



普通高等教育“十二五”规划教材

信息检索

XINXI JIANSUO (第二版)

戴建陆 张 岚 主编



中国电力出版社
CHINA ELECTRIC POWER PRESS



普通高等教育“十二五”规划教材

信息检索

XINXI JIANSUO (第二版)

主 编 戴建陆 张 岚
编 写 李宏宇 韩凤伟
主 审 王 宇



中国电力出版社
CHINA ELECTRIC POWER PRESS

内 容 提 要

本书为普通高等教育“十二五”规划教材。

本书注重理论与实践相结合,系统阐述了信息检索的基本原理、检索途径和检索方法,详细介绍了图书馆文献分类与书目检索的相关知识;相对于第一版,删减了部分利用率不高或停止出版的手工目录、文摘型检索工具内容;重点讲述了国内外重要全文数据库资源与互联网信息资源的检索方法与技巧,图文并茂、例证丰富、内容明晰,以中文检索工具为主,兼顾部分外文检索工具,另外对互联网上部分免费开放存取的学术信息资源进行了介绍。

本书可作为普通高等院校公共基础课教材,也可作为广大教师、科研人员、工程技术人员等在职人员参考用书。

图书在版编目(CIP)数据

信息检索/戴建陆,张岚主编.—2版.—北京:中国电力出版社,2012.4

普通高等教育“十二五”规划教材

ISBN 978-7-5123-2971-3

I. ①信… II. ①戴…②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(2012)第079036号

中国电力出版社出版、发行

(北京市东城区北京站西街19号 100005 <http://www.cepp.sgcc.com.cn>)

北京丰源印刷厂印刷

各地新华书店经售

*

2009年2月第一版

2012年7月第二版 2012年7月北京第四次印刷

787毫米×1092毫米 16开本 16.75印张 408千字

定价 29.50元

敬告读者

本书封面贴有防伪标签,刮开涂层可查询真伪

本书如有印装质量问题,我社发行部负责退换

版权专有 翻印必究

前言

“信息检索”是一门实践性很强的素质教育课,在普通高等学校已开设多年,该课程的主要教学目标在于培养大学生具有良好的信息素质,使其在校学习期间能够较快掌握文献信息检索的基本知识,了解一些常见检索工具的检索方法和使用技巧,从而能够充分利用各类文献信息资源为自己的学习提供帮助。

高等教育的任务不仅是给学生传授知识,更重要的是对学生进行全面素质的培养。大学阶段,学生的课余时间、自主学习时间更多,课堂上的许多知识、问题需要课后通过相关文献资料的阅读来深入理解、消化吸收。大学生除学好本专业知识外,还应广泛涉猎其他专业的相关知识,而这一切都离不开文献信息的查找与选择。通过“信息检索”课的学习,在拓宽学生获取知识渠道、增强学生获取信息能力的同时,也会强化学生探求新知的乐趣,提高学生学习新知的效率,对学生未来的终身教育与学习均有帮助。

本教材来源于几位编者多年的教学实践,在第一版的基础上,对部分检索工具检索功能的改进及检索页面的变化进行了修订,删除了部分利用率不高或已停止出版的手工目录、文摘型检索工具内容。为突出信息检索的时效性和针对性,修改了部分检索实例,对网络信息检索工具及网络开放存取资源链接地址进行全面审校。本书内容以介绍各种常用数据库检索工具与互联网信息检索工具为主,各类检索工具的介绍,以中文检索工具为重点,兼顾部分外文检索工具。

本教材理论与实践相结合,在满足学生掌握常用检索工具使用方法的同时,使其能够概括了解信息存储与检索的基本原理与技术,了解信息检索的相关知识,为未来部分学生求学深造、从事科学研究和工程实践的高层次信息检索需求提供入门导引。本书的检索实践部分,力求清晰易懂,读者通过书中大量图例,对照相应的检索工具实际操作,会很容易掌握其检索方法。

本教材由沈阳工程学院戴建陆、张岚主编,李宏宇、韩凤伟编写。戴建陆编写了第1~4章、第8章、第9章9.1和9.2节,张岚编写了第5~7章及附录部分,韩凤伟编写了第9章9.3和9.4节,李宏宇编写了第10章。全书由戴建陆统稿。

本教材由沈阳师范大学王宇主审。在编写过程中还参考了大量专家学者的论著、成果。在此一并致谢。

由于编者水平有限,书中难免会有疏漏、不足之处,恳请同行专家、读者批评指正。

编者

2012年3月

目 录

前言

第1章 概述	1
1.1 学习信息检索的目的和意义	1
1.2 信息检索基本知识	3
1.3 文献信息资源概述	8
第2章 图书馆与馆藏文献检索	18
2.1 图书馆的发展与现代图书馆	18
2.2 图书馆文献分类	21
2.3 图书馆目录与馆藏文献检索	27
第3章 信息检索的基本原理与方法	36
3.1 信息检索语言	36
3.2 信息检索工具	44
3.3 检索途径、检索方法和检索策略	47
第4章 中外文索引、文摘型检索工具	55
4.1 《全国报刊索引》	55
4.2 《电子科技文摘》	60
4.3 《中文社会科学引文索引》(CSSCI)	61
4.4 美国《工程索引》(EI)	64
4.5 英国《科学文摘》(SA)	72
第5章 计算机与互联网信息检索	80
5.1 计算机信息检索基础	80
5.2 Internet 概述	86
5.3 网络信息资源检索	92
5.4 各类信息网站	110
第6章 常用中文全文数据资源系统	121
6.1 中国知网(CNKI)中国知识资源总库	121
6.2 维普资讯数据系统	133
6.3 万方数据知识服务平台	135
6.4 超星数字图书馆	138
6.5 书生之家数字图书馆	139
6.6 方正阿帕比(Apabi)电子图书	141
6.7 中国资讯行	141

6.8	中国经济信息网	144
第7章	常用外文全文数据库系统.....	146
7.1	EBSCOhost 西文数据库	146
7.2	SpringerLink 数据库系统	151
7.3	LexisNexis 数据库系统	154
7.4	WSN 全文数据库系统	159
7.5	ScienceDirect 数据库系统	163
7.6	OCLC 数据库系统	168
7.7	IEEE/IET Electronic Library	172
第8章	专利文献检索.....	178
8.1	专利概述	178
8.2	中国专利检索系统	185
8.3	国外专利检索系统	201
第9章	其他特种文献检索.....	210
9.1	标准文献检索	210
9.2	会议文献检索	223
9.3	学位论文检索	229
9.4	科技报告检索	235
第10章	数据事实检索与工具书的使用	239
10.1	工具书概述.....	239
10.2	辞书.....	241
10.3	百科全书.....	244
10.4	年鉴.....	246
10.5	手册.....	249
10.6	名录.....	250
10.7	工具书指南.....	251
附录	各类信息资源网址.....	253
参考文献		261

第1章 概述

1.1 学习信息检索的目的和意义

“材料”、“能源”、“信息”是构成现代文明社会的三大支柱。人类进入文明社会以来,先后经历了游牧时代、农业时代和工业时代。在农业时代,主要劳动工具是简单农具,材料为铁和竹木,动力为人和牲畜。18世纪以来,蒸汽机和纺织机的出现标志着人类开始走入工业时代。在工业时代,由于动力和机械的应用,劳动生产力得到较大提高,在工业社会200余年的历程中,产生了几次技术革命,如化工技术革命、电力技术革命、机械技术革命,使社会大大向前发展。20世纪后半叶,人类社会爆发了历史上的又一次新的技术革命——信息技术革命。信息技术革命是继农业革命、工业革命之后的第三次产业革命,它正在引发人类历史上的又一次飞跃。

信息技术革命带来了社会的巨大变革。既非物质又非能量的信息,成为越来越多劳动者的劳动对象。人类社会的劳动成果也不再仅仅是农产品和传统工业品,还有大量涌现的信息内容产品以及信息技术产品,包括硬件产品和软件产品。围绕这些产品,信息服务和信息技术服务也应运而生。所有这些经济活动已经形成了新的、庞大的产业群体——信息产业。一些发达国家的统计表明,信息产业对国民经济增长的贡献率已经超过传统的第一、第二、第三产业,大量劳动者进入信息行业,有别于农业经济、工业经济的现代经济体系——信息经济全面走向成熟。与经济领域发生的巨大变革相对应,人们的社会活动和生活方式也在发生变化,社会管理的手段也出现了新的形式,信息社会已经到来。

1.1.1 信息社会主要特征

关于“信息社会”的显著特征,20世纪80年代较为流行的说法是“3C”社会(通信化、计算机化和自动控制化)、“3A”社会(工厂自动化、办公室自动化、家庭自动化)和“4A”社会(“3A”加上农业自动化)。20世纪90年代以后,关于信息社会的说法又增加了多媒体技术和信息高速公路网络在经济和社会生活中的普遍采用等条件。具体而言,信息社会还应具备几种主要特征,如劳动力结构中信息职业的人数占绝对优势,国民生产总值中信息经济占有较大比重,社会生活上拥有覆盖面极广的远程快速通信网络系统,生活文化模式多样化,社会观念上尊重知识的价值观念成为社会之风尚等。在信息利用与传递方面,信息社会的特征具体表现为如下几方面。

1. 信息传递方便快捷

信息时代,人与人之间的信息交流越来越方便快捷,广播电视、新闻出版、互联网成为公共信息传播的主要媒介。利用手机、电话、网络、寻呼等工具,个人之间的信息传播与交流已经摆脱了时空障碍。在光纤通信、卫星通信、移动通信、微波通信等现代通信技术支持下的全球高速互联网建设,极大地改变了人们现有的生活方式,并带来巨大经济与社会效益,现代信息处理技术使海量信息数字化成为可能,互联网已经成为信息传播与获取的主要媒介,全球范围内信息资源共享得以实现。

2. 信息爆炸

人类社会进入 21 世纪,科学技术以前所未有的高速向前发展,以生命科学、材料科学、计算机科学、新能源科学、空间科学以及大量新兴学科为代表的科学技术发展日新月异。社会的发展与科技的进步,带来记载科学成果与社会政治、经济、文化发展水平的各类文献生产量呈几何级迅速增长。

现代技术提高了信息的加工、存储与传播能力,但随之带来的问题和副作用是:面对汹涌而至、浩如烟海的文献信息,人们越来越感到无所适从,从信息海洋中迅速而准确地获取自己最需要的信息,变得非常困难,许多有用的信息还未来得及被发现和利用就自生自灭了。信息的增长和人们利用信息的能力两者之间出现越来越大的差距,面对这种现象,人们称为“信息危机”或“信息爆炸”。在这种形势下,信息检索作为一个专门的研究领域应运而生。

3. 信息混乱与信息烟尘

信息爆炸为人们带来了内容丰富、形式复杂、载体多样的海量文献信息,大到美国火星探测的技术资料、勇气号火星车的设计图纸,小到一个本科生的毕业论文,都有相应的文献记录,当然有的信息是不公开的。信息爆炸的负面影响即信息混乱和信息烟尘,关于同一客观事件的不同文献可能有截然不同的记述,互联网上更遍布着大量没有经过验证的虚假信息。人们周围充满着过量的信息,当追踪信息时,却发现难以找到真正需要的、经过科学归纳的、具有针对性的有效信息。在过剩的信息海洋中,人们的信息吸收利用率却在下降。面对信息混乱与信息烟尘的污染,人们除了要掌握一定的信息检索能力外,还应具备一定的分辨、识别和处理信息的能力。

4. 信息也是生产力

在历史的长河中,“材料”、“能源”、“信息”支配着人类最基本的社会活动,在不同历史时期三者所起的作用有所不同。能源推动人类社会的车轮前进,材料是建设人类社会的物质基础。那么信息是什么?没有信息,石油永远埋在地下,矿石将不能成为钢铁。

当今,信息资源作为一种重要的战略资源已经成为社会生产力的要素。信息资源的研究开发和利用在很大程度上决定了一个国家和地区的经济水平和竞争实力。信息不仅是科学技术的一个重要组成部分,也是工作、学习和生活中不可缺少的因素。人们必须重视信息的作用,重视文献信息资源的开发与利用。信息检索作为开发、利用信息资源的首要步骤,有助于提高工作效率,把知识和智慧成果转化为生产力,从而推动经济发展和社会的进步。

1.1.2 学习信息检索的意义

1. 有利于减少课题的重复研究,提高科研成功率

任何科学研究都是在继承前人知识的基础上有所发明、有所创新的。任何人从事某一特定领域的学术活动,或开始一项新的科研工作,都要花费大量的时间,对有关文献进行全面的调查研究,摸清国内外是否有人做过或者正在做同样的工作、取得了一些什么成果、尚存在什么问题,以便借鉴、改进和规划自己的工作。只有这样才能做到心中有数,才能有所发现、有所创新、有所前进,否则容易造成重复劳动,导致人力、物力、财力的浪费。正如牛顿所说:“假如我比别人看得略为远些,那是因为我站在巨人的肩膀上”。这里,巨人是指前辈科学家和他们积累下来的知识。

据我国一项调查统计表明,我国的科研成果重复率达 60%,这种重复劳动的损失是无

法估量的。

2. 有助于节约时间, 提高科研效率

随着科学技术的发展、文献数量剧增、学科间相互渗透, 科研人员在进行一项科研活动中, 需要花费大量的时间查询与该研究项目相关的文献资料。如果熟悉信息检索方法, 就能大大节省查找资料的时间, 从而提高科研工作的效率。在当今世界, 提高科研效率、加快科研速度的意义还在于使相同科研课题在国内外竞争中处于有利位置。专利法规定, 对相同的发明成果, 按先申请原则授予专利权, 即只授予第一个申请人专利权, 其后申请的发明作为已知技术处理。显然, 如果忽视科研速度, 即使科研获得了成功, 但由于发明失去了时间上的新颖性, 因此会变成无效劳动。

3. 有助于协助管理者正确决策

准确、可靠和及时的信息, 是正确决策的基础。在竞争激烈的今天, 如果不能通过查阅文献, 获得国内外有关本行业的发展动态, 做到知己知彼, 那么无论经营何种产业都如同“盲人骑瞎马, 夜半临深池”一样危险, 很难有成功的希望。

4. 有助于增加知识积累, 改善知识结构, 提高自身素质

英国情报学家布鲁克斯曾提出关于情报与知识的基本方程为

$$K(S) + \Delta I = K(S + \Delta S)$$

式中: $K(S)$ 为原有的知识结构; ΔI 为信息增量; $K(S + \Delta S)$ 为新的知识结构。

该方程式表明, 新的知识结构是随着吸收信息量而增加的, 而吸收信息量又取决于原有的知识结构。据美国工程教育协会统计, 美国大学毕业的科技人员所具有的知识, 只有 12.5% 是在大学阶段获取的, 而 87.5% 则来自于工作实践。大学毕业后, 5 年内不补充新知识, 原有的知识 50% 会失效。

掌握了信息检索的方法和技能, 就会找到一条吸收和利用新知识的捷径, 以最少的时间和精力, 继承前人的知识, 最大限度地增加知识积累, 改善知识结构, 不断提高自身素质, 在激烈的竞争环境中立于不败之地。

5. 有助于培养信息意识, 提高信息素质

信息素质 (Information Quality) 或信息素养 (Information Literacy) 实质是一个个人信息搜集、整理、筛选、判断、评价和利用方面的能力, 是个人综合素质的一个方面。信息检索课是对大学生进行信息素质教育的主要形式之一。信息素质教育内容主要由信息意识教育、信息道德与信息法规教育、信息能力教育等组成。其中, 信息意识教育主要培养大学生对信息的敏感度, 或捕捉、分析、判断和吸收信息的自觉度; 信息道德和信息法规教育要使学生防止信息垃圾与信息污染, 不借助网络进行人身攻击, 不侵犯他人的知识产权、隐私权; 信息能力教育包括信息认知能力、信息获取能力、信息处理能力和信息利用能力等的培养。

1.2 信息检索基本知识

1.2.1 基本概念

一、信息

1. 信息的概念

当今时代, 信息是一个得到广泛应用的概念。信息如同空气一样, 同人们的生活息息相

关。人的五官、身体无时无刻不在感受着信息,语言、文字是人们信息交流的最主要形式。信息是自然界、人类社会和思维活动中普遍存在的现象,而关于信息的概念定义,不同专家学者从不同的角度出发,却有着许多不同的认识和理解。

信息论创始人美国著名数学家香农(C. E. Shannon)于1948年在《通讯的数学理论》一书中,首次提出了信息量的概念和信息熵的计算方法,并给出了十分抽象的信息定义,即:信息是用来消除随机事件的形式的不确定性的东西。控制论创始人维纳(N. Wiener)在1948年出版的专著《控制论——动物和机器中的通信与控制问题》中认为:信息是人们在适应外部世界并且使这种反应作用于外部世界的过程中同外部世界进行交换的内容和名称。意大利学者朗高(G. Longo)于1975年在《信息论:新的趋势与未决问题》一书中认为:信息是反映事物的形式、关系和差别的东西,包含在事物的差异之中,而不在事物本身。

《辞海》(2009年版)将信息解释为: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。信息的量值与其随机性有关,如在接收端无法预估消息或信号中所蕴含的内容或意义,即预估的可能性越小,信息量就越大。

我国信息论专家钟义信教授在其《信息科学原理》一书中把信息定义为:信息是事物运动的状态和状态变化的方式,是事物的一种属性。

不同学科从不同的研究角度出发也赋予了信息不同的理解。从哲学的角度分析:信息是客观存在的,信息是物质世界的运动状态与转换方式,是物质的本质属性。从信息资源管理角度分析:信息是数据处理的最终产品,是经过搜集、记录、处理,能以检索的形式存储的事实或数据。信息的不同定义从不同的侧面、不同的层次揭示了信息的某些特征和性质。信息的定义之所以呈现多样化,主要有三方面的原因:第一,信息本身的复杂性,是一个多元化、多层次、多功能的综合物;第二,信息科学是一门新兴学科,是一门“大”学科,有许多分支学科,其内涵与外延不很确切,而且随着社会、经济和科学技术的发展处于不断发展之中;第三,人们出于不同的研究目的或使用目的,从不同的角度或层次出发,对“信息”必然做出不同的理解与解释。

综合各种信息定义的内涵和外延,认为:信息不是事件或物体本身,信息是事物某种属性的反映,是客观事物某些特征的表象和线索,是客观事物的存在方式及运动状态的陈述,通常以声音、语言、文字、图像、气味等方式传达其所表示的实际内容。信息不是物质实体,同时又不能脱离载体而存在,总是借助于一定的物质载体存储或传输。

2. 信息的特征

(1) 客观性。信息是事物运动变化和状态的客观反映,其实质内容具有客观性,信息客观性的特征是由信息源的客观性决定的。由于运动是普遍存在的,也就决定信息的普遍性。信息普遍存在于自然界和人类社会,同时,信息本身也具有客观实用性。

(2) 依附性。信息必须依附于一定的载体而存在,必须以文字、图形、音频、视频符号等形式依附于书籍、磁带、磁盘、光盘等载体上,信息与载体不可分割。

(3) 替代性。信息的内容可以物化在不同的载体上,传递过程中可以实现物质载体形态的变化。现代信息技术尽量减少由于信息载体变化而引发信息本身内容的衰变。

(4) 加工性。信息可以通过各种手段和方法加工处理,只有通过选择和提炼,去除无用

的信息才能使信息具有更大的价值。

(5) 传输性。任何信息,只有从信息源发出,经过信道传递,才能被信宿接收并进行处理和利用,不能传输的信息是无用的、无法存在的。

(6) 时效性。信息作为对事物存在方式的反映,随客观事物的变化而变化。信息如果不能反映事物的最新变化状态,它的效用将会降低,随着时间的推移将完全失去效用,成为历史记录。一般来说,信息的价值与其所处的时间成反比,也就是说,信息反映的内容越新,其价值越大;时间延长,价值随之减小,一旦超过其“生命周期”,价值就消失。信息相对于其反映的客观事实具有一定的滞后性,因此只有加快传输,才能减少信息的滞留时间。经济信息、军事信息等这一特征表现尤为突出。

(7) 价值性。信息是一种资源,信息的价值性主要表现为信息的效用。人们要想最大限度地实现信息价值,就必须在信息拥有者与信息需求者之间建立畅通的渠道。信息的价值性还表现为人们可以以较少的物质投入实现最大的社会效益和经济效益,即信息的投入可以使自然资源、人力资源和资本资源增值,从而使信息资源成为当今时代最重要的战略资源。

(8) 共享性。信息区别于物质的一个重要特性是它可以被共同占有,共同享用。也就是说,信息在传递过程中不但可以被信源和信宿共同拥有,而且还可以被众多的信宿同时接收和利用。信息与物质不同,不会因利用而减少,相反由于信息的传递、反馈、利用,信息的内容不断丰富。

二、知识

1. 知识的概念

《辞海》(2009年版)将知识解释为:人类的认识成果或结晶,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,是对现实的能动反应。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,成为人类共同的财富。

知识建立在信息基础上,人们在利用自然、改造自然的生产、生活实践中,将搜集到的信息进一步加工处理、分析浓缩、概括总结,形成对信息系统化的认知而成为知识。知识是认识产生飞跃的结果,是人们在改造客观世界的实践中所获得认识和经验的总结。知识是同类信息的深化、积累,是优化了的信息的总汇和结晶。

信息杂乱无章,是非难辨,而知识则往往经过实践的多次检验,具有系统化的特性。当然,知识的产生、发展与利用是一个不断更新的过程,知识本身始终是处于动态的发展过程中,它永远处于不断被充实完善和更新的状态。知识和信息如果仅仅拥有,还不能取得效益,只有通过利用,将知识和信息应用到实践创新中去才会有生命力。

依反映对象的深刻性,可将知识划分为生活常识和科学知识;依反映层次的系统性,可将知识划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;依反映知识的内容,可将知识划分为自然科学知识、社会科学知识、思维科学知识,哲学是三者的概括和总结。

2. 知识的特征

(1) 实践性。社会实践是一切知识产生的基础和检验知识的标准,科学知识对实践有重要指导作用。

(2) 规律性。人们对实践的认识,是一个无限的过程,人们获得的知识在一定层面上揭

示了事物及其运动过程的规律性。

(3) 渗透性。随着知识门类增多,各种知识可以相互渗透,形成许多新的知识门类,形成科学知识的网状结构体系。

(4) 继承性。每一次新知识的产生,既是原有知识的深化与发展,又是更新知识产生的基础和前提。知识被记录或物化为劳动产品后可以世代相传和利用。

三、文献

在我国“文献”一词始见于《论语·八佾》:“子曰: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。足吾能徵之矣”,大意是孔子能讲解夏、殷的典章制度,但杞、宋两国的典章制度因为缺乏足够的文献而无法证实。

宋代朱熹在《四书章句集注》中对“文献”一词解释为:“文,典籍也;献,贤也”。在这里“文献”指“典籍”和“贤人”两方面意思。贤人,即见多识广、熟悉掌故历史的人。

后来文献概念引申为“具有一定历史价值的图书文物资料,亦指与某一学科有关的重要图书资料”,如:历史文献、医学文献。

国际标准化组织《文献情报术语国际标准》(ISO/DIS5217)对文献的解释为:在存储、检索、利用或传递记录信息过程中,可作为一个单元处理的,在载体内、载体上或依附于载体而存储有信息或数据的载体。《国际标准书目著录(总则)》[ISBD(G)]将文献定义为:文献是指以任何实体形式出现,作为标准书目著录的书目文献实体。

现代文献一词可广义理解为:记录有知识的一切载体的统称,即用文字、图像、符号、音频、视频等手段记录人类知识的各种载体,如竹简、书籍、胶片、磁带、磁盘、光盘统称为文献。

四、情报

情报一词在英文中被翻译为 Intelligence。前苏联情报学家米哈依诺夫将情报解释为,作为存储、传递和转换的对象的知识;英国情报学家布鲁克斯认为,情报是使人们原有的知识结构发生变化的那一小部分知识;我国著名科学家钱学森指出,情报就是为了解决一个特定问题而需要的知识;《辞海》(1999年版)对情报的解释为,获取的它方有关情况以及对分析判断的成果。情报按内容性质分为政治情报、经济情报、军事情报和科技情报等。

1992年9月,国家科学技术委员会提出将情报改称信息后,情报一词逐步被信息所代替,如科技情报改称科技信息,情报检索改称信息检索,《辞海》(2009年版)也不再收录情报一词。但事实上,情报一词并没有消失,更多情况下,情报被理解为是对知识的加工与传递,强调知识信息所发挥的作用,即情报是指被传递的知识或事实,是知识的激活,是运用一定的媒体(载体),越过空间和时间传递给特定用户,解决科研、生产中的具体问题所需要的特定知识和信息。

五、信息、知识、文献和情报的关系

从信息、知识、文献和情报的定义可看出,信息、知识、情报同属一个范畴,相互之间存在包含与从属关系,即信息的领域最广泛,知识次之,情报最窄,文献则是固化三者的有效载体。可以理解为:信息经过大脑的系统化加工、经过实践的检验成为知识,知识经过加工传递被接收者利用则成为情报,文献是记载信息、知识、情报的最终载体。

四者之间存在交叉关系,它们可以相互转化,文献记录的知识传递给用户,并被用户所利用时,就转化为情报;情报虽大多来自文献,但也可通过语言、实物来传递,情报在不需

要利用时,则还原为客观知识。

1.2.2 信息检索

信息检索 (Information Retrieval), 是指将大量无序的信息按照一定的规则方法有序地组织起来, 信息服务人员或信息用户根据需要从有序化的信息集合 (检索工具) 中找出有关信息的过程, 其全称为信息存储与检索。狭义的信息检索则仅指整个过程的后半部分, 即从信息集合中找出所需信息的过程, 相当于人们所说的信息查询 (Information Search)。信息检索也称情报检索, 最早发展于传统图书馆的参考咨询工作。

一、信息存储与信息检索

信息存储就是将搜集到的原始信息经过加工处理, 著录其内、外部特征 (内部特征有主题词、分类号、内容摘要等, 外部特征有题名、责任者、出版者、开本、页数等) 而形成信息记录 (款目), 并将这些信息记录有序组织起来的过程。信息检索是在有序化的信息集合中找出有关信息的过程, 是信息存储的逆过程。

然而, 由于文化水平、思想方法、表达方式等方面的差异, 信息标引者 (存储人员) 与信息用户对同一信息的分析、理解也必然存在差异。例如“青少年犯罪问题”依《中国图书馆分类法》分类, 标引者可能将其归入“C 社会科学总论——青少年问题”类, 而检索者则可能在“D 政治、法律”类进行查找, 从而影响检索效果。因此, 必须使信息存储与检索依据一致的规则, 这样, 无论什么样的标引者, 对同一文献的标引结果都一致, 不论是谁来检索, 都能查到该文献。

信息存储与检索共同遵循的规则称为信息检索语言 (详见第3章)。只要标引者和检索者用同一种检索语言来标引要存入的信息特征和要查找的检索提问, 使它们变成一致的标识形式, 信息的存储过程与检索过程就具备了相符性。相应地, 存入的信息也就可以通过检索工具 (系统) 检索出来。如果检索失败了, 那么就要分析检索提问是否确切地描述了待查课题的主题概念; 在利用检索语言标引时是否出了差错, 从而导致检索提问标识错误。只有检索提问标识和信息特征标识一致时, 相关的文献才能被检索出来。

二、信息检索的分类

1. 按检索内容划分

(1) 文献信息检索 (Document Information Retrieval)。文献信息检索是以文献 (包括文摘、题录和全文) 为检索对象, 即围绕某一特定课题, 利用目录、文摘、索引、数据库、网络搜索引擎等检索工具, 获得大量相关文献线索的过程。文献信息检索以获取相关文献线索、获得原始文献为最终检索目的的完成。例如, 我国正在开展的特高压电网建设工程, 用户可以通过文献信息检索获得与之相关的建设规划、专著、学术论文、技术专利等方面文献。

(2) 数据信息检索 (Data Information Retrieval)。凡是以数据为检索对象, 将经过选择、整理、鉴定的数值数据存入检索系统中, 根据需要回答某一问题的具体数据的检索都称为数据信息检索。数值型数据可包括地理常数、统计数据、人口数据、经济数据、设备参数等。有些数据信息检索系统不仅能查出数据, 而且具备一定的运算推导能力。例如, 通过《中国人口统计年鉴 2002》可以查询到 2000 年第五次全国人口普查的相关数据; 通过《电力工程电气设计手册》可以查询到关于“软导线的技术性能和荷重”的资料数据。

(3) 事实信息检索 (Fact Information Retrieval)。凡是以事实作为检索对象, 即将存储

于文献或数据库中的关于某一事项内容相关信息查找出来的信息检索。其检索对象既包括事实、概念、知识、定义等非数值信息,有时也包含部分数据信息。有些检索系统(如智能搜索引擎)可以针对用户查询要求,由检索系统进行分析、推理、逻辑运算后,再输出检索结果。例如,通过《华夏妇女名人词典》可以查询到中国共产党1~12届中央委员会女委员名单;通过“CNKI(中国知网)学术定义搜索”可以查询到大量关于“信息检索”的学术定义。

2. 按检索手段划分

(1) 手工信息检索(Hand Information Retrieval)。利用传统纸载体检索工具进行的信息检索,检索工具有图书馆卡片目录、文摘、索引及各类工具书。

(2) 计算机信息检索(Computer Information Retrieval)。利用计算机信息系统进行的信息存储和检索,其检索过程是在人机协同作用下完成的。计算机会从其存储的大量数据中自动分拣出与用户提问相匹配的信息,而用户则是整个检索方案的设计者和操纵者。与手工检索相比,其检索的本质没有发生变化,发生变化的只是信息的载体形式、检索手段、存储方式和匹配方法。

3. 按组织方式划分

(1) 全文信息检索(Full Text Information Retrieval)。全文信息检索建立在全文数据库自动标引基础之上,可以实现文献题名、主题词、内容摘要、章节、正文、参考文献等任意字段组配检索。该项检索由于需要进行全文扫描,一般不宜使用常见的通用词汇进行检索,若查询新词或某段话的文献出处更具检索意义。网络搜索引擎在对网页信息全文自动标引的基础上提供网络信息的全文搜索,智能搜索引擎同时根据检索词在网页中出现的位置、频率,判断网页与检索词的相关性,并依此对检索到的网页进行排序。

(2) 超文本信息检索(Hyper Text Information Retrieval)。超文本信息检索是对互联网每个节点中所存信息及信息链构成的其他相关网页中的信息进行的检索。

(3) 超媒体信息检索(Hyper Media Information Retrieval)。超媒体检索是对存储的文本、图像、声音等多种媒体信息的检索。超媒体检索既包含基于信息特征描述的文本检索,也包含基于信息内容特征(颜色、纹理、音频等)的高级检索。

1.3 文献信息资源概述

1.3.1 信息资源

一、信息资源与信息源

狭义理解,信息资源是人类社会经济活动中经过加工处理有序化并大量积累起来的有用信息的集合,其内涵定义在信息的本质表现,即有用的记录、文献、数据;广义理解,信息资源是与人类社会信息活动相关多种要素的集合,包括信息本身、信息人员、信息技术、信息设施、信息管理机制等。在信息社会,信息资源是人类社会物质财富和精神财富的基本来源,是人类进行物质生产和精神生产的主要现实基础,是人类社会得以生存和发展的基本前提。

信息源不等于信息资源,信息源是蕴含信息的一切事物,人类对于自然界和社会的任何知识,全都来源于社会实践。科学研究人员、工程技术人员和一切参加物质资料生产的人们

所从事的科学技术活动,是信息产生的源泉。信息资源则是可以利用的信息的集合,信息资源可以是一种高质量、高纯度的信息源,但信息源并不全是信息资源。信息源是信息资源的来源,经过开发利用,信息源可以转化为信息资源。信息资源生产者的一部分工作就是通过对各类信息源的加工、整序,使其成为有用的信息资源。

二、信息资源的分类

1. 体载信息资源

体载信息资源是以人体为载体并能为他人识别的信息资源,按其表达方式又可分为口语信息资源和体语信息资源。通过体载信息资源获取信息时,一般可以与信源(信息传播者)进行实时的交流与互动,往往可以获得其他方式所不能获得的高价值信息。例如,某一学科、领域的知名专家学者往往是该学科的重要信息资源。

2. 实物信息资源

实物信息资源是以实物为载体的信息资源。依据实物的人工与天然特性又可分为天然实物信息资源(如矿石、土壤、生物标本等)和人工实物信息资源(如样品、模型、雕塑等)。相对于其他信息资源,实物信息资源大多是直接为生产服务的技术信息,具有直观、真实、易检验、易仿制的特点,同引进技术与设计相比,花钱少而见效快,是值得重视的信息源。例如,在家具、服装等以外观设计为主要内容的行业,其产品实物往往更能直接反映其本身的知识信息。

3. 文献信息资源

文献信息资源是以文字、图形、符号、音频、视频等手段记录在一切载体之上的信息资源,是最常见、利用率最高的一种信息资源。

4. 网络信息资源

网络信息资源是文献信息资源的一种新形式,其载体为连接于互联网上的海量计算机存储设备。网络信息资源是下一代信息资源的最主要类型,通过计算机技术、通信技术、多媒体技术相互融合,人们可以实现全球范围内实时的文献信息获取与传递。

1.3.2 文献信息资源的基本功能

1. 存储功能

文献是保存人类文化遗产,记录人类文明进步与社会科学文化发展水平的重要载体。科学知识的继承、历史文化的传播、生产实践经验的积累很多都是通过文献来完成的。例如,一个国家航天科技水平较高,就会有大量反映其航天技术的科技文献。

人们对古代文明、科学发展、社会生活状态的认识,都是通过对留存下来的古代文物、文献的研究而实现的。

2. 传递功能

文献是人类传递文化知识、科技信息的主要媒介。一般科技信息的传递分为正式传递渠道(在刊物上发表或出版专著等)和非正式传递渠道(讨论、参观、讲演、展览、通信、交换等),前者是主要的传递手段。

3. 再生功能

人们在研究某一课题时,首先要借助于已有的文献,从中获取大量的间接经验,以便使研究少走弯路,缩短周期,提高效率。同时,新成果被记录下来又成为新的文献。新文献在水平上超过了原有文献,所以,这是一个阶梯式发展的、推陈出新的过程。

4. 鉴别功能

科技文献往往是用来衡量科研成果质量高低的重要依据。在评价一项研究成果时,首先要与前人发表的有关文献进行比较,从已有文献提出的观点、方法、工艺和数据等方面逐项核实,以确定该成果的新颖性、先进性。

1.3.3 文献信息资源的类型和特点

从不同角度出发,可以将文献信息资源划分为几种不同的类型。

一、按载体形式划分

为了有效地记录、积累和传递信息,人类先后发明和运用了各种各样的物质材料记录科学知识,例如:刻在甲骨上,铸在青铜器上,焙烧在陶器上,写在简策、布帛上;后来印在纸张上,晒在蓝图上,摄在感光材料上,录在磁性材料上,刻录在光盘上等。总之,随着社会的发展和科学技术的进步,文献的载体形式越来越多样化。

1. 刻写型文献

刻写型文献是印刷术未发明之前的古代文献和当今未付印的手稿等,如甲骨文、金石文、碑刻、简帛文、手稿等。

2. 印刷型文献

印刷型文献是以纸张为存储介质,以印刷为记录手段而产生的具有悠久历史的一种文献形式,今后较长时间内,印刷型文献仍将是文献的主体。

3. 缩微型文献

缩微型文献是以感光材料为存储介质,缩微照相为记录手段而生产的文献形式,主要包括缩微胶卷、缩微胶片等。缩微型文献需借助于缩微阅读机阅读。目前,缩微型文献已逐渐被计算机可读型文献所取代。

4. 机读型文献

机读型文献即计算机可读型文献。它是以磁性材料或激光材料为存储介质,以磁录技术和激光技术为主要记录手段,并通过计算机处理而形成的一种文献形式。其优点是存储密度高,存取速度快,可根据需要来组织其体系结构,便于数据更新;缺点是有专用设备才能阅读,费用较高。便携型电子图书阅读器的普及可能在未来带来一场阅读方式的革命。

5. 声像型文献

声像型文献又称为视听资料。它是以磁性材料和感光材料为存储介质,借助特殊的机械装置,直接记录声音或图像信息而形成的一种文献形式。它主要包括唱片、录音带、录像带、幻灯片、电影片等。其优点是生动直观,能够帮助人们认识某些复杂或罕见的自然现象,能够展示物质结构和运动机制等方面难以用文字来描述的科技知识。传统声像型文献多以模拟信号为主要记录格式,其缺点是容易造成信息的衰减,声像文献的数字化存储是发展的主要方向。

6. 网络文献

网络文献是指在互联网上发布的一种机读型文献形式,如一些出版物的网络电子版、Web 页面及其链接等。相对于传统出版形式,网络出版发展快速,开放存取资源日益丰富,互联网已成为学术资源交流与传播的重要方式。但相对于正规网络出版物,网络论坛、博客、贴吧等也存在大量随意发布,内容准确性、完整性难以验证的信息,信息质量良莠不齐,因此在使用或引用网络文献信息时应注意筛选和甄别。

二、按内容特征及加工层次划分

1. 一次文献 (Primary Document)

一次文献也称为原始文献,是指由作者本人来源于科研和生产实践而撰写或发表的学术论文、科研报告、技术说明书等原始创新成果。它具有内容新颖、使用价值高等优点,是科研人员追踪的主要目标。

一次文献的判断不是依据其载体、存在形式及出版方式,而是根据它的内容性质,只要作者根据自己的研究成果而发表的原始创作,都属于一次文献。

一次文献具有创造性、原始性和分散性的特点。创造性指一次文献是作者创造性劳动的结晶,包含着新观点、新发明、新技术、新成果,具有直接参考和使用价值;原始性是指作者的首次发表;分散性是指成果的个别性,同一学科、主题的相关一次文献出版分散。

例 期刊论文《论民事审判监督制度之完善》,作者:江必新,发表于《中国法学》2011年第5期。文章提出“民事审判监督程序的再次修订需要调整建构的理念,准确定位审判监督程序的性质,坚持辩证施治的方针,注意平衡各种相关价值,并对当事人申请再审条件、启动再审主体、管辖、再审审理程序等方面的规定予以合理化改造。”此文为一次文献。

2. 二次文献 (Secondary Document)

二次文献是人们把大量的、分散的、无序的一次文献搜集起来,按照一定的方法进行加工、整理、提炼、浓缩,即著录其外部特征和内容特征,标引出文献的主题,编制成具有多种检索途径的检索工具,使之系统化和便于查找而形成的文献。二次文献是对一次文献的加工重组,并不是新的信息,类型主要有题录、文摘、索引等。

二次文献具有汇集性、检索性、系统性的特点。汇集性是指它是在大量的分散性文献的基础上加工整理而形成的,一般汇集某一单位或专业范围内的文献,能直接反映某一信息部门或专业学科的文献收藏出版状况;检索性是指它所汇集的不是文献本身,而是一定范围内的文献线索,其本身是检索工具,还需要通过文献线索来进一步获取原始文献;系统性是指二次文献实现了对不同专业学科文献的系统组织。

例 国内报刊检索工具《全国报刊索引》、英国的《科学文摘》均为二次文献。

3. 三次文献 (Tertiary Document)

三次文献是围绕某个主题,利用二次文献搜集大量相关的一次文献,经过研究、分析综合而编写出来的文献,如各种综述、述评、学科年度总结、年鉴、数据手册等。

三次文献具有综合性、价值性和针对性的特点。综合性是指需要对大量分散的有关课题的文献、数据、事实进行分析、综合、浓缩,其内容较为概括;价值性是指三次文献可以直接提供有价值的信息,供用户参考、利用,三次文献兼有一次文献的作用,有时可以替代原文;针对性是指三次文献大多数都是为特定目的而编写的,一般情况下是信息部门受用户委托而从事的研究成果,研究人员通过综述性文章的撰写也可以全面了解当前某一领域的研究状况。

例 论文《数据挖掘在电力系统中的应用综述》,作者:陈星莺。该文结合电力系统的自身特点,全面分析数据挖掘在电力系统当前的主要应用动态及应用前景。此文为三次文献。