

信息检索

INFORMATION RETRIEVAL

主编 ◇ 张静 陈秀华 金福臣



中国出版集团



世界图书出版公司

信息检索

张 静 陈秀华 金福臣 主编

中国出版集团

世界图书出版公司

广州·上海·西安·北京

图书在版编目(CIP)数据

信息检索 / 张静,陈秀华,金福臣主编. —广州: 世界图书出版广东有限公司, 2012. 9

ISBN 978-7-5100-5106-7

I. ①信… II. ①张…②陈…③金… III. ①情报检索
IV. ①G252.7

中国版本图书馆 CIP 数据核字(2012)第 203498 号

信息检索

责任编辑 杨力军 吴小丹

出版发行 世界图书出版广东有限公司

地 址 广州市新港西路大江冲 25 号

<http://www.gdst.com.cn>

印 刷 东莞虎彩印刷有限公司

规 格 880mm×1230mm 1/32

印 张 9.25

字 数 250 千字

版 次 2013 年 9 月第 2 版 2013 年 9 月第 2 次印刷

ISBN 978-7-5100-5106-7/G·1107

定 价 30.00 元

版权所有,翻印必究

前 言

知识经济时代,信息广泛渗透到经济、科技、文化的各个学科领域乃至人类生活的各个方面,信息检索与利用已经成为提高人们信息素质和综合素质的重要手段之一。随着信息社会化,以及信息技术和信息资源的迅速发展,信息检索与利用的重要性越来越突出。它既是一个平台,一种工具,也是一个“引擎”。如何快捷、准确、及时、有效地获取所需要的信息,是知识经济和网络时代对信息检索提出的新要求,也是进行素质教育不可缺少的重要组成部分。

爱因斯坦曾经说过“用专业知识教育人是不够的”,信息素养已成为现代人才的必备基本素质之一。由于信息载体发生的巨大变化,除传统纸质信息外,还有大量的磁载体信息、电子信息及各类网上信息。信息的多样性增加了信息检索的难度,这极大地影响了人们获取信息的质量与效率。通过对信息检索知识的学习,能够帮助学习者掌握有效的信息获取方法,充分利用已有的信息资源,为科学研究的预测和决策提供必要的信息支撑。

信息检索教育的目的就是培养学习者的信息素质,使其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,科学、准确地获取信息,提高自学能力,激发创新能力。为此,我们在总结长期从事信息检索教学、科研服务和实际工作的基础上,结合各大检索系统的发展变化,并吸收各类相关教材的优点,编写了《信息检索》一书。

本书力图紧扣信息时代的脉搏,突出现代网络信息的检索与利用,压缩了传统检索知识的篇幅,着重介绍了近几年新出现的检索方法,将最新信息检索方法与利用作为主要内容。在进行国内外大



型检索系统、大型检索工具、特种文献的介绍时,采用了理论结合实践操作的模式(图文结合),便于读者举一反三,触类旁通。本书具有内容新颖、实用性强的特点;在文字上力求简练,在取材上力求新颖、翔实。

全书共 12 章,具体分工如下:

第 3—9 章、第 11 章由张静编写;第 1 章、第 12 章由陈秀华编写;第 2 章、第 10 章由金福臣编写。全书各章节初稿完成后,由张静负责统改全部稿件。

由于经验不足,水平所限,内容难免有疏漏和不当之处,恳请同行专家和读者批评指正。

在本书编写过程中,我们参阅了大量文献,因篇幅所限,仅列出主要参考文献,请有关作者见谅,并借此机会向这些文献作者表示诚挚谢意。

编 者

2012 年 5 月

目 录

第 1 章 概 论	001
1.1 信息、知识与文献	001
1.2 信息源	005
1.3 信息检索及其类型	013
1.4 信息检索语言与工具	015
1.5 信息检索途径、方法和步骤	018
第 2 章 常用中文信息资源	023
2.1 CNKI 中国知网	023
2.2 维普资讯	032
2.3 万方数据	037
2.4 国家科技图书文献中心	042
2.5 超星数字图书馆	047
2.6 书生之家	050
2.7 方正 Apabi	053
第 3 章 常用英文信息资源	059
3.1 Elsevier ScienceDirect	059
3.2 SpringerLink	063
3.3 Scopus	067
3.4 WILEY InterScience	073
3.5 ISI Web of Knowledge	076
3.6 Ei Village	084
第 4 章 图书馆资源与利用	092
4.1 图书馆概述	092



4.2	馆藏资源利用	098
4.3	图书馆信息服务	103
第5章	专利信息检索	116
5.1	概述	116
5.2	国内专利信息检索	126
5.3	国外专利信息检索	130
第6章	标准信息检索	137
6.1	概述	137
6.2	国内标准信息检索	143
6.3	国外标准信息检索	151
第7章	学位论文检索	160
7.1	概述	160
7.2	国内学位论文检索	160
7.3	国外学位论文检索	167
第8章	会议信息检索	171
8.1	概述	171
8.2	国内会议信息检索	172
8.3	国外会议信息检索	179
第9章	科技报告检索	185
9.1	概述	185
9.2	国内科技报告检索	186
9.3	国外科技报告检索	191
第10章	事实数据检索	196
10.1	概述	196
10.2	百科全书检索	199
10.3	字典、词典检索	206
10.4	名录信息检索	210
10.5	年鉴、手册检索	213
10.6	表谱、图录检索	218
10.7	工具书指南	222

10.8	中国工具书网络出版总库	224
第 11 章	Dialog 国际联机检索	227
11.1	概述	227
11.2	常用检索平台	232
11.3	常用检索指令	238
11.4	基本检索技术	243
11.5	联机检索的步骤	247
第 12 章	网络信息资源检索	250
12.1	概述	250
12.2	搜索引擎	258
12.3	学科网络信息资源检索	269
参考文献	285

第 1 章 概 论

1.1 信息、知识与文献

1.1.1 信息

信息与人类的关系十分密切,可以说人类从诞生开始就从事着信息活动。在文字发明以前,人们的手势、眼神、乃至捎口信等都是表达和交流信息的方式。人们用结绳记事来编码和存储信息,用烽火来传递信息。“信息”一词的出现已有上千年的历史,唐代李中的《暮春怀故人》一诗中就有“梦断美人沉信息,目穿长路倚楼台”的绝句。今天,“信息”一词的使用频率越来越高、越来越频繁。但究竟“信息”是什么?至今各说不一,没有一个统一的定论。

信息的英文 information,是消息、情报的意思,这是对信息的狭义解释。把信息作为一门科学来解释,“信息”一词的定义多达几十种。概括起来有以下几个方面。

信息的创始人申农(C. E. Shannon)在其经典论著中定义:“信息是用来消除不确定性的东西。”他把信息看做一个抽象的量,提出能否定义一个量,这个量在某种意义上能度量一个过程所产生的信息是多少。

控制论创始人维纳(N. Wiener)认为,信息是“我们对外界进行调节并使我们的调节为外界所了解时而与外界交换来的东西”。

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



我国大多数学者认为: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,伴随有信息的产生,信息借助于物质载体记录和媒体传播。

实际上,信息无处不在,它与万事万物相联系,在我们日常生活中每天都与信息打交道,我们通过感觉器官接受各种各样的外界信息,通过语言和手势表达和传递信息。信息是层出不穷的,人类在社会实践又不断地产生新信息。因此,我们说信息是客观存在的一切事物,通过物质载体所发出的消息、情报、指令、数据及它们中所包含的一切可传递的和可交流的内容,并随着自然界和人类社会的发展而不断产生新的信息。

1.1.2 知识

信息、知识与文献有一定的内在联系,但它们之间又有着明显的差别。

知识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反映。人类在社会实践中通过信息对自然界和人类社会发展的运动规律产生认识,再通过大脑的重新组合和系统化,就获得知识。人类通过世间存在的信息来认识世界和改造世界,人类的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。

1.1.3 文献

文献是用符号、图形、文字、视频、声频等方式记录人类知识的载体。《中华人民共和国国家标准文献著录总则》(GB 3792. 1-83)对文献的解释为:“文献是记录有知识的一切载体。”这里,文献具有两个基本的要素,一是含有知识;二是一种载体。文献记录知识,而这些知识又依附于载体而存在。正是文献的这些特点使人类的知识得以保存和传播,人类的科学和文化得以继承和发展。所谓载体,不仅包括印刷型出版物,还包括声像制品和电子出版物。因此,文献的范围非常广泛,古代的甲骨文是文献,碑刻、竹简、帛书是文



献,青铜器上的铭文是文献,包括今天的磁带、磁盘、光盘、缩微胶片等都是文献。

文献中有可能蕴藏着大量的信息,是信息检索的主要信息源。

信息的内涵范围很大,把一般的信息经过加工处理、系统化之后,信息就可转化成知识。知识是从信息积累中找出的规律性的东西。把知识用载体记载下来就是文献,因此,知识和文献是信息的组成部分,而有价值的信息转化成知识并促使信息再生。

1.1.4 信息的分类

分类是人们认识事物的一种最为基本的方法,也是科学研究活动中常用的一种方法。由于信息在信息界和人类社会生产中存在和流动的范围极其广泛,因此对信息的分类也相对复杂。不同学科领域的研究人员依据不同的分类标准,可对信息进行不同的划分。

(1)按信息内容划分,可将信息分为人类信息和非人类信息。

(2)按信息产生的先后和加工深度划分,可将信息分为一次信息、二次信息、三次信息。

(3)按信息的存在形式划分,可将信息划分为内储信息和外化信息。

(4)按信息的流通方式和传递范围划分,可将信息分为公开信息、半公开信息和非公开信息。

(5)按信息载体形式划分,可将信息分为感官载体形式、语言载体形式、文字载体形式、电磁波载体形式、缩微载体形式、光波载体形式、声像载体形式和计算机载体形式。

另外,我们还可以按信息管理组织、信息的反映面、信息的时态、信息的特征、人对信息的感知方式、信息运动形态等方式将信息划分成不同的类型。

1.1.5 信息的特性

信息定义所揭示的是信息的本质属性,但信息本身还存在许多由本质属性派生出来的一般特性,包括普遍性、时效性、传递性、共享性、开发性、变换性、转化性和可伪性等,它们大都从某一个侧面



体现了信息的基本特点。正是有了这些新颖的特性,才使信息成为既不同于物质又不同于能量的一类新的研究对象。同时,通过研究信息的这些新颖性质,又会有助于加深人们对信息基本概念的理解。在这里,我们给大家介绍几种信息的常见特性。

1. 普遍性

信息的普遍性是指信息的存在是普遍的,即信息是无处不在的,也是无时不在的。信息普遍存在于自然界、人类社会之中,也存在于人类的思维或精神领域之中。

2. 时效性

信息的时效性是指信息从发出到被接收的时间间隔及其效率。信息的功能、作用都是随着时间的改变而改变的,信息在获取、加工、传递和利用的过程中,时效性可以表明信息的经济价值。如果信息传递的时间慢,信息很容易失去效益,即使有用的信息也会失去价值,因此,信息的时效在很多时候就决定了信息的价值。

3. 传递性

传递性是信息的本质特征之一,是指信息可以通过一定的传输工具和载体进行传递,从而形成信息联系,被人们感知和接受。信息是物质存在方式的直接或间接显示,这种显示靠光、声、磁以及语言、表情、文字符号等得以呈现和传递。信息的可传递性使信息有可能在短时间内广泛扩散、传播。

4. 共享性

信息的共享性是指信息可被众多主体共享。信息共享是信息的运动规律之一,也是信息的一个重要性质,同时还是信息与其他资源的一个重要区别。信息在交换和转让过程中,其原有信息一般不会丧失,还可能同时获得新的信息。正是由于信息可以被共享,所以它是人类社会的宝贵财富。

5. 开发性

信息是一种可开发的资源,存储与传递信息的目的是为了开发信息资源。信息是点燃智慧的火种,它能给人们以新的知识,引导人们开拓更新的、更高层次的、更广阔的知识领域。人们正是通过



对信息资源的开发利用,来促进科学技术的进步和社会的发展。

6. 价值的不定性

信息的价值就在于将一人、一事、一时、一地的信息传递给需要者,从而创造新的物质财富和精神财富。由于人们对信息的需求和理解及判断能力的不同,信息的价值也有很大的差别。同样的信息对于不同的使用者可能有不同的价值,同样的,不同类型的信息对于不同的使用者也不一定有相同的价值。

1.2 信息源

信息无处不在,无时不有。人们的生产实践、日常生活,以及科研、贸易、金融、军事、政治等活动都是产生信息的源泉。

总体来看,信息来源有语言信息源、实物信息源和载体信息源。

实物包括自然实物和人工实物(人类文化的创造物,如文物、产品等),内含大量科技文化信息,具有文献所不具备的种种优点:直观、实用。实物是现实的商品,除了本身的信息价值外,还具有商品价值(转让)和使用价值。在不作为信息载体使用时,实物可投入流通或作为一般物品发挥它本身的使用价值,这是其他信息载体所不及的。

口头信息指通过交谈、讨论、报告等方式交流传播的信息。多数时候的口头交流传播都属于非正式交流,午餐、沙龙、公关活动等渠道都可能成为主要的交流方式,因而信息分散、随意,交流的信息多属于非公开出版,不具有记载性,不易获取,而且信息质量良莠不齐,存在大量谬传信息有待分析再加以利用。目前,随着通信技术的发展,我们可通过正式会议、视频会议、广播电视、固定与移动电话、网络语音对话等获得此类信息源。

实物信息和语言信息是还没有加工的信息,用户只能根据自己的需要去细心地、有目的地搜集、加工、整理、分析和利用。载体信息源是我们获取信息最主要的来源,根据不同的标准载体信息源又划分为多种不同的类型。



1.2.1 按出版发行方式形成的文献信息源

印刷型文献信息源是一种传统而又古老的记载信息的信息源, 这些信息数量庞大、种类繁多、内容丰富, 是获取信息的重要来源。印刷型文献信息源有如下几种。

1. 图书

图书(Book)即书籍, 有专著、丛书、论文集、工具书、教科书、参考书等, 每种图书都有一个中心主题或系统论述。图书的内容比较成熟、全面、系统、完整。尤其是科技图书, 一般是对已发表的科研成果、生产技术、科技知识和经验, 或某一知识领域进行系统论述或总结之后重新组织编写而成。它通常以会议论文、研究报告、期刊论文以及第一手资料为基本素材, 经过作者分析、鉴别而重新组合。它是综合、积累和传递科技知识的一种重要文献信息源。由于图书的出版周期较之期刊长, 其报道的信息比期刊论文和专业学科文献迟。但科技图书中提供的知识全面可靠, 有的内容还包括一些从未发表过的研究成果和资料, 查阅这些图书有利于加强对综合性问题的了解以及对陌生问题的认识。近年来, 电子图书大量面世, 且数据量越来越大, 内容十分丰富, 通过计算机就能方便快捷地获得相关图书信息。

2. 期刊

期刊(Journal/Magazine/Periodical)是一种有固定名称的连续出版物, 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, 报道的内容新颖。

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。有人曾做过统计, 从期刊得到的信息占了整个信息来源的 65%~75%, 有些学科和专业所占的百分比或许还要高些。

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, 流通面广泛, 连续性强, 能及时反映各国的科学技术水平。



有的期刊还登有文献述评、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。

3. 报纸

报纸(Newspaper)是新闻的主要载体,它的出版周期最短。报纸的特点是内容新、涉及面广、读者多、影响大。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,能让读者及时了解社会政治、经济、文化、科技等各个方面的信息。报纸信息具有连续性和完整性,通过报纸可以了解社会的当前信息和即将发生的事情。但报纸信息零乱、材料分散、知识不系统,是一种难以保存和累积的信息文献。

报纸的数量十分庞大,这给查阅资料带来困难,因此,一般须利用相关索引。电子版报纸的出版发行给查找报纸信息带来了极大的便利。

4. 科技报告

科技报告(Science & Technical Report)是科研工作成果或进展的研究报告和技术报告,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。最早科技报告是在 20 世纪之初出现的,尤其是在第二次世界大战期间及战后,各国政府加紧军事科研活动并不断增加经费投入,这使科技报告得到快速发展。科技报告大致可以分为基础理论研究和生产技术两大类,所反映的科研水平和成果比期刊论文快,且内容新颖、详尽、专深、具体、完整可靠,是科研过程的真实记录,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之中。

科技报告具有保密性,发行是受控制的,一般以单行本形式出版,只有公开与解密的科技报告,流传范围才会较广。目前,国外常见的科技报告主要有:军事系统的 AD 报告,由美国武装部队技术情报局收集、整理和报道;行政系统的 PB 报告,由美国商务部国家技术信息服务处负责搜集、整理、报道和发行;原子能和能源管理系统



的 AEC/ERDA/DOE 报告,由美国原子能委员会出版;航空与宇航系统的 NASA 报告,由美国国家航空和宇宙航行局出版。以上也就是常说的四大科技报告。

5. 学位论文

按国际标准化组织(ISO)的定义,学位论文(Dissertation/Thesis)是指作者为取得专业资格的学位而提出的介绍他本人的研究发现或某种结论的调查的文献。学位论文在美国通称为“dissertation”,在英国则称为“thesis”,上面注明了授予的学位头衔,授予学位的单位名称、时间、地点等。

学位论文,尤其是高层次的学位论文质量较高,它们带有一定的独创性,探讨的问题比较专深,对问题的阐述较为系统详细,其参考价值不亚于科技报告,因此是一种重要的信息源。学位论文属非卖品,一般不对外发行,这就造成了利用学位论文的不便。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。

6. 会议文献

会议文献(Proceedings/Conference Paper)是指在国际或各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。它分为会前、会中、会后出版物,即会前发言预印本或发言预摘、会议期间的论文汇编和会后整理出版的论文集或会议资料。学术会议的类型很多,概括起来有研究特定问题的专题讨论会;有交流学术观点、发表研究成果的学术研讨会;有一年一度讨论工作、规划来年活动的年会。在学术会议上交流的文献常常反映出科学技术的最新成就和发展趋势,探讨的问题专业领域集中、针对性强、内容专深。全世界每年都有上万次的各种学术会议召开,每年发表科技会议记录数千种,文章十多万篇。这些会议文献有许多不再在其他出版物上发表,即使要发表也需经过很长一段时间,因此,很多国际国内的科技工作者都把学术会议视作重要的信息交流场所,以收集和传递各种最新信息。会议文献是了解国际和各国的科技水平、动态及发展方向的重要信息资源。



7. 专利文献

专利文献(Patent Literature)是专利制度下的产物,它是一种知识产权。广义的专利文献指一切记录有关发明创造的文献,包括专利申请书、专利说明书、专利公报以及专利二次文献等。狭义的专利文献仅指各国专利局公布出版的专利说明书,它是一种受法律保护的文献。发明人因研制出某项新技术、创造出某种新的设计等,都可向政府主管部门提出专利申请,经审批后,就能获得一定时间限制(如年限)的专利权。

专利说明书中要指出发明实用新型和外观设计的名称、所解决的核心技术问题、解决的方法、使用后的效果、实施的最佳方案或实例等。专利说明书的内容很具体,同时附有图形,通过它可以了解该项专利的技术内容、经济价值和特殊用途等。据统计,全世界的专利文献虽仅占期刊文献的10%,却能提供40%左右的新产品信息,世界上90%~95%的新技术是通过专利文献公诸于世的。因此,专利文献是科技工作者极为重要的信息源,是了解科技发展沿革的途径之一。利用专利文献,通过专利文献披露出来的技术资料,从中仔细挑选和权衡,常常可获得质优价廉的技术信息,从而帮助人们顺利地进行技术贸易。专利文献涉及的技术内容比较广泛,从日常的生活用品到高科技领域,涉及面广、数量巨大。

8. 灰色文献

灰色文献(Gray Literature)是近年发现的一种新型信息源,一般指非公开出版发行的文献。灰色文献品种繁多,包括非公开出版的政府文献、学位论文;不公开发行的会议文献、科技报告、技术档案;不对外发行的企业文件、企业产品资料、贸易文件(包括产品说明书、相关机构印发的动态信息资料)和工作文件;未刊登稿件以及内部刊物、交换资料、赠阅资料等。灰色文献流通渠道特殊,制作份数少,容易绝版。虽然有的灰色文献的信息资料并不成熟,但所涉及的信息广泛、内容新颖、见解独到,具有特殊的参考价值。

9. 政府出版物

政府出版物(Government Publication)是指各国政府部门以及