用的过程中是不会减少的,只有实现信息资源的共享才能充分发挥信息的作用。

1.1.3 信息源

信息源即用户获取信息的来源,是信息产生的源头。可以这样说,一切信息来源于自然界,来源于人类社会。人们的生产实践、日常生活以及政治、经济、军事、科研等活动都是信息产生的源泉。

根据信息的载体和表达方式的不同,信息源可分为语言信息源、实物信息源和文献信息源。在这三种信息源中,文献信息源是检索与利用的主体。随着计算机技术和网络技术的发展,数字型文献信息源有了很大的发展,数据库、网络信息已

成为重要的文献信息源。

1. 语言信息源

语言信息源,即通过语言进行传播和交流的各种信息,在情报学领域称之为“零次信息”。其特点是传递迅速、互动性强、主观随意性强,其可信度得不到保证。语言信息源是一种不可忽视的信息源,通过语言交流如聊天、授课、讨论等方式获得的信息可能包含较高价值,有时能达到提供新思路、新设想、解决难题的目的,但许多信息并不十分准确与可靠。

2. 实物信息源

实物信息源是指某种实际存在的物品所揭示出来的信息,如产品样本、样机、模型、雕塑等实物形式表示的信息。其特点是直观性强、可信度高、信息量大,但需要通过观察、分析、研究才能揭示出其隐含的大量信息。

3. 文献信息源

文献信息源是指以文字、图形、符号、声频、视频等记录手段,将系统化的信息内容存储在各种载体上而形成的一种信息源。它包括印刷型、缩微型、数字型和声像型信息源。这类信息经过加工、整理,较为系统、准确、可靠,便于保存和利用,是目前数量最多、利用率最高的信息源。

按照划分标准的不同,文献信息源可划分为不同的类型。按文献信息的加工程度划分,有一次信息、二次信息、三次信息;按文献信息载体的物理类型划分,有印刷型、缩微型、声像型、数字型;按文献信息的出版类型划分,有图书、期刊、报纸、学位论文、科技报告、会议文献、专利文献、标准文献、产品资料、政府出版物、档案文献等。

1) 按文献信息的加工程度划分

(1) 一次信息(primary information)。一次信息也称原始信息,即第一次书写、第一次报道、第一次出版的信息。一次信息是以人们的社会实践和科学研究成果为依据而创作出的信息,它的载体形式有图书、期刊论文、科技报告、会议论文、专利说明书以及不宜公开发表的科技日记、杂记、实践记录等。一次信息也称为第一手资料,信息内容具体,具有新颖性、创造性、系统性等特点,它的数量十分庞大,常常需要通过利用二次信息才能获得相关信息。

(2) 二次信息(secondary information)。二次信息是把大量分散无序的一次信息依一定的方法和原则进行加工提炼、浓缩而形成的信息,其目的是为了有效地管理和利用一次信息。二次信息是查找一次信息的工具,它的形式包括目录、题录、文摘、索引、书目等。二次信息的特点是浓缩性、汇集性、有序性,对一次信息有

指引和报道的作用。

(3) 三次信息(tertiary information)。三次信息是指在对一次信息和二次信息进行广泛深入的研究之后,通过汇集、综合、分析等深度加工而形成的信息,它的形式有百科全书、年鉴、指南以及综述、述评、进展报告等,还包括指引利用二次信息的书目之书目、文献指南等,其特点是综合性、针对性和科学性。三次信息可以帮助我们检索各种“事实”或“数据”。

从一次信息到三次信息,是为了帮助人们管理和有效地利用信息,对信息进行不同程度的加工而形成的信息,实际上是把分散的信息集中起来,使无序的信息有序化。掌握了信息的不同级别,可以帮助人们有效地利用信息。

2) 按文献信息载体的物理类型划分

为了有效地存储、传播知识,人类先后发明了各种各样的物质材料来记录信息,根据物质载体形式的不同,文献信息可分为以下几种类型。

(1) 印刷型。印刷型也称纸介质型,它是以纸质材料为载体,以印刷为记录手段而形成的文献形式。例如,各种纸质图书、期刊、报纸、学位论文、专利以及标准等。它是目前整个文献中的主体,也是有着悠久历史的传统文献形式。其优点是便于阅读和使用,不需要特殊的设备;缺点是存储密度小,体积大,不利于保管。

(2) 缩微型。缩微型信息是以感光材料为载体,利用光学技术将文字、图形、影像等按比例缩小而产生的一种文献形式,如缩微胶片、缩微胶卷、缩微平片等。其优点是体积小、容量大、成本低、便于保存;缺点是阅读不方便,需要借助专门的阅读机。目前在整个文献中,缩微型文献数量较少。

(3) 声像型。声像型信息又称视听资料或声像资料,是采用特定的设备,以磁性和感光材料为介质记录声音、图像等信息的一种文献形式,如录音带、录像带、唱片、幻灯片、影视片等,其特点是能给人以直观形象的感受。

(4) 数字型。也称电子型,是指以数字代码方式将图形、文字、声音、影像等信息存储到磁、光、电等介质上,通过计算机或类似设备阅读和使用的一种文献形式,如电子图书、电子期刊、各种数据库、网络信息资源等,其特点是存储容量大、传播更新速度快、检索方便。数字型文献是当今重要的文献信息源。

3) 按文献信息的出版类型划分

根据出版形式的不同,文献信息可分为图书、期刊、报纸、学位论文、科技报告、会议文献、专利文献、标准文献、产品资料、政府出版物、档案文献等。

(1) 图书(book)。图书是对某一学科或知识领域进行系统论述或介绍的一种出版物。图书包括专著、教科书、科普读物、参考书、工具书等。图书往往是作者长期研究的成果和知识的积累,是传播知识、教育和培养人才的主要文献类型。

图书的特点是内容比较成熟、系统、全面、可靠,但出版周期较长,报道速度相对较慢,一般不能反映最新的信息,时效性差。近年来,电子图书大量面世,且数量

越来越多,内容十分丰富,通过计算机就能方便快捷地获得相关图书。

图书的著录特征一般有书名、著者、出版地、出版社、出版时间、价格、页码、国际标准书号(ISBN)等。

(2) 期刊(journal/magazine/periodical)。期刊又称连续出版物,是一种有固定名称、定期或不定期连续刊行,每期载有不同著者、译者或编者所编写的文章,用连续的卷期和年月顺序编号出版,每期的内容不重复。与图书相比,期刊出版发行速度快,内容新颖。

期刊是重要的信息来源,经常查阅期刊,可以了解科研动态、前沿信息和进展情况,以拓宽思路,汲取有用的成果。据统计,科研人员所获取信息的65%以上来源于期刊,其中科技期刊的利用率最高。

期刊的特点是:品种繁多,内容丰富多样,出版周期短,报道及时、迅速,流通面广,连续性强,能及时反映国内外各学科领域的发展动态。有的期刊还登有文獻述评、动态介绍、学术协会通告、书评和商品广告等。期刊全文数据库为期刊的利用带来极大的方便,受到越来越多用户的欢迎。期刊的著录特征有期刊名称、期刊出版的年卷期、国际标准刊号(ISSN)等。

(3) 报纸(newspaper)。报纸是出版周期最短的定期连续出版物。及时性、新闻性是报纸的主要特征,发生的事件有的仅几个小时就能见诸报端。报纸的内容十分丰富,能及时反映社会政治、经济、文化、科技等各个方面的信息。报纸的数量十分庞大,给查阅资料带来困难,因此,一般需利用相关索引,电子版报纸的面世为查找报纸信息带来了极大的方便。报纸的特点是报道及时、时效性强、涉及面广、影响大,其缺点是知识不系统、信息分布零乱,是一种难以保存和积累的文献信息。

(4) 学位论文(thesis dissertation)。学位论文是高校或科研机构的毕业生为获取某种学位而撰写的论文,包括学士论文、硕士论文和博士论文。

学位论文,尤其是高层次的学位论文质量较高,带有一定的独创性,探讨的问题比较专深,对问题的阐述较为系统详细,其参考价值较高,是一种重要的信息源。学位论文属非卖品,一般不对外发行,这就造成利用纸本学位论文的不便。但目前已有不少博硕论文数据库面世,要利用这些信息已不再困难。学位论文的著录特征有论文名称、作者、学位名称、导师姓名、学位授予机构等。

(5) 科技报告(sciectific report)。科技报告是指国家政府部门或科研单位关于某项研究的总结报告或阶段性进展报告。科技报告一般都有编号,供识别报告类型使用。科技报告通常以正式报告、进展报告、技术札记和备忘录等形式发表。有的报告涉及尖端技术或国防技术等保密问题,所以又分为绝密、秘密、内部限制发行和公开发布等几个等级。

科技报告的特点是内容新颖、专深、详细、可靠,是科研过程的真实记录,每份报告单独成册,统一编号,有严格的发行范围。科技报告能反映一个国家在某一学

科领域的科技水平,许多尖端学科的研究信息首先就反映在科技报告中,所以它是一种重要的信息源。

目前国际上较著名的科技报告有:美国政府的四大报告,即行政系统的 PB (Publication Board)报告、军事系统的 AD(ASTIA Documents)报告、原子能与能源管理系统的 DOE(Department of Energy)报告、航空与宇航系统的 NASA(National Aeronautics and Space Administration)报告;英国航空委员会的 ARC 报告;法国原子能委员会的 CEA 报告;德国航空研究所的 DVR 报告等。

科技报告的著录特征有报告名称、报告号、报告入藏号、报告完成单位、完成时间等。

(6) 会议文献(conference paper)。会议文献是指在国际或各国国内各种专业学术会议和科学技术研究会议上宣读的论文和报告。按出版时间的先后分为会前文献、会期文献、会后文献;按会议召开的范围分为国际会议文献、全国性会议文献、地区会议文献。

会议文献学术性强,往往代表着某一领域当时的最新成果和信息,是一种重要的信息源,但其内容与期刊相比可能不太成熟。会议文献的著录特征有论文名称、作者、会议名称、会议召开时间地点等。

(7) 专利文献(patent literature)。专利文献是指专利形成过程中产生的一系列官方文件及出版物的总称。它包括专利说明书、专利公报、专利分类表、专利文摘、专利检索工具以及与专利有关的法律文件及诉讼资料等。但通常所说的专利文献是指各国专利局公布出版的专利说明书,它是专利文献的主体,是一种受法律保护的文献,也是科研人员检索的主要对象。专利说明书是专利申请人向专利局递交的有关发明、目的、构成和效果的技术文件,主要涉及发明的技术内容和权利要求。专利说明书的内容具体、可靠、详细,具有新颖性、实用性和创新性。专利文献的著录特征有专利号、专利名称、国际专利分类号、专利公布日期等。其中专利号按国际规定由两个字母表示的国家名称和顺序号组成,如 US7371568、CN1016502 等。

(8) 标准文献(standard literature)。标准文献是对工农业产品和工程建设的质量、规格及其检验方法等方面所作的技术规定。标准文献作为一种规章性文献,具有一定的法律约束力。一个国家的标准文献反映了该国的生产工艺水平和技术经济政策,而国际现行标准则反映了当前世界科学技术水平。按照使用范围,标准通常分为国际标准(如国际标准化组织制定的 ISO 标准、国际电工委员会制定的 IES 标准)、地区标准(如欧洲标准委员会制定的 CEN 标准)、国家标准(如美国的 ANSI、法国的 NF、日本的 JIS、中国的 GB 等标准)、部颁标准等。标准文献的著录特征有标准颁布国家(组织)、标准名称、标准号、颁布时间等。其中标准号由国别(组织)代码+顺序号+年代号组成,如 ISO3256—1986,GB17930—1999 等。

(9) 产品资料(product literature)。产品资料是指各国厂商或经销商为介绍或推销产品而印发的商业宣传品。产品资料包括产品样本、产品目录、产品说明书、厂商介绍、厂刊或贸易刊物、技术座谈资料等。产品样本是对定型产品的性能、构造、原理、用途、使用方法、操作规程、产品规格等所作的具体说明,其数据可靠,技术成熟,有详细的外观照片和结构图,直观性强,故是查阅和分析国内外产品的重要资料。产品资料的著录特征有公司名称、产品名称、内容和厂家联系方式。

(10) 政府出版物(government publication)。政府出版物是指各国政府部门以及所属机构发表、出版的文献。政府出版物的内容十分广泛,按其性质通常分为两大类:一类是行政性文献,如国会记录、立法司法资料、方针、政策、规章制度、决议以及调查统计资料等;另一类是科技文献,主要指政府部门出版的科技政策文件、科技档案、经济规划等。政府出版物是了解各个国家政治、经济、科技发展政策的重要信息源。

(11) 档案文献(archives)。档案文献是指国家机关、社会团体及组织在从事政治、军事、经济、科技、文化、宗教等活动以及个人从事社会活动中所形成的具有保存价值的各种形式的历史记录。档案材料具有原始性特点,其客观真实地反映了历史。档案文献的可靠性和稀有性使其具有特殊的价值,但一般为内部使用,不公开出版发行,有的档案有密级限制。

1.2 信息素质

1.2.1 信息素质概述

1. 信息素质的定义

信息素质(information literacy),又称信息素养、信息教养、信息鉴别能力、信息文化。我国台湾地区则称之为资讯素养。信息素质的概念是从图书馆素质(library literacy)等概念发展演化而来的,最早由美国信息产业协会(IIA)主席保罗·泽考斯基(Paul Zurkowski)在1974年提交的一份报告中提出。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。他们掌握了利用大量的信息工具及主要信息源使问题得到解答的技术和技能”。经过5年的研究,IIA于1979年更新了该定义,认为具备信息素质的人即“掌握了信息工具利用的知识与技能,并能够将其应用于解决实际问题的人”。20世纪80年代,信息素质的内涵得到了进一步扩展和明确,不仅包括各种信息技术和技能,而且涉及个体对待信息的态度,如信息需求意识、确定与利用信息的愿望、对信息价值的评价与判断、对信息合理与准确的利用、对信息的接受与评估等。

进入90年代后,随着计算机网络等技术的飞速发展和广泛应用,信息素质的

内涵也变得更加深刻。2003年9月,联合国信息素质专家会议发表了《布拉格宣言:走向信息素质社会》。会议宣布:信息素质是终身学习的一种基本人权。信息素质是人们投身信息社会的一个先决条件。信息素质可以广义地理解为是个体在先天所赋予的生理素质基础上,通过后天学习实践形成的信息品质,它是社会成员的信息意识、信息道德、信息能力、信息潜能等多项基本素质的有机结合。信息素质不仅仅是诸如信息的获取、检索、分析、交流等技能,而且包括以独立学习的态度和方法,将已获得的信息用于信息问题的解决,进行创新性思维的复合的信息能力。信息素质能够引发、保持和延伸终身学习,这是通过技术的利用实现的,但又是独立于技术之外的一种基本素质。

2. 信息素质的内涵

信息素质的内涵是比较广泛的,综合现有研究对信息素质概念的理解,信息素质的内涵主要包括三个方面的内容:信息意识素质、信息能力素质和信息道德素质。

1) 信息意识素质

“意识”是指人的头脑对于客观世界的反应,信息意识是指人们对信息需求的自我意识,即人们能从信息角度出发,去感受、理解和评价自然界和社会中的各种现象、行为与理论,并具有捕捉、判断有用信息的能力。它是指人们对信息需求的自我感悟,即人们对信息的捕捉、分析、判断和吸收的自觉程度,包括信息主体意识、信息传播意识、信息保密意识及信息更新意识等多种形式。信息意识表现了人们对信息的敏感性和重视程度。在当今信息大爆炸的时代,人们获取知识的愿望与能力显得非常重要。

2) 信息能力素质

所谓信息能力是指人们对信息确认、评价、加工、利用、交流的综合能力。它是信息素质的重要组成部分,主要包括:广泛了解信息的性质;存取和检索信息的能力;分析和综合信息的能力;评价信息的权威性、可靠性和有效性的能力;能反复修改检索策略直到问题得到解决;能够采用各种技术手段交流,等等。面对成千上万的信息,有了信息意识、欲望、积极性,而没有信息能力,是不可能很好地利用信息的。信息社会为我们每个人都提供了大量的信息资源和共享机会。这些数量庞大的信息也给我们带来了麻烦,社会信息量剧增,混杂无序、良莠不齐、交叉重复,涉及众多学科和领域,这就需要我们了解各种信息来源,掌握一定的信息检索技术。对快速获取的大量信息,还要具备选择、分析、利用信息的能力,从众多信息中选择相关信息,从相关信息中选择有用信息,从有用信息中选择核心信息、增值信息,才能掌握信息。

3) 信息道德素质

信息道德是指整个信息活动中的道德规范,是调节信息创造者、信息服务者、信息使用者之间相互关系的行为规范的总和。它的目的是促使社会个体成员遵循一定的信息伦理与道德准则来规范自身的信息活动行为。信息道德的基本内容包括人们的信息信念、信息态度、信息行为义务、信息纪律和信息良心等。信息道德规范信息行为,主要包括:不制作、传播、消费不良信息;不侵犯他人的知识产权、商业秘密、隐私权,正确处理信息创造、信息服务、信息使用三者之间的关系;恰当使用与合理开发信息技术等。

一定社会的道德取决于一定社会的现实物质生活条件,社会生活条件的不同,人们的精神追求往往会有差异。伴随信息经济而出现的信息道德,是人类在信息技术演化中形成的一种本能上的自我要求和自我限制。

由于信息技术的高度发达,当今社会不良信息泛滥,信息侵权、信息犯罪时有发生,信息环境正遭到严重污染。因此在未来的信息社会中,每一个人都应遵循一定的信息伦理与道德准则来规范自身的信息行为。在目前我国网络立法方面尚存许多空白、网络伦理弱化的情况下,信息道德素质显得尤为重要。

1.2.2 大学生信息素质教育的意义

信息素质教育为大学生获取终身学习的能力奠定基础。在知识经济时代,作为第一生产要素的知识的增长速度非常快,由此导致的新旧知识的更替速度也非常快。据统计,现代社会知识的半衰期已缩短至5~7年。从知识的生产和老化状况看,近50年来,人类获得的知识等于过去两千年的总和,今天的知识到2050年仅为届时总量的1%,99%的知识是今后才创新的。这就意味着我们必须不断学习,获取新知识,才能跟上时代步伐。学习成为人类生存的一种方式,大学生要学会生存的本领,要学会终身学习,首先就是要具有自我学习的能力。因为除了通过学校教育所能学习到的知识外,还有很多的知识,而这些知识大多是需要终身学习的过程中通过自我学习获得的。终身学习的目的就变得不仅在于获得已有的知识、经验,其更高的目标在于发现、吸收知识和信息以创造新的知识信息。于是,能够在不同的社会环境下敏锐地捕捉所需信息,主动、有效地检索和吸收信息的能力就显得尤为重要。

通过信息素质教育可以使大学生在研究性学习和创新活动中,随时收集和获取相关信息,并及时掌握相关研究的现状和发展方向,在科研中少走弯路,最终取得创新性成果。通过信息素质教育可以培养大学生正确的信息道德观,以及与知识产权相关的著作权法、专利法等法制观念,增强信息安全意识,树立信息道德观来规范自身信息行为;通过信息素质教育,可以使大学生在学习上拥有更大的自主权,并赋予大学生自我学习的能力,为其终身学习打下基础。

第 2 章 信息检索的基础知识

2.1 信息检索概述

2.1.1 信息检索的定义

信息检索(information retrieval)是指从众多的信息集合中迅速准确地查找出用户所需信息的全过程。信息集合是指有组织的信息资源整体,可以是馆藏目录、数据库、印刷本检索工具等。广义的信息检索又叫信息的存储与检索,包含两个方面的含义:一是信息的存储,即将信息按一定的方式组织和存储起来;二是信息的检索,即根据用户的需要找出有关的信息资料的过程。狭义的信息检索仅指该过程的后半部分,即从信息集合中找出所需信息的全过程。

信息作为一种重要的资源,它在人类经济生活和社会交往中的地位越来越重要,通过检索获取信息逐渐成为人们进行教学、科研、生产、贸易和社会交往的一种基本技能。因此,信息检索的目的是为了解决庞大的信息资源与人们特定需求之间的矛盾。

2.1.2 信息检索的原理

人们的信息检索需求是多种多样的,获取信息的方法也各不相同,但信息检索的原理却是相同的。简单地说,信息检索的原理就是将检索提问标识与存储在检索系统中的信息标引标识进行比较,两者一致或信息标引标识包含着检索提问标识,则具有该标识的信息就从检索系统中输出,输出的信息就是检索找到的信息。

根据广义信息检索的含义,信息检索的全过程包含存储和检索两个过程。信息的存储就是信息标引人员将收集到的大量的、分散无序的原始信息进行分析,找出能代表文献信息的特征,对文献信息进行标引,按照检索语言规定的原则和方法将信息的外表特征和内容特征(如题名、著者、主题等)转化为一定的信息标引标识(如分类号、主题词等),再将这些标识按一定的顺序编排后存储在检索系统(工具)中,并提供多种检索途径;信息的检索则是检索人员对检索课题进行全面分析,找到其特征,形成检索提问,根据检索语言规定的格式和要求编制检索提问,形成检索提问标识,根据存储所提供的检索途径,将检索提问标识与存储在检索系统(工具)中的信息标引标识进行比较,与检索提问标识相符的信息作为检索结果从检索系统中输出,如图 2-1 所示。