

信息资源

检索与利用

程发良
陈伟
主编



國防工業出版社
National Defense Industry Press



信息資源

校書利用

一、二、三、四





普通高等教育“十二五”规划教材

信息检索与利用

编 著 曹丽娟 白首晏
主 审 李小菲



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。全书共分十一章，系统介绍了信息检索与利用的基本方法，主要内容包括信息检索概论，信息源与信息检索工具，计算机信息检索基础，图书与期刊信息检索，特种文献信息检索，重要中文信息检索系统，国外著名信息检索系统，图书馆信息资源利用，Internet 信息资源检索，信息的收集、整理与分析，科技论文写作等。

本书重点突出新颖性和实用性，着重介绍最新的、实用的信息检索技术与信息资源类型。

本书可作为普通高等院校的公共课教材，也可作为广大教师、科研人员、工程技术人员等信息检索的参考用书。

图书在版编目 (CIP) 数据

信息检索与利用/曹丽娟，白首晏编著. —北京：中国电力出版社，2011.8

普通高等教育“十二五”规划教材

ISBN 978-7-5123-2069-7

I. ①信… II. ①曹…②白… 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 (2011) 第 177838 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>)

航远印刷有限公司印刷

各地新华书店经售

*

2011 年 9 月第一版 2011 年 9 月北京第一次印刷
787 毫米×1092 毫米 16 开本 15 印张 362 千字
定价 26.00 元

敬告读者

本书封面贴有防伪标签，加热后中心图案消失
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

序 言

《信息检索与利用》是为高校在校生的和有关技术人员编写的一本教材式的工具书，或者说是为生活在现代社会中的知识分子在当下信息化时代的信息海洋里遨游提供的技术指导书。众所周知，现代社会的一个最基本的特征，就是信息化。信息已同物质、能源一起，成为人类社会生产与生活不可缺少的三大资源之一，而且还是在重复使用的传递流通过程中发挥作用并显现其价值，是可传递的资源能量。信息的占有者不会因信息被传递而失掉信息，也不会因信息的重复使用而改变信息的自身价值。虽然，由于信息时效的非恒定如一性（表现为人们认识落后于客观存在的信息的滞后性，以及人们在把握客观规律的前提下，能够对可能发生的事物进行预测的信息的超前性），决定了其价格、价值多元（变）性，即其价格、价值或是因滞后使用的时差而减值，或是因能够及时把握有效信息而获得最佳价值；但信息是有载体，可以识别、可以转换、可以储存、具有可塑性，并可以共享的本质则是恒定如一的。信息检索与利用，就是建立在前述信息的多元（变）性和恒定性对立而又统一的基础上、现代社会的人们对现代社会的一种适应性选择的自觉行为。

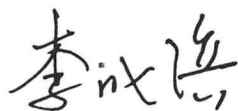
选择是什么？选择首先是提出问题，是人根据客观现实和需要对解决具体问题的方向、目标、方式、方法等自觉的设问和破解，所谓提出问题就是问题成功解决的一半。2007年8月1日，美国明尼阿波利斯市（位于明尼苏达州）密西西比河上的I-35W桥梁坍塌了，坠入河中，事故发生后的第一时间明尼苏达州的交通部通过招标来确定重建单位对桥梁进行重建。虽然其建议书要价最高、交货期也最长，而且还是参加竞标企业中唯一没有在明尼苏达州修建过桥梁的单位，这项价值2.5亿美元的工程最终还是被美国科罗拉多州的一家单位获得。它是如何竞争出线、最终取胜的呢？事后，明尼苏达州交通部负责此项决策的小组透露说，价格和速度不是唯一的考虑因素。参加竞标的另外8家单位事后都抱怨说：“你们从来没有把这些告诉过我们！”“可是，你从来都没有问过我们！”该项目组回应说：“而中标的一家这么做了。”换句话说，这家成功中标的单位获得这样一大单工程的秘诀是提问，而那些竞争者却想当然地认为采购方必定会以价格和速度为主要标准，并据此设计出投标方案。科罗拉多人没有这样去想当然，他们把自己的工作定位成找准客户的需求和需求背后的原因，继而提出问题，并通过这种提问对客观现实存在的各种可能中蕴藏的各种信息，进行筛选、放大和选择，发现重要并具有决定性的信息资源而为我所用，并转换成为现实的生产力。获得信息的一般手段就是提问，而它的对手却仅仅把其看成是一个单纯工程价格报价和工程建设速度的竞标活动，未能从提问这一最简单的手段入手去进行信息检索与利用。这种以不同的认识、态度对待信息检索与利用而获取不同信息的结果，提供了现代条件下不同的信息检索与利用的思维模式、思想方法在商战中被自觉成功应用和被动无效应用，也即不同适应性选择的生动案例。

信息检索与利用，在形式上表现为本书所全面介绍的“从信息集合中迅速、准确的查找出所需信息的程序和方法”的各种实践载体、操作路径、技术手段；在实质上，则是由上述各种实践载体、操作路径、技术手段所体现的那种信息检索与利用的思维模式、思想方法养成和应用的过程，是本书作者特别强调的对读者“自学能力、获取信息的能力、解决问题的能力”和创

新能力”的培养过程，是提升人们现代信息意识观念、信息素质能力的过程。密西西比河 I-35W 桥梁坍塌重建的案例，主要是从思维模式、思想方法的角度，对信息检索与利用技术是现代社会中人们生存、发展的一种基本方式这一本质的说明。以这样的视角来思考，本书的写作与出版，不仅填补了针对建筑类理工院校信息检索教材建设工作上的一个空白，更为重要的是，还在于其从技术的角度，具体揭示了现代社会中人们应该如何运用信息检索与利用的知识和智慧去生存、发展的文化内涵，指出了现代社会发展的一种文化自觉的发展方向。

选择的本质是放弃。在同一时空内，选择了寒窗苦读，就意味着对尘世喧嚣的放弃；选择了登高望远，就意味着对临水思幽的放弃；选择了中央电视台的信息，就放弃了辽宁卫视台信息；选择了东，就背离了西；选择了坚定，就杜绝了动摇……本书的编著者曹丽娟和白首晏同志，在沈阳建筑大学图书馆从事了多年信息检索课程的教学与参考咨询的服务工作，岗位本职并无编写相关教材的工作要求。但是，作为普通教师，她们放弃了工作之余的休息、寒暑假整块休闲娱乐的时间，选择了对大学生如何开展信息检索与利用技术，提高信息素质教育及其方式、方法和有效载体的摸索与探究之路，在长期教学、科研、管理实践中，历经参加多项科研课题研究和反复的教学应用，多方总结、几易其稿，最终形成了集严谨的科学性、新颖的实用性于一身的这本《信息检索与利用》。放弃其实是一种奉献、牺牲。曹丽娟和白首晏两位普通大学教师的这种奉献、牺牲，也许很平凡，可是，这种牺牲是以《信息检索与利用》教材成书出版的方式实现的，是以对信息是构成世界的基本要素、信息是人类认识世界的中介、信息是知识生产的基本原料、信息是科学技术转化为生产力的桥梁与工具等功用做出“信息集合”式的创新论述而实现的，是以对精于钻研的崇高事业心、社会良知的自觉责任感躬身践行而实现的，这就有了不平凡的意义和价值。1984 年国家教育部颁布《印发〈关于在高等学校开设文献检索与利用课的意见〉的通知》后至今，已近 30 年。我相信，作为高校选修课、公共课或必修课教材，《信息检索与利用》的成书出版，对于我国高校，特别是建筑类理工院校，建设形成从大学新生入学阶段掌握大学图书馆的“入门钥匙”，到覆盖本科生文献检索课、研究生专业信息资源查找与利用课的教育教学体系，必定会起到创异推动、创新发展的作用。

“为有牺牲多壮志，敢教日月换新天”。毛泽东当年对英勇献身于中国革命英雄们歌颂和缅怀的壮烈诗句，理应成为当代改革创新、科学发展主旋律的主题词，亦应成为我们为推动改革创新、科学发展而开展和进行信息检索和利用的主题词。由于工作上有直接联系的原因，曹丽娟和白首晏老师嘱我为她们编著的《信息检索与利用》教材的成书出版写上几句话，不揣学识的浅薄，是有文如上，权作序言。“十一”国庆在九月揭开序幕，九月还有许多重要的纪念日，又恰逢新中国第 27 个教师节来临之际，谨一并以此作为对创立新中国的先辈们和当代各位“灵魂工程师”们的致敬！



写于 2011 年教师节前的 9 月 7 日

前言

信息检索与利用是高等学校培养大学生信息素养,提高人才素质的一门重要课程。高校信息检索课程的建设与发展,对培养学生的自学能力、获取信息的能力、解决问题的能力 and 创新能力发挥着重要作用,是当前大学教育的重要内容之一。

自 20 世纪 90 年代以来,信息技术的突破性前进和知识的更新,促使信息检索技术和手段飞速地演变,传统的媒体和检索工具、检索方法不断更新。信息检索教材需及时调整内容和结构,以更好地适应新的信息环境和学生的信息观念及信息需求。

本书重点突出新颖性和实用性,着重介绍最新的、实用的信息检索技术与信息资源类型。在关注通用教材完备性的前提下,对内容进行精简和优化。全书共分三大部分:第一部分介绍信息检索的相关理论,共三章。第一章是信息检索相关概念的界定与辨析,确定研究对象的外延与内涵,在综合国内外研究结果的基础上,提出个人的观点。第二章从整体上介绍了常用的信息源类型,以利于读者对其有一个整体的认识与把握。第二章的结构安排突破了现有模式,省略了人为设置的关系层次,从应用的角度展开铺陈式介绍,更加符合普通用户的信息利用习惯。第三章介绍计算机信息检索理论,以深入浅出的语言加实例讲解,化解专业术语的抽象。第二部分集中介绍信息资源类型及其使用方法,共六章。第二部分是本书的重点论述部分,也是篇幅最长的部分,因此编者从多维的视角进行资源框架组织,各章内容既独立,又不乏联系。此外,在介绍资源使用方法时,作者结合教学实践经验,将检索实例嵌入理论性介绍当中,使平面的介绍生动、立体起来,便于学生接受与掌握。第三部分介绍信息的综合利用,共两章。该部分内容属于信息检索的目的与应用,顺承了前面的章节,使全书结构完整。

本书编写分工如下:第一~四章、第七章、第九章由沈阳建筑大学曹丽娟编写,第六章、第八章、第十章由沈阳建筑大学白首晏编写,第五章、第十一章由曹丽娟、白首晏共同编写;全书由曹丽娟统稿。

辽宁大学李小菲审阅全书,提出了许多宝贵意见,沈阳建筑大学徐亚丰、辽宁大学郑伟和沈阳建筑大学於英德对本书的编写提供了帮助,在此一并表示感谢!

书中疏漏和不足之处在所难免,恳请读者和同仁批评指正。最后衷心希望本书能对读者的信息检索和利用能力的提高有所帮助!

编者

2011 年 8 月

目 录

序言

前言

第一章 信息检索概论	1
第一节 基本概念	1
第二节 信息检索	7
第二章 信息源与信息检索工具	11
第一节 常用信息源	11
第二节 信息检索工具	18
第三章 计算机信息检索基础	20
第一节 计算机信息检索概述	20
第二节 计算机信息检索技术	23
第三节 计算机信息检索方法	28
第四章 图书与期刊信息检索	33
第一节 图书信息检索	33
第二节 期刊信息检索	43
第五章 特种文献信息检索	57
第一节 专利文献检索	57
第二节 标准文献检索	68
第三节 科技报告检索	77
第四节 学位论文检索	85
第五节 会议文献检索	89
第六节 政府出版物检索	94
第六章 重要中文信息检索系统	97
第一节 中国知网	97
第二节 维普期刊资源整合服务平台	106
第三节 万方数据知识服务平台	113
第七章 国外著名信息检索系统	120
第一节 工程索引	120
第二节 Web of Science	125
第三节 联机信息检索系统	134
第八章 图书馆信息资源利用	152
第一节 图书馆概述	152
第二节 图书馆信息资源介绍	154

第三节	图书馆文献资源的选择与利用·····	156
第四节	图书馆服务类型及项目·····	157
第五节	网上图书馆·····	165
第六节	数字图书馆·····	168
第九章	Internet 信息资源检索 ·····	175
第一节	Internet 资源概述 ·····	175
第二节	Internet 信息资源检索 ·····	179
第三节	网络资源评价·····	209
第十章	信息的收集、整理与分析 ·····	212
第一节	信息的收集·····	212
第二节	信息的整理·····	216
第三节	信息的分析·····	217
第十一章	科技论文写作 ·····	220
第一节	科技论文概述·····	220
第二节	科技论文写作·····	222
参考文献	·····	229

第一章 信息检索概论

在介绍何为信息检索之前,我们必须先了解什么是信息。大多数人每天都在使用“信息”、“数据”、“知识”这些概念,并将其视为同义词,认为从概念上严格准确地加以区分似乎没什么必要。当我们说到这些词的时候,说者与听者似乎都能理解对方所指而不太会出现误差。但是,作为学科概念,它们从自身属性来说还是有区别的。

第一节 基本概念

一、信息

(一) 信息的概念

尽管人们每时每刻都在与信息打交道,然而,何为“信息”却众说纷纭,至今仍未出现一个被公众普遍接受的定义,对信息本质的认识也差别甚大。人们站在不同的学科角度,给出信息各种迥然不同的界定。信息一词在不同的词典中有不同、甚至矛盾的解释。《辞海》(1989年版)中解释为:“泛指消息和信号的具体内容和意义”,而《简明不列颠百科全书》(1991年版)中称:“信息论中的‘信息’与信号中固有的意义无关”。《牛津英语词典》对“信息”一词的诠释为:“告知……的行为;某种事实或情况的知识或消息的传播;告知的行为或被告知某事的事实;有关某种事实、专题或事件的所传播的知识;某人被告知的东西;情况、消息”。

控制论创始人维纳(Norbert Wiener)先生给信息下的定义是:“信息是人们在适应外部世界,并且这种适应反作用于外部世界的过程中,同外部世界进行互相交换的内容的名称。”这是迄今为止关于“信息”最重要的定义,但对于这个定义,人们仍存在许多争议:维纳能看到用机械唯物论的概念不可能说明信息的本质,这是对的,但他本人对信息的本质也未能做出明确的回答。

目前学界一种较普遍的表述为:“信息是指具有新内容、新知识的消息。”仔细考察我们会发现,这个概念作为一个科学词汇不够严谨。首先,“新”的程度难以界定;其次,“消息”一词似乎只对人或智能系统有意义,而且与对象的具体内容相关,如果将消息改为信号,如何解释信号又成为问题,信息、信号、消息这几个词的解释存在循环现象。还有一种说法称“信息是物质的基本属性之一,反映物质运动状态和变化程度”,这被称之为客观信息。可是物质的运动状态和变化程度以及其他属性,都可以用物理量来表示,同样的内容可以构成信息,也可以不构成信息,所以,将信息定义为物质的基本属性并不确切。以上各种定义对信息的界定都不十分准确与完整,而且,随着人们对事物认识的逐步深化,概念也会随着发展变化,有时还会发生根本性的变化。

其实,信息这一概念本身是模糊的,许多词汇都传递出信息的意思:数据、知识、文献、智慧等,许多事物如一个名称、一首歌、一幅画或一个想法也包含了一个共同的特征,即称为“信息”的东西。信息究竟是何种事物,它的本质特征是什么呢?人们一直在试图给

出一个确切的答案，但也许是因为信息与人类生活的关系太密切了，是最基础的概念，反而很难为其下一个定义，这也符合逻辑的共识：最基本的概念是不能给予确定的定义的。至少目前为止，作者主张对信息这一概念进行描述，而不强求确定一个准确的定义。

在英语中与“信息”相对应的单词是“Information”，这一词来源于拉丁语“Informare”，在拉丁语中，它的意思是给予形式（give form to）、告知（informing）、形成一种思想（to form an idea of）或描述（describe）的意思。Informare 一词的这些含义构成信息的基本概念，依据“信息”一词的基本含义和用途，我们对其进行这样的描述：

1. 信息是一个过程

当一个人接收信息后，他原有的知识体系就会改变，从这一层意思上来讲，“信息是一种告知的行为或事实；是知识、新闻的交流活动”。控制论创始人、美国科学家维纳（N. Wiener）在他的《控制论》一书中这样描述：“信息就是信息，既不是物质，也不是能量。”“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，同外部世界进行互相交换的内容的名称。”维纳的信息概念是从信息在发送、传输和接收的过程中，客体和接收（认识）主体之间的相互作用来定义的。信息只能在相互作用中体现，交互性是信息的本质之一。

2. 信息是抽象的

信息是指在信息交流过程中被感知到的东西。人是信息交流过程中最活跃的因素，被传递和交流的内容要被人感知和理解才能称为信息。人的感知和理解是一种抽象的思维活动，因此，信息的内容是一种抽象的东西。牛津词典中的解释为：“信息是被交流的关于一些事实、主题、或事件的知识。”韦伯词典的解释为：“用来通信的事实、在观察中得到的数据、新闻和知识。”日本《广辞苑》：“信息是所观察事物的知识。”

3. 信息是物质的

这主要指信息要记录在纸张、泥版等载体上。信息的一个特性是：它是无形的、主观的和概念上的。我们无法触摸到它，也无法用直接的方法测量它。因此，为了交流，信息必须被表达或描述出来或以物理的方式替代表示，如语言、图像、符号、文本等。所有这种表达、描述与替代都是信息的物质形态，信息是包含在其中的抽象内容。牛津词典（Oxford English Dictionary, 1989, vol. 7, p. 946）有这样的描述：“信息一词也用于指特定的事物，比如，数据、文献等概念表示信息的意思，因为它们具有传授知识或交流信息以及说明的功能。”许多事物都可以作为信息的书写和存储设备，比如绳子上打的结、泥板上刻写的符号，所有这些标记毫无疑问都表达一定的信息，现代信息技术提供了更加丰富的信息表达和存储的方式。

（二）信息的特征

“信息既不是物质，也不是能量”，它与我们熟知的材料与能源迥然不同，信息具有如下特征：

1. 客观性

信息不是虚无缥缈的事物，它的存在可以被人们感知、获取、传递和利用。信息是现实世界中各种事物运动与状态的反映，其存在是不以人的意志为转移的。客观性与广泛性是信息最重要的本质特征。

2. 普遍性

信息的普遍性是指信息无处不在、无时不在。信息普遍存在于自然界、人类社会和人类的思维领域中,无论是自然界的四季变换、电闪雷鸣,还是人类社会活动中的语言文字、城墙古堡,无一不是信息的表现形式。

3. 时效性

信息的时效性是指信息的功能、作用和效益都是随着时间的改变而改变的。信息价值的大小与提供信息的时间成反比,时间的延误会使信息的使用价值衰减甚至最后完全消失。例如,时效性很强的天气预报、股票信息、交易信息、科学信息等,如不能适时使用将会失去其原有的价值。在现代社会中,信息的使用周期越来越短,信息的价值实现取决于对其及时地把握和运用。当然,信息的时效性也存在着相反的情形,有些信息资源随着时间的推移可以不断增值,因此,利用者只有把握好时机,才能发挥信息资源最大的效益。

4. 寄载性

信息必须依附于一定的载体(如声波、电磁波、纸张、化学材料、磁性材料等)才能流通和传递。信息可以存储在不同的载体上,但其内容并不因记录手段或物质载体的改变而发生变化。例如关于筑路机械的信息,不论是刊登在报刊上、发布在电视节目中还是存储在光盘数据库中,其信息内容和价值是同样的。

5. 非同一性

信息可以脱离它所反映的事物而被存储和传播。也就是说信息生成之后,可以与产生信息的事物分离,单独被保存和传播。这也就是为什么我们可以通过书籍、录音或影像看到或听到很久以前的信息而不需要再把以前真实的事物保留下来。信息可以与实体分离并被传递的特性是十分重要的,它可以帮助我们了解很久以前的事情及人类无法亲自触及的事情。例如,对遥远的星体的认识,人类只需利用各种仪器接收它们发出的信息(如光波、引力波等)便可以了解它们。

6. 传递性

信息的传递性是指信息可以借助一定的工具或载体进行空间上和时间上的传递。空间传递是指信息的利用不受地域的限制,可以由此地到彼地;时间传递是指信息的传递不受时间的限制,可以由古到今。这种跨越时空的传递特性是实现信息资源共享的基础,是将信息最大化利用的保证。信息的传递主要依靠声、光、电、磁、语言、表情以及文字符号等得以实现。

7. 可塑性

信息在流通和使用过程中,人们借助于先进的技术,可以对其进行分析、综合及加工处理。也就是把信息从一种形式变换成另一种形式,如可以将一本图书加工为题录或文摘等形式,从而方便用户的选择和利用。不过,在信息的加工过程中,信息总量会减少或增加。

8. 共享性

信息的共享性是指信息可以由不同的个体或群体在同一时间或不同时间共同享用。与物质交换不同,如果两个人各有一份信息,互相交换之后,会使两个人各自拥有两份信息。另外,同一份信息可以被使用无数次,但信息的总量不会因使用次数的增多而减少甚至消失。我们对信息的存储、加工和整理主要是为了使信息能够共享。当然,信息的共享性并不排除某些特殊信息的独占性和利用的有偿性,如军事、商业、专利信息等。

9. 无限性

信息是事物的状态、变化和特征的表现，而事物的发展与变化是永无穷尽的，因此，产生的信息必然是无限的。此外，信息资源一旦被保留便可永不消失，而事物的变化又会不断产生新的信息，新旧信息不断叠加，信息的总量是无限的。

(三) 信息的表示形式

信息的表示形式主要有四种，即数据、文本、声音和图像。数据是指计算机能够生成和处理的所有事实、数字、文字和符号等；文本是指书写的语言，文本可以手写，也可以是印刷的；声音是指人们用耳朵听到的信息，如说话的声音、音乐等；图像是指人们能用眼睛看到的信息，如影像、照片等。

(四) 信息与数据、知识、智慧的关系

许多词汇都传递出信息的意思：数据、知识、智慧等，信息与这些概念具有什么样的关系呢？如图 1-1 所示是知识管理学中著名的知识阶层图，它反映了这些相关概念之间的关联。

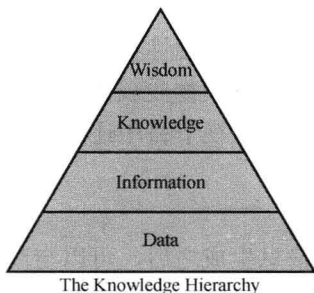


图 1-1 知识管理中的知识阶层图

从图 1-1 中我们可以看到，数据（Data）是知识阶层中最底层也是最基础的一个概念。我们经常说“水的温度是 100℃、木头的长度是 2m、大楼的高度 100 层”，通过水、温度、100℃、木头、长度、2m、大楼、高度、100 层这些符号，我们的大脑里就形成了对客观世界的印象。数据是载荷或记录信息并使之按照一定规则排列组合的物理符号，通过对这些符号的应用，将人类认知的物质世界清晰的描述出来，它可以是数字、文字、图像、声音、视频数据，也可以是计算机代码。数据是中性的，同样一组数据对一个人可能至关重要，但对另一个人可能毫无意义。只有当数据被置于合适的语境中，我们可以对其进行分析 and 处理时才有意义。人们对信息的接受始于对数据的接收，对信息的获取只能通过对数据背景和规则的解读。背景是接收者针对特定数据的信息准备，即当接收者了解物理符号序列的规律，并知道每个符号或符号组合公认的指向性目标或含义时，便可以获取一组数据载荷的信息，即数据转化为信息。

信息（Information）是从原始数据形成知识的一个步骤。信息不仅仅是简单的数据，从另一方面来说，信息是我们互相交流的内容。信息是数据载荷的内容，对于同一信息，其数据表现形式可以多种多样，例如，你可以打电话告诉某人某件事（利用语言符号），也可以写信告诉某人同一件事（利用文字符号），或者干脆画一个图（利用图像符号）。信息是具时效性的、有一定含义的、有逻辑的、经过加工处理的、对决策有价值的信息流，当数据改变了一个人的知识结构时，我们称它是信息。作为知识层次中的中间层，有一点可以确认，那就是信息必然来源于数据并高于数据。我们知道像 7℃，50m，300t，大楼，桥梁这些数据是没有联系的、孤立的，只有当这些数据用来描述一个客观事物和客观事物的关系，形成有逻辑的数据流，他们才能被称为信息。

信息虽给出了数据中一些有一定意义的东西，但它的价值往往会在时间效用失效后开始衰减。只有通过人类利用归纳、演绎、比较等手段对信息进行挖掘，使其有价值的部分沉淀下来，并于已存在的人类知识体系相结合，这部分有价值的信息就转变成知识（Knowledge）。例如，北京 7 月 1 日气温为 30℃。在 12 月 1 日气温为 3℃。这些信息一般会在时效

性消失后,变得没有价值,但当人们对这些信息进行归纳和对比就会发现,北京每年的7月气温会比较高,12月气温比较低。因此,知识是信息接收者通过对信息的提炼和推理而获得的认识,是人类通过信息对事物运动规律的把握,是人的大脑通过思维重新组合的、系统化的信息集合。知识并不简单等同于信息,将信息转化为知识,要增加可用性、重要性、价值、可理解性、观点和逻辑要素。知识存在于人的大脑中,这也解释了一个人为什么要经过漫长时间努力才能够成为专家或长大成人。

从图1-1可知,智慧(Wisdom)是处于知识阶层中最高层的一个概念。我们经常看到一个人满腹经纶,拥有很多知识,但不通世故,被称做书呆子;也会看到有些人只读过很少的书,却能力超群,能够解决棘手的问题,我们会认为后者具有更多的智慧。因此,我们认为智慧是人类基于已有的知识,针对物质世界运动过程中产生的问题,根据获得的信息进行分析、对比、演绎找出解决方案的能力。这种能力运用的结果是将信息的有价值部分挖掘出来,并使之成为已有知识架构的一部分。

概言之,数据是信息的原材料,而信息则是知识的原材料,智慧来源于知识但高于知识。

二、文献

(一) 文献的概念

上一段文字探讨了信息的概念和特点,我们知道信息具有寄载性,即抽象的信息内容必须借助于一定的符号系统记载在物质载体上才能够得以表现和传播。在信息的各种载体中,文献是记录、积累、传播和继承的最有效手段,是人类社会活动中获取信息的最基本、最主要的来源,信息检索也主要指文献检索,即检索文献当中记载的信息内容。

“文献”一词自古有之,最早见于《论语·八佾》篇:“子曰:夏礼,吾能言之,杞不足征也。殷礼,吾能言之,宋不足征也。文献不足故也。足,则吾能征之矣。”在这里,“文献”一词有两层含义,文指典籍,献指人才。朱熹《论语集注》中解释:“文,典籍也。献,贤也。”随着历史的发展,其概念不断变化,现在偏重于“文”,单指典籍,而原来含有的贤人一义逐渐消失了。

国际标准化组织《文献情报术语国际标准》(ISO/DIS 5217)对文献的定义为:“在存储、检索、利用或传递记录信息的过程中,可作为一个单元处理的,在载体内、载体上或依附载体而存储有信息或数据的载体”。我国国家标准《文献著录总则》(GB 3792.1—1983)给文献所下的定义为:“文献是记录有知识的一切载体”。可以看出,文献概念的中外表述在实质上是一致的,都准确地反映了文献的内涵。为了便于理解,我们对其稍加解释,即文献是记录有知识和信息的一切载体的统称,是用文字、图形、图像、符号、声频、视频等技术手段记录知识和信息的物质载体。文献的概念具备三个要素:一是知识或信息内容,这是文献的核心要素,是文献价值的集中体现;二是载体材料,即用于记录知识或信息的物质材料;三是记录方式,指用文字、图形、声音、视频等方式和技术手段将知识或信息记录在一定的物质载体上。

(二) 文献的类型

科学技术的进步促进了文献的发展,现代文献具有数量庞大、类型复杂、文种多样、出版分散、内容重复交叉、新陈代谢频繁等特点。由于文献是最主要的信息来源,因此,为了实现快速查找文献尤其是获取文献原文,我们必须首先了解和掌握文献的类型及主要特点。

依据不同的标准,文献可以被划分成不同的类型,这些文献类型之间往往在内容上有相互交叉,本书选介三种常见的划分方法。

1. 按文献的载体划分

(1) 刻写型文献。刻写型文献以泥板、纸草、甲骨、青铜、简牍、缣帛、纸等书写材料和天然材料为载体,记录方式为刻画和书写。刻写型文献包括印刷术发明之前的古代文献和当今未正式付印的手稿、草稿、日记、信件、原始档案、碑刻等。

(2) 印刷型文献。印刷型文献以纸张为载体,记录方式为油印、铅印、胶印、静电复印等印刷技术,主要指印刷术发明之后印刷出来的文献,包括书、报、刊等。其优点是便于阅读、传播和交流,但存储密度低,存储空间大,很难实现自动化管理和服务。

(3) 缩微型文献。缩微型文献以胶卷等磁性材料为载体,记录方式主要是光学记录技术,主要类型有缩微胶片、缩微平片、缩微胶卷等。其优点是忠于原件、体积小、存储密度高;缺点是阅读不便,需要借助缩微阅读器方能阅读。

(4) 声像型文献。声像型文献以感光材料和磁性材料为载体,记录方式为机录技术、磁录技术和光录技术,主要类型有唱片、录音带、录像带、电影胶卷、胶片、幻灯片等。这类文献可以让人们通过自己的视觉、听觉感受直观、形象、生动、逼真的信息世界,充分体现了其直观性和真实性,但它也需要借助于一定的设备进行阅读。

(5) 电子型文献。电子型文献指采用电子手段并以电子形式存在,利用计算机及现代通信方式提供信息的一种新型载体。电子型出版物内容丰富,类型多样。按照出版物类型划分,主要有电子期刊、电子图书、电子报纸、各类联机数据库以及磁带、软盘、光盘等产品;而如果按照媒体的信息结构组织形式划分,主要有文本型出版物、超文本出版物、多媒体出版物和超媒体出版物等。电子型文献的优点是信息容量大、出版周期短、易更新、方便检索、容易复制、具备可交互性、低成本、可共享等;缺点是需要借助计算机等电子设备阅读。

2. 按文献的出版方式划分

按文献的出版方式划分,常见的文献类型有:图书、期刊、科技报告、会议文献、政府出版物、专利文献、标准文献、学位论文、产品样本、科技档案、报纸、统计资料、图谱、地图、乐谱、广告资料、新闻稿等。在以上文献类型中,除图书和期刊分别作为独立的两种类型外,其他文献类型统称特种文献。所谓特种文献就是介于图书和期刊之间,似书非书、似刊非刊的文献类型,它通常在出版发行方面或获取途径方面比较特殊,因而称为特种文献。比较重要的特种文献包括会议文献、学位论文、专利文献、标准文献、政府出版物和科技报告。

3. 按文献内容的整理加工深度划分

文献按照其内容整理加工的深度,可以分为零次文献、一次文献、二次文献和三次文献。

(1) 零次文献。零次文献又叫灰色文献,主要包括两个方面的内容:一是形成一次文献以前的知识信息,即未经记录,未形成文字材料,是人们的“出你之口,入我之耳”的口头交谈,是直接作用于人的感觉器官的非文献型的信息;二是未公开于社会即未经正式发表的原始的文献,或没正式出版的各种书刊资料,如书信、手稿、记录、笔记以及一些内部使用通过公开正式的订购途径所不能获得的书刊资料。

零次文献具有信息内容新颖、不成熟不定型的特点,在内容上有一定的价值,此外,零次文献还能弥补一般公开文献从信息的客观形成到公开传播之间费时甚多的弊病。由于不公开交流,故零次文献的获取有一定难度,一般是通过口头交谈、参观展览、参加报告会等途径收集零次文献信息。

(2) 一次文献。一次文献是人们直接以自己的生产、科研、社会活动等实践经验为依据生产出来的文献,也常被称为原始文献。一次文献在整个文献集合中是数量最大、种类最多、创新内容最多、使用最广、影响最大的文献,期刊论文、专利文献、科技报告、会议录、学位论文等都属于一次文献范畴。一次文献所记载的知识、信息比较新颖、具体、详尽,具有直接参考、借鉴和使用的价值,是人们检索和利用的主要对象。

(3) 二次文献。二次文献是对一次文献信息进行整理、加工的产品,即把大量的、分散的、无序的一次文献信息资源收集起来,按照一定的方法进行整理、加工,使之系统化而形成的各种目录、索引和文摘。二次文献没有新的知识、信息产生,它只报道某个特定范围的一次文献线索,具有汇集性、检索性的特征。它的重要性在于提供了打开一次文献信息资源知识宝库的钥匙,节省人们查找一次文献所花费的时间。

(4) 三次文献。三次文献是根据一定的目的和需求,在大量利用一、二次文献信息资源的基础上,对有关知识、信息进行综合、分析、提炼、重组而生成的再生信息资源,如各种教科书、技术书、综述、百科全书等都属于三次文献范畴。三次文献具有综合性高、针对性强、系统性好、知识面广的特点,有较高的使用价值,能直接提供问题的答案。

总之,从零次文献、一次文献、二次文献到三次文献,是一个由分散到集中,由无序到有序,由博而精的对知识信息进行不同层次的加工过程。它们所含信息的质和量是不同的,对于改善人们的知识结构所起到的作用也不同。零次和一次文献是最基本的信息源,是文献信息检索和利用的主要对象;二次文献是一次文献的集中提炼和有序化,它是文献信息检索的工具;三次文献是把分散的零次文献、一次文献、二次文献,按照专题或知识的门类进行综合分析加工而成的成果,是高度浓缩的文献信息,它既是文献信息检索和利用的对象,又可作为检索文献信息的工具。

第二节 信息检索

一、信息检索的概念

信息检索 (Information Retrieval) 这一概念是在 1949 年国际数学会议上由 Galvin W. Mooers 首次提出的。检索即查找之意,广义的信息检索包括信息的存储过程和查找过程,而狭义的信息检索仅指信息的查找。本书取其狭义概念,可以这样定义信息检索:从信息集合中迅速、准确地查找出所需信息的程序和方法。这里所说的信息集合指的是有组织的信息整体,它可以是数据库的全部记录,也可以是某种检索工具,还可以是某个图书馆的全部馆藏。一句话概括信息检索的原理为:对信息集合与需求集合的匹配和选择。

信息检索活动可以划分为四个步骤:定义、判断、获取和评价,如图 1-2 所示,而且这是一个需要重复执行的循环过程。

在社会科学化的进程中,信息检索经历了手工信息检索、机械信息检索和计算机信息检索的发展历程。

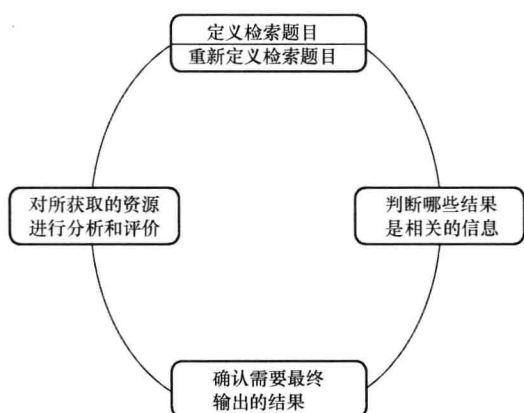


图 1-2 信息检索流程图

手工检索向计算机信息检索的过渡阶段。

机械信息检索系统通过控制机械动作，借助机械信息处理机的数据识别功能代替部分人脑，实现信息检索的自动化。但它并没有发展信息检索语言，只是采用单一的方法对固定的存储形式进行检索的工具，而且过分依赖于设备，检索复杂，成本较高，检索效率和质量都不理想。机械信息检索系统很快就被迅速发展的计算机信息检索系统所取代。

3. 计算机信息检索阶段

计算机信息检索起源于 20 世纪 50 年代初，是指利用计算机查找文献信息的过程。计算机信息检索由最初的单机批处理系统，经过联机计算机检索系统，发展到今天的网络信息检索系统。计算机信息检索的优点在于速度快、耗时少、查全率高、检索入口多等，是目前主要的信息检索方式。缺点是误检率较高，且要依赖一定的设备。

二、信息检索的意义

信息检索的作用与意义主要体现在以下三个方面：

1. 避免重复研究

在研究工作中，任何一个课题从选题、试验直到出成果，每一个环节都离不开信息。研究人员在选题开始就必须进行信息检索，了解别人在该项目上已经做了哪些工作，哪些工作目前正在做，谁在做，进展情况如何等，以使我们避免重复前人（他人）的研究。

在我国专利法实施的初期，上海的一位发明者，花费 20 多年时间潜心研制了一种用于机械运动的“同心轮”。“同心轮”问世后，发明者去专利局申请专利，但经专利局检索，发现这种技术早在 60 年代，也就是在他开始研制的时候，国外就已经有了。而申请专利的技术必须具备新颖性，即提出申请日以前，没有同样的发明或者实用新型在国内外出版物上公开发表过、在国内公开使用过或者以其他方式为公众所知。因此，这位发明人潜心研究了 20 年的成果，最后却无法获得专利权。专利法实施前，辽宁省曾对机械系统 280 个企业的科研情况作过一次调查，重复的科研项目达 340 多个。甚至同居一城，研究同样课题的部门竟然也互不相闻。

通过获取和评价目前已有的研究成果可以帮助我们：衡量即将开展的研究课题当前处于何种水平；发现新的相关信息从而进一步明晰研究题目；在整个研究过程中不断更新自己的知识和见解；确定研究计划和研究方法；产生新的研究设想。

1. 手工信息检索阶段

早期，用户在获取信息时，主要利用图书馆的馆藏目录、印刷型的检索工具（文摘、索引等）和参考工具书，这些工具都是卡片型或者印刷型的，检索的方式基本上是以手工翻看为主。手工检索的优点是具有概念思维性和随机应变性，用户在检索时大脑不断思考与判断，可以随时调整检索策略。但手工检索比较耗费时间，而且查全率比较低。

2. 机械信息检索阶段

机械信息检索系统始于 20 世纪 40～50 年代，利用各种机械装置进行信息检索，是