

XINXI JIANSUO YU LIYONG
SHIYONG JIAOCHENG

信息检索与利用

实用教程

主编 陈苗新
主审 赵 尘 丁其祥

011110000111111101010011100011101000111001010011100011

011111000011111110101001110001111010001111001010011100011

011111000011111110101001110001111010001111001010011100011

0110011000111000111000111010101001100110000011110101

011111000011111110101001110001111010001111001010011100011

011110000111111110101001110001111010001111001010011100011

011111000011111110101001110001111010001111001010011100011

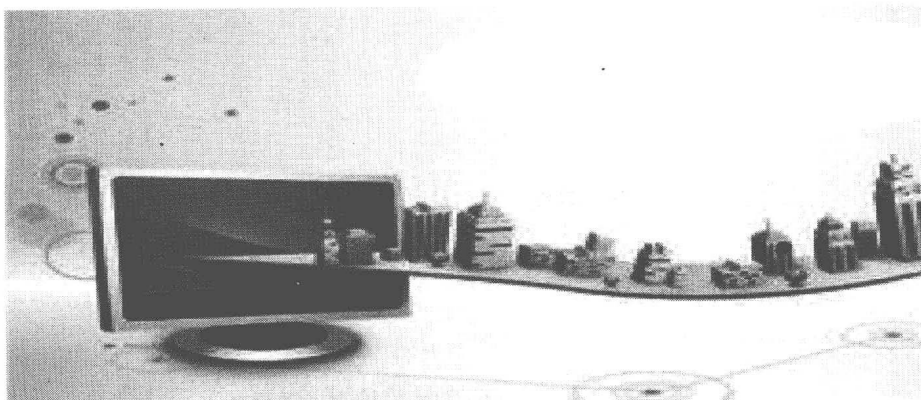


国防工业出版社
National Defense Industry Press

信息检索与利用实用教程

陈茜新 主编

参 编 刘刚 陆燕 梅嘉
主 审 陈伟 朱剑峰
赵尘 丁其祥



国防工业出版社
National Defense Industry Press

内 容 简 介

针对网络环境下信息资源与信息服务的的发展趋势与特点、学生信息素养的培养以及学生学习、研究与创新能力的需要,本书内容涵盖了包括信息检索理论、信息检索技术、数据库的使用方法与检索技巧等基本知识、信息素养教育、信息资源的合理利用、学术论文写作与学术规范等实用知识方面的内容。本书力求体现“易用、实用”,是一部内容全面、结构合理、重点突出、新颖直观、实用性和可操作性俱强的信息检索课教材。

本书可作为高等院校本科生、研究生信息检索学课程的教材,也可供图书馆与信息管理人员、科研人员使用。

图书在版编目(CIP)数据

信息检索与利用实用教程/陈苗新主编. —北京:国防工业出版社,2011.12
ISBN 978-7-118-07820-6

I. ①信... II. ①陈... III. ①情报检索—教材
IV. ①G252.7

中国版本图书馆 CIP 数据核字(2011)第 235099 号

※

国防工业出版社出版发行

(北京市海淀区紫竹院南路 23 号 邮政编码 100048)

三河市鑫马印刷厂

新华书店经售

*

开本 787×1092 1/16 印张 11 1/4 字数 283 千字

2011 年 12 月第 1 版第 1 次印刷 印数 1—5000 册 定价 22.00 元

(本书如有印装错误,我社负责调换)

国防书店:(010)88540777

发行邮购:(010)88540776

发行传真:(010)88540755

发行业务:(010)88540717

前 言

随着全球信息化、网络化的发展,以数字化、网络化为特征的信息技术在全社会各个领域广泛渗透和运用,人类已经处于以计算机网络为核心的新的信息环境——网络信息时代。网络环境下,信息内容无所不包,信息资源数量巨大,印刷型信息与电子型信息各具特色,其种类繁多,分布、传播、存取和利用的方式多样化,信息传递速度在不断加快。如何在这浩瀚的信息海洋中捕捉有用信息,挖掘知识,迅速查找出无论是对于科研、教学还是实验最需要或最有用的信息资源,就成为信息时代面临的重要课题。

大学图书馆是大学生汲取知识营养和更新专业知识的文献宝库与知识基地,处于数字时代的大学图书馆,更兼具传统图书馆与数字图书馆的双重特征。网络环境改变了人们获取信息的方式及思维与认知模式,信息素养成为信息社会人们必备的一项技能。作为大学生专业教育“第二课堂”的大学图书馆,如何使高校学生尽快适应这种日新月异的数字化新环境,改变学生的学习观念与教育模式,培养学生在网络环境下高效学习的方法与能力,培养大学生的信息能力和创新能力,是我国大学课程与教学改革必须着重研究和思考的重大课题。

信息素养最早是由美国人泽考斯基在 20 世纪 70 年代提出来的,1989 年,美国图书馆协会(American Library Association)的“信息素养主席委员会(Presidential Committee on Information)”在总结报告中给信息素养的定义是:“要成为一个有信息素养的人,他必须能认识到何时需要信息,并具有检索、评价和有效使用必要信息的能力。”美国大学与研究型图书馆协会指导委员会批准的《高等教育信息素养能力标准》中明确提出信息检索、分析、组织和利用的能力是一个具备信息素养的人(学生)应达到的目标。“开展信息素质教育,培养读者的信息意识和获取、利用文献信息的能力”,也是我国《普通高等学校图书馆规程(修订)》所规定的高校图书馆的任务之一。所有这些,对《信息检索学》课程的教学内容、教学方法和教学手段都提出了新的要求。

“信息检索学”是一门根植于图书馆学和情报学,又与其他学科如计算机科学、数学、逻辑学等有千丝万缕联系的学科;它既是一门理论方法论课程,又是一门实践性、技能性很强的课程。掌握“信息检索学”的基本理论、基本方法和基本技能不仅可以提高学生收集、整理、组织、分析和利用各种信息资源的能力,更能促进与培养他们在信息意识、信息价值、信息获取中所涉及的学术道德、信息安全等信息素养观念的形成与发展。同时,从长远看,有利于培养学生的创新精神,提高学生的学习、研究与创新能力。

本书编者是一些多年从事“信息检索学”课程教学的教师,具有丰富的理论知识和实践经验。针对网络环境下信息资源与信息服务的的发展趋势与特点,对学生信息素养的培养以及学习、研究与创新能力的需要,本书的编者共同研究大纲、分工合作,逐步形成了自己的特色。

(1) 内容全面。全书共分 8 章,主要内容包括信息与信息素养、信息检索概论、计算机信息检索基础、常用中外文学学术数据库资源、网络信息资源、特种文献资源、信息资源合理利用。同时,与时俱进,补充了大量最新的信息。

(2) 结构合理。既有信息检索理论、信息检索技术、数据库的使用方法与检索技巧等基本知识,又有信息素养教育、信息资源的合理利用、学术论文写作与学术规范等培养教育学生学习、研究与创新能力的实用知识。同时,每章后附加思考题,以利于学生在更深层次上理解和掌握信息检索学知识。

(3) 重点突出。把握时代脉搏,顺应时代特征,重点突出电子资源的使用方法与检索技巧以及对信息资源的合理利用,全面系统地介绍了当前电子资源(包括网络信息资源、中外文数据库)的所有内容。

(4) 新颖直观。全书配有大量数据库检索界面和检索结果示意图,读者可通过图示直观模拟各种数据库检索系统的操作流程。同时,增加了实际检索案例,寓理论传授于实际应用,联系实际,贴近学生。

本书由陈茁新主编,他对全书写作大纲、结构及内容进行了策划与设计,并最终对全书修改、定稿。各章节分工如下:本书第8章、第1章第1节、第5章第1节由陈茁新编写,第2章、第5章第6节由刘刚编写,第3章、第6章、第5章第4节、第5节由陆燕编写,第4章、第7章、第5章第2节、第3节由梅嘉编写,第1章第2节由陈伟编写,朱剑峰参与了第8章第3节的部分编写工作。

本书的编写得到了国防工业出版社的大力支持和帮助。南京林业大学图书馆赵尘馆长不仅审阅了全书,而且还给予了人力、物力的帮助,在此表示深深的谢意。

由于作者能力及水平有限,书中难免存在一些疏漏和不足乃至错误之处,特别是本书在编写过程中参考了大量的国内外文献及同类图书,恐未能在书后参考文献处全部列出,在此向作者表示歉意和感谢。同时,敬请专家和读者给予批评指正,以便使教材能够日臻完善。

陈茁新

2011年8月于南京

目 录

第1章 信息与信息素养	1		
1.1 信息概述	1		
1.1.1 信息的概念	1		
1.1.2 信息的特征	2		
1.1.3 信息的分类	3		
1.1.4 科技文献的特征与类型	4		
1.2 信息素养	8		
1.2.1 信息素养的概念与主要 表现能力	8		
1.2.2 信息素养的评价标准	9		
1.2.3 图书馆与信息素养教育	11		
思考与训练1	14		
第2章 信息检索概论	15		
2.1 信息检索概述	15		
2.1.1 信息检索定义	15		
2.1.2 信息检索原理	15		
2.1.3 信息检索类型	15		
2.2 信息检索语言	17		
2.2.1 分类检索语言	18		
2.2.2 主题检索语言	19		
2.3 信息检索工具	20		
2.3.1 检索工具定义	20		
2.3.2 检索工具类型	20		
2.3.3 检索工具结构	21		
2.4 信息检索的途径、方法与步骤	21		
2.4.1 信息检索途径	21		
2.4.2 信息检索方法	22		
2.4.3 信息检索步骤	23		
思考与训练2	24		
第3章 计算机信息检索基础	25		
3.1 计算机信息检索概述	25		
3.1.1 计算机信息检索的概念、 原理与特点	25		
3.1.2 计算机信息检索的发展 历程	26		
3.1.3 计算机信息检索系统构成	27		
3.1.4 数据库	28		
3.2 计算机信息检索基本技术	30		
3.2.1 布尔逻辑检索技术	30		
3.2.2 截词检索技术	31		
3.2.3 词位限制检索技术	32		
3.2.4 检域限制检索技术	33		
3.3 计算机信息检索步骤	34		
3.3.1 信息需求分析	35		
3.3.2 检索策略制定	36		
3.3.3 检索策略的实施和调整	37		
3.3.4 检索结果的筛选、整理和原始 信息的获取	38		
思考与训练3	39		
第4章 常用中文学术数据库资源	40		
4.1 期刊全文数据库	40		
4.1.1 CNKI 中国知网——中国期刊 全文数据库	40		
4.1.2 万方数据资源系统——中国 数字化期刊群	44		
4.1.3 中文科技期刊全文数据库—— 维普数据库	46		
4.2 电子图书数据库	49		
4.2.1 超星数字图书馆	49		
4.2.2 读秀学术搜索	51		
4.3 其他中文电子资源选介	52		
4.3.1 中国林科网	52		
4.3.2 华通数据库	53		
4.3.3 国家科技图书文献中心电子 资源	53		
思考与训练4	54		
第5章 常用外文学术数据库资源	55		
5.1 Elsevier 电子期刊 (ScienceDirect)	55		
5.1.1 数据库简介	55		

5.1.2 检索方法	57	6.5 免费网络学术资源	106
5.1.3 检索历史	62	6.5.1 开放存取资源	106
5.1.4 检索设置及结果处理	63	6.5.2 电子预印本	109
5.1.5 检索实例	64	思考与训练6	112
5.2 SpringerLink	66	第7章 特种文献资源	113
5.2.1 SpringerLink 简介	66	7.1 专利文献	113
5.2.2 SpringerLink 检索方法	67	7.1.1 专利的基本概念	113
5.3 ProQuest 博硕士学位论文全文数据库(PQDD)	72	7.1.2 专利的种类	113
5.3.1 资源简介	72	7.1.3 专利授予的条件	114
5.3.2 检索方式	72	7.1.4 专利权的特点	114
5.3.3 检索结果	74	7.1.5 专利申请、审批、授权及 保护程序	115
5.4 Engineering Information Village	75	7.1.6 专利文献	115
5.4.1 资源简介	75	7.1.7 国际专利分类法	117
5.4.2 资源获取	76	7.1.8 专利的检索	118
5.5 AGRICOLA、AGRIS、CAB Abstracts	79	7.2 标准文献	120
5.5.1 AGRICOLA、AGRIS、CAB Abstracts 简介	79	7.2.1 标准和标准文献的概念	120
5.5.2 OvidSP 检索平台使用	80	7.2.2 标准的种类	120
5.6 美国《化学文摘》光盘网络版	89	7.2.3 标准级别、代号与编号	122
5.6.1 概述	89	7.2.4 标准的有效期和标准的 复审	123
5.6.2 美国《化学文摘》光盘网络版 检索方法	89	7.2.5 国际标准化组织	123
思考与训练5	95	7.2.6 标准文献分类	123
第6章 网络信息资源	96	7.2.7 标准的检索	124
6.1 网络信息资源概述	96	7.3 会议文献	125
6.1.1 网络信息资源特点	96	7.3.1 会议文献的定义	125
6.1.2 网络信息资源类型	97	7.3.2 会议文献的特点	125
6.2 网络信息资源检索方法	99	7.3.3 会议文献的类型	126
6.2.1 浏览方法	99	7.3.4 会议文献的检索	126
6.2.2 检索方法	99	7.4 学位论文	127
6.3 网络信息资源检索工具	99	7.4.1 学位论文的概念	127
6.3.1 网络目录	99	7.4.2 学位论文的特点	127
6.3.2 搜索引擎	100	7.4.3 学位论文的收藏单位	127
6.3.3 特殊检索平台	100	7.4.4 学位论文的检索	128
6.4 搜索引擎	101	思考与训练7	129
6.4.1 搜索引擎的工作原理	101	第8章 信息资源综合利用	130
6.4.2 搜索引擎类型	101	8.1 科学评价工具——国内外著名 检索系统	130
6.4.3 主流搜索引擎介绍	102	8.1.1 美国科学引文索引	131
		8.1.2 科学技术会议录索引	134
		8.1.3 工程索引	134

8.1.4 中文社会科学引文索引 (CSSCI)	135	8.3.1 学术规范	146
8.2 原始文献获取	136	8.3.2 文献的合理使用	149
8.2.1 原始文献获取的原则	137	8.3.3 国内外学术文献投稿 指南	155
8.2.2 原始文献获取的渠道	137	思考与训练 8	168
8.2.3 原文传递服务	143	参考文献	169
8.3 学术论文写作规范与投稿	146		

信息与信息素养

1.1 信息概述

1.1.1 信息的概念

追根溯源,“信息”一词有着很悠久的历史,早在两千多年前的我国西汉,即有“信”字出现。“信”常可作消息来理解。作为日常用语,“信息”经常是指“音讯、消息”。南唐李中《暮春怀故人》诗:“梦断美人沉信息,目穿长路倚楼台。”以及宋陈亮《梅花》诗:“欲传春信息,不怕雪埋藏。”等都是例证,但至今信息还没有一个公认的定义。

1948年,美国数学家、信息论的创始人香农在题为“通信的数学理论”论文中指出:“信息是用来消除随机不定性的东西”。这里,香农从通信科学的角度给出了“信息”的定义:信息是人们对事物了解不定性的减少和消除,是两次不定性之差。

同年,美国著名数学家、控制论的创始人维纳在《控制论》一书中指出:“信息是人们在适应外部世界并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总称。”认为“信息就是信息,既非物质,也非能量。”

我国《辞海》中对“信息”一词的注释是:“信息是指对消息接受者来说预先不知道的报道。”

我国国家标准《情报与文献工作词汇基本术语 GB 4894—85》这样定义信息:“信息是物质存在的一种方式、形态或运动状态,也是事物的一种普遍属性,一般指数据、消息中所包含的意义,可以使消息中所描述事件的不定性减少。”例如,“城市交叉路口的红绿灯”只是一种交通信号灯,其本身并不是信息,但我们所得到的“红灯停,绿灯行”这一信息;“电子体重秤”只是客观存在的物体,但人站上去之后显示的数字却向我们传达了身体重量的信息。

可见,信息无所不在,可以感知,但它不是事件和物质本身,它是客观事物的存在方式或运动状态,以及关于客观事物存在方式或运动状态的陈述。信息是原料,经过人类的认识活动,成为已知的知识。

信息可以从广义和狭义两个层面理解。

广义的信息概念是指一切事物存在方式及其运动规律、特点的外在表现形式,以及信息活动中各种要素的总称,包括信息本体、与信息有关的人员、设备、技术和资金等各种资源。简言

之,信息是关于一切事物运动状态和运动方式的反映。

狭义的信息概念是指文献资源或数据资源,包括任何媒体中的片段、文章、图书、情报、观念等。

1.1.2 信息的特征

1. 普遍性

信息是物质固有的普遍属性,哪里有物质,哪里就有信息。人类社会以前,无机界运动变化过程在大自然中留下的痕迹和印记是信息。例如,我们的祖先可以从一块石头的外观、硬度、质地等方面来获取相关信息,从而判断它能否作为制造工具的材料或狩猎的武器。而如今的地质学家则可从石头的成色、光泽及元素含量等方面获取信息,从而确定它是否是矿石。可以说,从宇宙天体到微观世界,从客观自然到主观现实,到处都充满着信息。在社会实践和日常生活中,在不同的领域内,人们都在时刻自觉或不自觉地创造、传递、接收和利用信息。信息作为社会的一种基本要素,其作用已渗透到人类社会生活的各个方面。

2. 传递性

物质是运动的,因而信息也是运动的。只不过信息的运动是一种传递与反馈过程。例如,红绿灯提示你是否可以过马路,打领带时领带结的皱折可以显示领带的质地(有皱折的为真丝领带),树叶的生长与凋落向人们传递春天与秋天来临的信息,等等。人们在个体信息阶段主要通过自然语言直接进行信息的传递,而在文字发明之后,信息的传递可以间接进行,即利用信息载体进行传递。从最早的甲骨文、陶文到竹简、纸,到现在的软盘、光盘、多媒体、电视等。由此可见,信息传递载体可以分为瞬时型和存储型两大类。人们完全可以根据自己的目的选择适当的物质载体及高效、低耗、快速的信息传递方式。

3. 时效性

信息的时效性是指从大众媒介发出到受众接收、利用的时间间隔及其效率。随着大众传播科技的飞速发展,人们对信息时效性有增无减的需求将会得到进一步满足,信息传播与接收将会越来越快。在时间面前,信息是易碎品。即使是十分真实的、很有价值的信息,一旦失去了时效,它就会变成无人问津的东西。大众媒介中的昨日消息、上午新闻,正迅速地被刚刚发生的、正在发生的甚至即将发生的信息所取代。人们对时效性的追求,是没有止境的。然而,信息的时效性又与一般物质通常概念意义上的时效性并不完全一致,换言之,信息的时效性并不完全表现为其使用价值随时间的延后而降低。信息的时效性更多的表现在人们的需求性。例如,时尚(Fashion)给人感觉理应是时髦、前卫的,但仿古的未必就不时髦,不时髦,甚至流行。典型的例子就是电影《花样年华》,由于其中女主角张曼玉不断变换的各种各样的旗袍,使得本该是20世纪30年代的服装在20世纪末又重新流行起来。

4. 共享性

同一内容的信息,可以在同一时间或不同时间、同一地点或不同地点被若干个用户使用,此时,信息载体本身的信息量并不因此而磨损、消耗、消失,这就是信息的共享性。教师可以把某一门学科的知识传授给千百个学生、Internet上的信息可以同时有成千上万人浏览或引用。路甬祥院士在上海第二届亚太地区城市信息化高级论坛上作题为“共享信息资源缩小数字鸿沟”的发言时说:“信息资源有许多不同于物质资源的特性,最重要的一点就是它的共享性。信息资源可以以极低的成本进行克隆,也就是无差别的复制。这不仅意味着共享信息资源的代价很小(与信息本身的价值相比基本可以忽略不计),非常易于实行,而且信息复制品不存

在质量和功能上的缺陷,完全和信息源具有同样的功用。另外,信息的共享并不对信息源造成任何的伤害(保密性的问题除外),信息的拥有者并不需要为与他人分享而做出牺牲。因此,信息资源是一种特殊的资源,本质地具有易于共享的特性。”“信息资源还有一个适于共享的特点,就是可开发利用的多样性。相同的信息资源可以在众多不同的场合发挥作用,例如,遥感卫星的数据在地学、环境、气象、军事等诸多领域都有重要的应用。现实中人们开发某项信息资源的时候往往有特定的目的,但这些信息其实可以有更多的用途。通过信息的共享,大大丰富了信息的开发利用,也就大大地增加了信息的价值。”但是,需要强调的是,虽然信息具有共享性的特点,同一个信息,对于不同素质的接收者来说会产生不同的实用价值。人们掌握信息也是要付出代价的,因为对于无所不在的信息,每个人必须经过获取、学习、消化,才能成为自己的东西。

5. 可转换性

只要信息含义、内容不变,其存在形式可以相互转换。换言之,信息可从一种形态转换成另一种形态,如物质信息可以转换成语言、文字、影像、数据信息,文字信息可以转换成计算机代码及无线电信号等。同样的信息内容在信息形态的不断转换中使信息的使用价值得到更好的体现。并且通过信息形态的转换,人们实现了对信息的去粗取精和去伪存真,将信息以一种更易辨识和更高度综合的方式表达出来,从而使信息更加具有使用价值。

6. 可知性

信息存在于宇宙中,人们可以通过不同的方式来感知信息。然而,由于人们认识水平的差异,对于同一事物,不同的人获得的信息量不同,更有许多信息尚未被人们认知,但这并不意味着未被认知的信息永远不会被认知。随着科技进步与发展,人类会认知越来越多的信息。

1.1.3 信息的分类

依据不同的分析方法及划分标准,信息类型可以从以下几个方面进行划分。

1. 按信息来源划分

按信息来源分,信息可分为直接信息和间接信息。直接信息是指从信息源直接获得的信息,也有的将之称为体载信息。比如口口相传的信息,它具有迅速、及时的优点,但不易收集和管理,因而其检索利用受到限制。间接信息是相对直接信息而言的,它不是直接从信息源获得,而是通过某种中介(往往是信息的载体)获得,比如图书、期刊、网络,因而具有全面、系统等优点,也便于检索利用。

2. 按信息的作用划分

按信息的作用分,信息可分为有用信息、无用信息和干扰信息。有用信息是指对人们的科学研究、教学工作或生产管理具有借鉴参考价值的信息,反之就是无用信息或似是而非的干扰信息。最直接明显的例证是,在搜索引擎 Google 中检索信息时,检索结果所呈现给我们的是几百万甚至上亿条的记录,而这其中,真正对我们的需要有帮助的信息不多,甚至非常少。因此,信息检索过程中,我们应尽量剔除无用信息,排除干扰信息。

3. 按信息的传递方向划分

按信息的传递方向分,信息可分为前馈信息和反馈信息。所谓前馈信息是指在输出信息没出现偏差之前,控制部分可发出控制信息来纠正即将发生的偏差的这一类信息,是预测的或期望的信息。在科学研究领域,是指科学研究或教学或生产管理中获得的研究成果、经验积累、学术提炼等可供人们进行学习、研究或借鉴的有用信息。反馈信息则是对前馈信息的回

应,是指把实施的行为结果回输给相关部门的情报信息。

4. 按信息的应用部门划分

按信息的应用部门分,信息可以分为工业、农业、政治、军事、经济、文化、管理等信息。这是一种具有重要现实意义的分类,对使用者最具有帮助价值。

1.1.4 科技文献的特征与类型

1. 文献的概念

依据中华人民共和国国家标准(GB/T 4894—2009 信息与文献、术语),文献(Document),是记录有知识的一切载体。

以这个定义而言,文献的构成可以从以下三方面来理解:

第一,必须具有体现历史价值或学术价值的信息或知识。

第二,必须具有用来表述信息或知识所使用的记录符号或语言符号。

第三,必须具有记录信息和知识的载体。从古代的甲骨、陶、竹简、纸到现在的计算机用软盘、光盘等。

据此,凡是用文字、图形、符号、声像等手段记录科技活动或科学知识的文献,都被称作科技文献。

2. 科技文献的特征

1) 数量众多,增长迅速

据估计,在现代科学发展初期,科学文献的年增长率为6%~8%。20世纪初,每10年左右数量翻一番。进入20世纪80年代,每3年左右翻一番。据不完全统计,目前全世界每年出版的图书在80万册以上;科技期刊15万种左右,发表的期刊论文在600万篇以上;登记专利100万件;国际会议文献1万种以上。据《乌利希国际期刊指南》(Ulrich's International Periodicals Directory)统计,截止2006年底,全世界现仍出版(Active)或停刊(Ceased)的定期及不定期期刊、年刊(Annual)、连续性出版品(Monographic Series)、报纸等,种数总计超过250000种。我国期刊总量目前也已超过9000种。现代文献增长的速度可以美国化学学会编辑出版的《化学文摘(Chemical Abstracts)》为例进行说明。《化学文摘》连续报道100万篇文献所用的时间从前至后依次为:32年,18年,8年,4.75年,3.3年。现在仅为两年左右。

2) 类型复杂,文种多样

随着科学技术的发展,声、光、电、磁等现代技术和化学塑料新材料的广泛应用,使现代文献载体发生了重大变化,文献载体早已突破了传统方式,出现了除印刷型文献之外的缩微型、视听(声像型)、机读型等多种文献并存的趋势。从其内容上可分为图书、期刊等11大类型。按其加工层次又可分为三个加工层次。除此以外,文献语种也从过去的英、德、法等几种语种发展到现今的近百种语种,比较集中的也不下十余种。例如,“美国化学文摘”涉及的语种多达52个,“美国工程索引”涉及20多个语种,而“国际农业和生物科学中心文摘”涉及的语种也有近50种。据有关调查报告称,世界各国出版的期刊中,有1/2的期刊是用一半左右的科学家所不懂的语言文字出版的。

3) 分布不均,交叉重复

由于科学发展与技术的进步,现代科技各领域之间呈现相互渗透,相互交叉的发展趋势,导致科技文献的分布越来越离散。例如,遗传学方面的论文可能出现在医学杂志上,也可能出现在农学杂志上。控制论方面的论文可能会发表在神经科学的期刊上。1934年英国文献计量学家

S. C. 布拉德福在第 3550 期《工程》(Engineering)杂志上发表了“专门学科的情报源(Sources of Information on Specific Subjects)”一文,文中首次提出了文献集中与离散规律,即布拉德福文献离散定律。该定律指出:如果将科技期刊按刊载论文数量递减顺序排列,则可分出一个核心区和相继的几个区域,每区刊载的论文量相等,此时核心期刊和相继区域期刊数成 $1:a:a^2$ 的关系。具体就某个学科而言,其论文发表数量与刊登论文的期刊数量呈现这样的比例关系:

1/3 的论文刊载在该学科专业期刊上,这部分称之为核心区(n_1)。

1/3 的论文刊载在与该学科相关的期刊上,这部分称之为相关区(n_2)。

1/3 的论文刊载在与该学科不相关的期刊上,这部分称之为边缘区(n_3)。

而刊载论文的期刊数量之间的比例则为:

$n_1:n_2:n_3=1:a:a^2$ (其中 $a\approx 5$, 通常称为布拉德福常数)

据《中文核心期刊要目总览》(2008 版),林业类核心期刊为 17 种,而林业类专业期刊为 147 种,核心期刊仅占到其中的 11.56%。

此外,一种期刊常刊载众多学科的论文,同一科技内容会出现在专利、会议录、期刊等多种文献上。有人统计,美国 AD 报告的 60%,农业部科技报告的 80% 会转为期刊论文再次发表。专利文献重复率更大,有报道称,加拿大国家专利的 87% 与国外专利说明书的内容重复。

4) 文献信息有效期缩短

社会的进步,科技的发展,使科技文献有效使用时间日益缩短,失效周期明显加快。有资料表明,大部分科技文献的使用寿命一般为 5 年~7 年,甚至更短。

部分文献的平均时效为:

图书著作:10 年~20 年

期刊论文:3 年~5 年

科技报告:10 年

学位论文:5 年~7 年

技术标准:5 年

产品样本:3 年~5 年

科技文献的以上特点,一方面为利用文献带来了极大的便利,另一方面也带来了查找文献的困难。本教程的目的就是解决文献多——看不完;文献杂——找不全;文献无序——找不快的矛盾。由此我们也可以看出文献检索的作用:可以作为知识的更新手段,通过研究、参考借鉴,开发智力资源;可以避免重复劳动,节省时间。

附:美国科学基金会和日本国家统计局的统计调查。

一个科研人员在—项科研活动中的科研时间分配情况为:

思考研究计划(开题)

7.7%

查阅文献资料

50.9%

进行实验研究

32.1%

编写研究报告

9.3%

3. 科技文献的类型

1) 按文献的载体分

(1) 印刷型文献。以纸张为存储介质,以手写、石印、油印、铅印等为记录手段的传统文献类型,是人类交流传递信息的主要形式,也是迄今为止占据主导地位的类型。它具有阅读便利,流传广泛的优点。然而其存储密度较低,且占用空间较大,保存也有一定的难度。

(2) 缩微型文献。以感光材料为存储介质,以缩微照相为记录手段而产生的一种文献类型。包括缩微胶卷、胶片和缩微卡片等。此类文献体积小,存储密度高,便于保存和收藏。缺点是需借助昂贵的缩微阅读机方能阅读。

(3) 电子型文献。其前身被称为机读型文献。它以磁性材料作为存储介质,以穿孔或电磁、光学字符作为记录手段,并通过计算机处理而产生出来的一种文献形式。它主要包括磁带、磁盘、光盘数据库、文献数据库、电子文献(包括电子图书、电子期刊)等,具有存储密度高、更新快、查找文献迅速等优点。但同样也存在设备投资大的缺点。

(4) 视听型文献。也称声像资料。是以磁性材料或感光材料为载体,以磁记录或光学技术手段直接记录声音、视频图像的一类文献,主要包括唱片、录音带、录像带等。存储密度高,直观生动,但成本也较高且不易更新和检索。

2) 按文献加工层次分

(1) 零次文献。零次信息资源指的是信息内容尚未经加工整理的信息,是未经发表和有意识处理的最原始的资料。包括私人笔记、书信、手稿、试验记录、设计草图、口头交流的信息等。零次信息一般是通过口头交谈、参观展览、参加报告会等途径获取。

(2) 一次文献。也称原始文献。是信息生产者根据自己的观察、科学实验及研究成果而第一次撰写并第一次出版的那些信息。一般指公开出版的图书、期刊报纸、科技报告、会议论文、学位论文、专利、技术档案等。

(3) 二次文献。对一次文献进行加工、提炼和浓缩之后得到的文献,具体而言,是指对一次信息加工、整理后形成的各种检索工具,包括目录、题录、文摘、索引等检索类出版物。它不对一次文献提供评论,只是把分散、无序状态的一次文献变成集中、有序化的二次文献状态,以此提供一次文献的检索线索,是人们迅速、准确检索一次文献的重要工具。

(4) 三次文献。又称参考性文献。是指利用二次文献线索,系统地检索出一批相关文献,并对其内容进行综合、分析、研究和评述而编写出来的文献。可分为综述研究类和参考工具类两种类型。前者如动态、综述、专题述评、进展报告等,后者如(字)词典、百科全书、年览、手册等。

3) 按出版形式分

(1) 图书。通常分为阅读类和参考类两大类,如教科书、学术专著、参考工具书等。从信息检索的角度考虑,我们应了解图书的一些基本常识。

图书在版编目数据(Catalogue In Print, CIP):是指对已出版的图书进行预编目而形成的MARC(机读代码)数据,其中有一项,称之为国际标准书号。

国际标准书号(International Standard Book Number, ISBN),是国际上通用的出版物标识编码系统。一部专题出版物在制作、销售和发行过程中只能有一个ISBN编码。

国际标准书号的前身是英国标准书号。1970年,国际标准化组织(ISO)颁布了ISO 2108《信息与文献——国际标准书号(ISBN)》,这项标准在国际上获得了广泛应用。经过多次修订ISBN的使用范围已由传统的印刷出版物扩展至视听资料、缩微制品、电子、网络等非印刷出版物领域。

现行国际标准书号由10位数字组成,前面冠以英文字母“ISBN”。10位数字分为组号、出版者号、书序号、校验码四部分,各部分之间用“—”隔开。

[例1-1] ISBN 7-5038-2565-0

其中:7—地区号,又叫组号,此处代表中国大陆出版物使用的代码—汉语(0,1—表英语,2—表法语,3—表德语);

5038—表示出版社代码,此处代表中国林业出版社;

2565—表示种次号,此处代表中国林业出版社出版的第2565种书;

0—计算机校验码。

2005年6月1日国际标准化组织(ISO)颁布第4版国际标准书号,ISO 2108:2005《信息与文献——国际标准书号(ISBN)》。新版国际标准书号将ISBN由10位升至13位,ISBN与EAN-UCC(全球开放的物流信息标识和条码表示系统)相结合,使用了国际编码协会为ISBN分配的专用编码978、979,增加为前缀码;规范了ISBN系统的管理;增加了与ISBN定位有关的元数据及其元数据的交换方式。ISBN升位后系统的编号容量得到了极大的扩展。新版《国际标准书号》已于2007年1月1日全面实施。

新版国际标准书号由13位数字组成,前面冠以英文字母“ISBN”。13位数字分为前缀号、组号、出版者号、书序号、校验码五部分,各部分之间用“-”隔开。

[例1-2] ISBN 978-7-301-04815-3

其中:前缀978(代表图书),中间一部分与ISBN相同,最后一位是校验码。

中国于1982年参加ISBN系统,并成立中国ISBN中心(设在国家新闻出版署)。中国标准书号(China Standard Book Number)自1987年1月1日起实施,至1988年1月1日完全取代原用的统一书号。

中国标准书号共分两部分,第一部分为ISBN,是主体部分;第二部分为《中国图书馆图书分类法》基本大类类号和种次号。类号除工业技术诸类图书用两个字母外,其他各学科门类图书均用一个字母。种次号是同一出版社出版同一学科门类图书的顺序号,由出版社自行编定。类号与种次号之间用中圆点(·)隔开。第一部分和第二部分分两行排列,也可用斜线隔开,排成一行。例如,ISBN 7-144-00316-X/TP·340。

(2) 连续出版物(serials)。指具有统一的题名,以连续分册形式出版,每个分册有卷期和年月标识,并且计划无限期地出版下去的印刷型或非印刷型出版物。包括期刊、报纸、年度报告、年鉴、年刊、指南、学会报告、丛刊和会刊、连续出版的专著丛书和会议录等,但不包括在一个预定有限期内以连续分册形式出版的著作。各期由多篇文献汇集而成,内容具有原始性和新颖性,学术性较强。由于连续出版物报道及时,出版连贯,数量、种类庞大,成为现代文献的一种主要类型。而期刊又是其中最主要的一种。

期刊,又称杂志,在英语中用periodical, journal, magazine表示。最早的期刊是1665年1月5日在巴黎创办的《Journal des savans(学者)》。现在我们较为熟悉的期刊有《Science》,《Nature》,《读者》等。期刊可以按其属性分为学术类期刊、技术类期刊、综合类期刊、检索类期刊、科普类期刊五种,其中的学术类、技术类以及检索类,对于我们进行学术研究,查找科研线索有十分重要的作用,尤其是各学科或各专业的核心期刊。

1931年著名文献学家布拉德福首先揭示了文献集中与分散规律。布拉德福发现学科文献在期刊中的分布是有规律的,少量期刊集中了大量某学科文献(某时期某学科1/3的论文刊登在3.2%的期刊上),而其他期刊则很少出现该学科文献。他将期刊按发表学科论文的数量排序,划分出对学科最有贡献的核心区和随后的若干区。1967年联合国教科文组织研究了二次文献在期刊上的分布,发现75%的文献出现在10%的期刊中;1971年,SCI的创始人加菲尔德统计了参考文献在期刊上的分布情况,发现24%的引文出现在1.25%的期刊上等,这些研究都表明期刊存在“核心效应”。后人的进一步研究证实了布氏揭示的学科文献在期刊中的分布规律是客观存在的,因而将布氏发现的规律称为布拉德福文献集中与分散定律,将位于核心区的期刊称之为核心期刊,从而衍生了“核心期刊”的概念。

国外对核心期刊的研究始于1934年布拉德福用载文量对期刊进行评价,后人又用文摘

量、引文量等对期刊进行评价研究,20 世纪 60 年代加菲尔德用引文量对期刊进行评价。我国国内对核心期刊的研究则开始于 20 世纪 70 年代末 80 年代初研制的“国外科技核心期刊表”,后期又由一些研究机构相继研制了《中文核心期刊要目总览》、《国外人文社会科学核心期刊总览》、《中国人文社会科学核心期刊要览》,目前国内最具影响力的科技类核心期刊表是由北京大学图书馆、北京高校图书馆期刊工作研究会研制的《中文核心期刊要目总览》,从 1992 年开始研制出版,每隔 4 年修订一次,现在使用的是 2008 版。

核心期刊(Core Periodical)不是一个广义的概念,而是基于学科、专业或专题而言的。确切说,某学科(某专业或某专题)的核心期刊,是指在该学科所涉及的期刊中,刊载相关论文较多(或信息量较大),论文学术水平较高的,并能反映该学科最新研究成果及该学科前沿研究状况和发展趋势,较受该学科读者重视的那些期刊。

核心期刊是期刊中学术水平较高的刊物,是学术评价体系的一个重要组成部分。它的作用体现在这几个方面:①作为期刊采购的参考工具;②作为图书馆导读和参考咨询的参考工具;③作为评价学术研究成果的参考工具;④作为读者投稿的参考工具;⑤作为文献数据库选择来源刊的参考工具;⑥对提高期刊质量有促进作用。

(3) 特种文献。是指出版、发行和获取方式都比较特殊的一类文献,也称之为非书非刊。这部分文献包括 8 种,具体为:专利文献、标准文献、会议文献、学位论文、科技报告、产品资料、技术档案、政府出版物。特种文献特色鲜明、内容广泛、数量庞大、参考价值高,是非常重要的信息源。这一部分的内容会在第 7 章中专门论述。

1.2 信息素养

1.2.1 信息素养的概念与主要表现能力

1. 信息素养的概念

“信息素养(Information Literacy)”的本质是在全球信息化下需要人们具备的一种基本能力,是从图书馆检索技能发展和演变过来的。信息素养最早由美国信息产业协会主席保罗·泽考斯基(Paul Zurkowski)在 1974 年提出,当时将信息素养定义为:人们解决问题时利用信息工具或主要信息源的技术与技能。美国图书馆学学会(American Library Association,ALA)会议 1989 年对信息素养的定义为“To be information literate, a person must be able to recognize when information is needed and have the ability to locate, evaluate, and use effectively the needed information.”意思是说,具有信息素养的人有能力去察觉何时需要信息,且能有效寻找、评估与使用所需要的信息。它包括:能够判断什么时候需要信息,并且懂得如何去获取信息,如何去评价和有效利用所需的信息。

这样的看法目前已成为一种共识。除美国外,一些西方发达国家的观点大同小异。美国高等教育图书研究协会(Association of College & Research Libraries,ACRL)提出的“高等教育信息素质能力标准”(Information Literacy Competency Standards for Higher Education),(包括 6 个标准指标和 22 个具体评价标准),在美国已被广泛认可和接收,英国、澳大利亚等根据国情,稍有补充和修改。

1992 年“信息素养”一词正式被 ERIC 数据库纳入索引词典,标志着信息素养教育得到

确认。

2. 信息素养的主要表现

美国教育技术 CEO 论坛 2001 年第 4 季度报告提出 21 世纪的 5 种能力素质,包括基本学习技能(指读、写、算)、信息素养、创新思维能力、人际交往与合作精神、实践能力。信息素养是其中一个方面,它涉及信息的意识、信息的能力和信息的应用。换言之,信息素养是一种对信息社会的适应能力。因而,信息素养是一种基本能力。

信息素养涉及各方面的知识,是一种特殊的、涵盖面很宽的能力,它包含人文的、技术的、经济的、法律的诸多因素,和许多学科有着紧密的联系。信息技术支持信息素养,强调对技术的理解、认识和使用技能。信息素养的重点是内容、传播、分析,包括信息检索以及评价,是一种了解、搜集、评估和利用信息的知识结构,因此,信息素养的培养既需要熟练的信息技术,也需要通过完善的调查方法、鉴别和推理来完成。信息技术是信息素养培养的一种工具,信息素养是一种综合能力。

信息素养主要表现为以下 8 个方面的能力。

(1) 运用信息工具。能熟练使用各种信息工具,特别是网络传播工具。

(2) 获取信息。能根据自己的学习目标有效地收集各种学习资料与信息,能熟练地运用阅读、访问、讨论、参观、实验、检索等获取信息的方法。

(3) 处理信息。能对收集的信息进行归纳、分类、存储记忆、鉴别、遴选、分析综合、抽象概括和表达等。

(4) 生成信息。在信息收集的基础上,能准确地概述、综合、履行和表达所需要的信息,使之简洁明了、通俗流畅并且富有个性特色。

(5) 创造信息。在多种收集信息的交互作用的基础上,迸发创造思维的火花,产生新信息的生长点,从而创造新信息,达到收集信息的终极目的。

(6) 发挥信息的效益。善于运用接受的信息解决问题,让信息发挥最大的社会和经济效益。

(7) 信息协作。使信息和信息工具作为跨越时空的、“零距离”的交往和合作中介,使之成为延伸自己的高效手段,同外界建立多种和谐的合作关系。

(8) 信息免疫。浩瀚的信息资源往往良莠不齐,需要有正确的人生观、价值观、甄别能力以及自控、自律和自我调节能力,能自觉抵御和消除垃圾信息及有害信息的干扰和侵蚀,并且完善合乎时代的信息伦理素养。

3. 信息素养培养的核心内容

(1) 培养用户具有能够认识到何时需要信息,能够有效地检索、评估和利用信息的综合能力。

(2) 培养用户能够将获取的信息与自己已有知识相融合,构建新的知识体系,解决所遇到的问题与任务。

(3) 培养用户能够了解利用信息所涉及的经济、法律和社会问题,合理、合法地获取和利用信息。

1.2.2 信息素养的评价标准

1. 美国高等教育信息素养能力标准

2000 年 1 月 18 日,美国高等教育图书研究协会在德克萨斯州的圣·安东尼召开了美国