

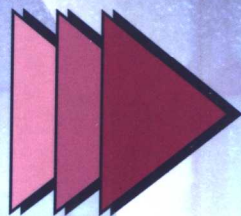
高等学校教材

信息存储与检索

张景元



高等教育出版社
HIGHER EDUCATION PRESS



高等学校教材

信息存储与检索

张景元

高等教育出版社

内 容 简 介

本书介绍了信息存储与检索的基本理论知识、原理与方法;介绍了计算机技术和网络技术在信息存储与检索中的应用、最新成果和发展趋势;介绍了工程索引、科学引文索引、科技会议录索引、化学文摘、金属文摘、机械工程文摘、计算机文摘、数学评论等国外主要信息检索工具的基本知识;讲述了 DIALOG、STN、《万方数据资源系统》、维普信息资源系统、专利信息检索系统、中国学术期刊数据库及国内主要科技信息服务机构的基本内容。

本书内容新颖、丰富,理论与实践密切结合,可作为高等学校理工科专业信息存储与检索课程的教学用书,亦可作为其他科技人员提高信息检索能力的参考书。

图书在版编目 (CIP) 数据

信息存储与检索/张景元. —北京:高等教育出版社, 2004.12

ISBN 7-04-014623-1

I.信… II.张… III.①信息存贮—高等学校—教材②情报检索—高等学校—教材 IV.①TP333②G252.7

中国版本图书馆 CIP 数据核字 (2004) 第 122726 号

策划编辑 陈红英 责任编辑 陈红英 特约编辑 丁 波
封面设计 李卫青 责任印制 孔 源

出版发行 高等教育出版社
社 址 北京市西城区德外大街 4 号
邮政编码 100011
总 机 010-58581000

购书热线 010-64054588
免费咨询 800-810-0598
网 址 <http://www.hep.edu.cn>
<http://www.hep.com.cn>

经 销 新华书店北京发行所
印 刷 北京四季青印刷厂

开 本 787×1092 1/16
印 张 17.5
字 数 400 000

版 次 2004 年 12 月第 1 版
印 次 2004 年 12 月第 1 次印刷
定 价 23.00 元

本书如有缺页、倒页、脱页等质量问题,请到所购图书销售部门联系调换。

版权所有 侵权必究

物料号: 14623-00

前 言

信息检索是一门既古老又年轻的技术,它伴随着科学技术的发展而发展,是科技工作者和工程技术人员必须具备的基本技能之一。早期的信息检索工具多数是针对印刷式的,通过手工进行检索。电子计算机诞生以后,20世纪50年代开始用于科技信息检索,随着计算机技术和网络通讯技术的不断发展,信息存储与检索理论、方法有了新的发展,在实现技术上也从脱机检索、联机检索发展到今天的计算机网络检索。经过40多年的研究与发展,信息检索逐步形成了一门新的交叉学科。它的发展与完善为科技工作者及社会提供了极大的方便,有效地促进了科教事业和社会的发展。

20世纪80年代以前,我国几乎没有高校开设文献检索课程,随着科教事业的不断发展,作为新时期的大学生必须得到科技信息检索技能的训练。国家意识到了问题的重要性,原国家教育委员会于1984—1992年先后颁发了4个有关文件,从此我国高等教育规划体系中增设了“文献检索与利用”这门课程。到20世纪90年代中期,大部分院校不同程度地开设了“文献检索”相关课程,进入21世纪以来,几乎所有的高校都开设了此课程。然而,与蓬勃发展的信息存储与检索技术相比,教材建设相对滞后,其内容较为陈旧,形式单一,无法适应实际教学的要求,尤其不适合非图书馆专业广大理工科学生学习。从20世纪90年代末开始,先后出版了几套比较实用的教材,从某种程度上缓解了这一矛盾。

计算机信息存储与检索技术发展速度惊人,信息存储方法与技术不断更新,各类信息资源数据库不断出现,并且发展迅速。信息检索方法更不断更新,新检索工具软件与检索系统不断涌现,更新换代速度很快(只要与计算机技术相关皆如此),基于Internet的信息检索技术已经相当成熟。网络环境下的信息检索所具备的多样性、灵活性远远超出了传统意义上的信息检索,除此之外,网络是个信息海洋,在这个浩瀚的信息海洋中,如何能够准确、及时、有效地找到、获取与自身需求相关和适用的信息是十分重要的。网络信息检索代表着在网络信息环境下发展起来的一种新型检索模式,而且已发展成为信息检索领域的一个重要分支。

本书主要针对理工科学生而编写,整个体系由三大模块组成:基本理论与基础知识、主要检索工具与检索系统、网络资源与利用。全书围绕着科技信息的产生、存储与利用展开,阐述了信息存储与检索的基本理论、概念、技术和方法,检索工具与检索系统。以计算机技术和Internet技术在信息存储与检索中的应用为主线,紧跟信息资源数据库和计算机(网络)信息检索系统的发展潮流,力求体现该领域的最新技术和最新成果。全书共分10章:第1~3章主要讲述了信息及其信息存储与检索的基本概念、知识与原理;第4~6章主要讲述了现代信息存储技术、存储系统的构成、数据库技术及其在信息检索系统中的应用;第7章主要讲述了一些专业性检索工具、信息检索系统及其使用方法;第8章重点讲述了《万方数据资源系统》、维普信息资源系统、专利检索系统的资源及其使用方法,同时介绍了国内

其他主要信息资源数据库的检索方法和信息服务机构;第9章主要讲述了 Google、Yahoo!、天网搜索等网络搜索引擎及其检索方法和免费资源的使用;第10章主要讲述了中国数字图书馆、超星数字图书馆、中国学术期刊网及其主要资源和检索方法。

在本书的编写过程中,参考了国内外大量的相关专著、教材、论文和网站,直接和间接地引用了许多学者的有关观点与信息,在此对他们表示诚挚的感谢。在整个编写过程中,得到了高等教育出版社陈红英老师的大力指导和帮助,在此深表谢意。谨此对在编写过程中给予帮助和支持的所有单位和个人一并表示深深的感谢。

由于编者水平有限,错误和不当之处在所难免,敬请专家和读者批评指正。

作者

2004年8月

目 录

第1章 概论	(1)	3.4.2 间接检索	(29)
1.1 科技信息的特点	(1)	3.4.3 间接检索原理	(30)
1.2 信息检索的作用和意义	(3)	3.5 信息检索方法	(31)
1.2.1 信息检索的定义	(3)	3.5.1 信息检索的具体方法	(31)
1.2.2 信息检索的作用	(3)	3.5.2 信息检索的步骤	(32)
1.2.3 信息检索的意义	(5)	3.6 检索性能的评价	(33)
思考题	(6)	3.6.1 查全率和查准率	(33)
第2章 信息资源及其分类	(7)	3.6.2 查全率的估计	(34)
2.1 基本概念	(7)	3.6.3 检索性能评价曲线	(35)
2.1.1 信息	(7)	思考题	(35)
2.1.2 知识	(8)	第4章 信息存储技术	(36)
2.1.3 文献	(9)	4.1 信息存储的发展历史与现状	(36)
2.2 信息资源的类型	(10)	4.2 印刷存储	(38)
2.3 文献信息的类型	(12)	4.3 缩微存储	(39)
2.3.1 印刷文献	(12)	4.3.1 缩微存储技术的发展简介	(39)
2.3.2 缩微文献	(21)	4.3.2 COM 的工作原理	(40)
2.3.3 音像文献	(22)	4.3.3 缩微品的存储介质及特点	(40)
2.3.4 机读文献	(22)	4.3.4 缩微品的特点及检索	(40)
2.4 信息的层次类型	(22)	4.4 信息的磁存储	(41)
2.4.1 零次信息及零次文献	(22)	4.5 信息的半导体存储	(42)
2.4.2 一次信息及一次文献	(23)	4.5.1 RAM	(42)
2.4.3 二次信息及二次文献	(23)	4.5.2 ROM	(42)
2.4.4 三次信息及三次文献	(24)	4.5.3 主要技术指标	(43)
思考题	(24)	4.6 信息的光存储	(44)
第3章 信息传播与检索	(25)	4.7 信息铁电存储系统	(45)
3.1 信息的产生过程	(25)	思考题	(47)
3.1.1 确定课题阶段	(25)	第5章 信息检索系统	(48)
3.1.2 科学实验阶段	(26)	5.1 信息系统及其种类	(48)
3.1.3 发表论文阶段	(26)	5.1.1 信息系统的特点	(49)
3.2 信息的传播渠道	(27)	5.1.2 信息系统的种类	(49)
3.2.1 正式传播渠道	(27)	5.1.3 信息系统的发展趋势	(49)
3.2.2 非正式传播渠道	(27)	5.2 信息检索系统的发展	(51)
3.3 信息的传播与检索过程	(28)	5.2.1 信息检索工具与信息 检索系统	(51)
3.3.1 信息的传播过程	(28)	5.2.2 信息检索系统的发展过程	(51)
3.3.2 信息的检索过程	(28)	5.2.3 我国信息检索系统的发展 及现状	(54)
3.4 信息检索原理	(28)		
3.4.1 直接检索	(28)		

5.3 信息检索系统的类型	(57)	6.4.4 信息的标引与采集	(102)
5.4 中文书本式检索工具	(59)	6.4.5 系统软件设计	(102)
5.4.1 中文检索工具体系	(59)	思考题	(103)
5.4.2 中文文摘类主要期刊	(60)	第7章 国际主要信息检索系统	(104)
5.4.3 我国出版的常用国外报刊 索引与文摘	(63)	7.1 《工程索引》	(104)
5.5 计算机信息检索系统	(65)	7.1.1 基本概况	(104)
5.5.1 计算机信息检索系统的组成	(65)	7.1.2 Ei 的出版形式	(105)
5.5.2 计算机信息检索系统的工作 原理	(66)	7.1.3 Ei 的著录格式	(106)
5.5.3 计算机信息检索系统的类型	(66)	7.1.4 Ei 的检索实例(手工检索 实例)	(107)
5.5.4 检索策略与技术	(67)	7.1.5 Ei 网络检索系统	(108)
5.5.5 制定检索策略的步骤	(68)	7.1.6 EngineeringVillage 检索	(110)
5.5.6 计算机信息检索系统的服务 方式	(69)	7.1.7 Ei Compendex Web 收录的中国 期刊名录	(115)
思考题	(70)	7.2 《科学引文索引》	(119)
第6章 数据库在信息检索系统中的 应用	(71)	7.2.1 基本概况	(119)
6.1 数据库的发展过程	(71)	7.2.2 SCI 的出版形式	(120)
6.1.1 历史的回顾	(71)	7.2.3 SCI 的著录格式	(121)
6.1.2 数据库的分类	(72)	7.2.4 SCI 的检索途径	(125)
6.1.3 实时数据库	(73)	7.2.5 SCI 检索实例(手工检索 实例)	(127)
6.1.4 移动数据库	(74)	7.2.6 SCI 网络检索系统	(127)
6.1.5 时态数据库	(75)	7.2.7 Web of Science 2002(SCI、SSCI、 A&HCI)年数据库收录国内 期刊名录	(133)
6.1.6 面向对象数据库	(75)	7.3 《科技会议录索引》	(137)
6.1.7 分布式数据库	(76)	7.3.1 基本概况	(137)
6.1.8 并行数据库系统	(76)	7.3.2 ISTP 的检索	(138)
6.1.9 演绎数据库	(77)	7.4 DIALOG 系统	(140)
6.1.10 模糊数据库	(77)	7.4.1 基本概况	(140)
6.1.11 智能数据库系统	(78)	7.4.2 联机方式	(141)
6.1.12 数据仓库	(78)	7.5 STN 系统	(147)
6.2 传统数据库在信息检索系统中 的应用	(81)	7.5.1 概况	(147)
6.2.1 文献型数据库	(81)	7.5.2 STN 系统的联机方式	(148)
6.2.2 非文献型数据库	(84)	7.5.3 STN 系统的功能	(149)
6.3 新型数据库在信息检索系统中 的应用	(87)	7.6 常用专业性检索工具简介	(150)
6.3.1 多媒体数据库	(88)	7.6.1 美国《金属文摘》	(150)
6.3.2 非结构化数据库	(94)	7.6.2 美国《机械工程文摘》	(152)
6.4 信息检索系统的设计步骤	(98)	7.6.3 美国《化学文摘》	(153)
6.4.1 信息检索系统的构成	(98)	7.6.4 美国《生物学文摘》	(154)
6.4.2 数据库的设计	(99)	7.6.5 英国《计算机文摘》	(155)
6.4.3 信息的组织	(99)	7.6.6 美国《环境文摘》	(155)
		7.6.7 美国《应用力学评论》	(156)

7.6.8 美国《数学评论》	(156)	9.1.1 我国网络建设现状	(214)
7.6.9 美国《情报科学文摘》	(157)	9.1.2 我国网络资源现状	(215)
7.6.10 美国《商业期索引》	(157)	9.2 网络搜索引擎概述	(216)
思考题	(158)	9.2.1 搜索引擎发展简史	(216)
第8章 国内主要信息检索系统	(159)	9.2.2 搜索引擎的基本结构及工作 原理	(218)
8.1 我国信息检索系统发展概况	(159)	9.2.3 搜索引擎的分类	(221)
8.2 《万方数据资源系统》	(162)	9.3 主要网络搜索引擎介绍	(222)
8.2.1 基本情况	(162)	9.3.1 Google	(222)
8.2.2 主要资源	(165)	9.3.2 Yahoo!	(226)
8.2.3 检索方法	(167)	9.3.3 天网搜索	(230)
8.2.4 服务方式	(179)	9.4 搜索引擎的发展趋势	(235)
8.3 《维普信息资源系统》	(181)	9.4.1 目前搜索引擎存在的主要 问题	(235)
8.3.1 基本情况	(181)	9.4.2 人工智能技术在搜索引擎中 的应用	(237)
8.3.2 主要信息资源	(181)	9.5 网上主要免费资源	(238)
8.3.3 中文科技期刊数据库的 检索步骤	(183)	9.5.1 WWW 网页	(239)
8.3.4 VIP 全文浏览器	(192)	9.5.2 地理信息	(239)
8.4 《专利信息检索系统》	(192)	9.5.3 免费专利信息	(239)
8.4.1 专利文献的种类与特点	(192)	思考题	(243)
8.4.2 国家知识产权局专利检索 系统	(196)	第10章 网络数字图书馆	(244)
8.4.3 中国专利网	(198)	10.1 中国数字图书馆	(244)
8.5 北京文献服务处	(200)	10.1.1 概述	(244)
8.5.1 概述	(200)	10.1.2 中国数字图书馆信息资源	(245)
8.5.2 主要信息资源数据库	(201)	10.1.3 网上读书服务	(249)
8.5.3 检索步骤	(203)	10.2 超星数字图书馆	(252)
8.5.4 服务方式及收费	(203)	10.2.1 概述	(252)
8.6 《中国高等教育文献保障系统》	(206)	10.2.2 检索方法	(252)
8.6.1 概述	(206)	10.3 中国期刊网数字图书馆	(255)
8.6.2 CALIS 的服务功能	(208)	10.3.1 概述	(255)
8.6.3 CALIS 的子项目建设	(209)	10.3.2 信息资源	(255)
8.7 中国科学技术信息研究所	(210)	10.3.3 检索方法	(256)
8.7.1 基本情况	(210)	思考题	(263)
8.7.2 主要资源及服务	(211)	主要参考文献	(264)
思考题	(213)	参考网站	(267)
第9章 网络搜索引擎及网络资源利用	(214)		
9.1 我国网络建设及网络资源现状	(214)		

第1章 概 论

信息检索是一门古老的技术,它随着科技工作的开展而产生,不论过去还是将来,都是工程技术人员及科技工作者必须具备的基本技能之一,其作用及重要性无须赘述,只是在不同时期所采用的方法与手段不同而已。1946年第一台电子计算机诞生,从20世纪50年代就开始将之用于信息检索,从而产生了计算机信息检索的相关理论与技术。50年来,随着计算机技术和通信技术的不断发展,计算机信息检索在理论和实践两方面都得到了极大的发展,在世界范围内相继出现了联机检索系统和检索终端。20世纪90年代,Internet技术得到了较大发展,各国家相继建设或正在筹建信息高速公路,使得计算机信息检索技术又有了新的发展,联机检索方法已基本被淘汰,逐步以Internet新方法所取代。

1.1 科技信息的特点

现代科学技术的发展速度之快,用“日新月异”来形容也不过分。科学技术的发展具有连续性和继承性,同时伴随着科技情报的利用和新的科技情报产生,在科研人员的整个业务时间中,查找资料和阅读文献的时间大约占50%,可见科技信息在科学研究中的重要作用。

目前,国外一些专家、学者关于科技信息的论述有3种观点:第一种为科技信息资源论,即把科技信息看成是第二资源、无形的财富、解决问题的钥匙;第二种为科技信息经济论,即认为研究是高价的,成果是昂贵的,信息是便宜的,它们的比例是100:10:1;第三种为科技信息时代论,即认为欧、美、日之间的技术水平旗鼓相当,相距如一纸之隔,一点就破,谁得到最新信息,谁就占据市场优势,谁得到最先进的技术,谁就可以很快超过其他国家。这表明,信息已经成为一种战略武器,谁掌握了它,谁就掌握了主动权,甚至可以用它作为一种制裁手段,制约对方的发展。

尽管以上观点的侧重点不同,但对科技信息的重要性的认识是一致的。科技信息是科学技术的直观体现,它反映了科学研究和生产技术领域的成果、经验、动态、政策、规划等。它包括科学信息和技术信息两部分。科学信息一般是指自然科学方面的信息,与基础理论研究关系密切;技术信息主要指应用技术方面的信息,与生产部门关系密切。科技信息是科技人员确定科研方向、完成科研项目、评价科研成果的重要保证,也是制定科学技术政策和经济发展政策的重要依据。科技信息的主要内容如下:

- (1) 科技成果,即已经完成的科学技术研究成果。
- (2) 科技工作和科研管理的经验教训。
- (3) 科技动态,包括科技进展情况、预测性论断、科技成果推广情况、科技合作情况等。
- (4) 科技政策、科技规划与计划。

第二次世界大战以后,科技信息发展速度之快,积累规模之大,包含内容之多,作用影响之广,是前所未有的,科技信息这种发展态势显示出如下特点。

1. 文献数量巨大

随着科学技术的迅速发展,科学技术成果增长迅速,同时记载科学技术成果的科技文献也急剧增加,从而出现了“信息大爆炸”的局面。20世纪60年代有资料显示,“近30年来,人类所取得的科技成果,即科学新发现和技术新发明的数量,比过去2000年来的总合还要多。”在20世纪80年代曾有人估算过,截至1980年,人类社会获得的科学知识的90%是第二次世界大战以后30余年获得的,到了20世纪末,几乎又翻了一番。从科技知识数量的翻番周期来看,19世纪翻一番花了50年的时间,20世纪初花了20年的时间,20世纪50~60年代用了10年时间,20世纪70~80年代只用了5年时间,进入20世纪90年代,尤其是在网络环境下,不到2年就会翻一番。19世纪初,全世界只有几十种期刊,到了20世纪60年代增加了近1000倍,达到40000多种,到20世纪八九十年代增长更快。据有关资料不完全统计,目前全世界每年出版的各种图书可达六七十万种之多,出版的科技期刊也有数十万种,发表的科技论文达数百万篇,申请专利多达上百万件。科技文献总量每年以40%~50%的速度增长,其中计算机技术领域的文献大约以60%的速度增长实现年增长。

2. 涉及面广,文献分布分散

传统的学科界限不断被打破,学科越分越多,越分越细。与此同时,综合学科、边缘学科和交叉学科大量出现,这些学科的出现和发展不仅填补了传统学科之间的空白,而且加强了传统学科之间的联系,促使它们相互渗透,将现代科学技术紧密地联结成一个有机的整体。各学科相互联结的同时也造成了科技文献异常分散的情况出现,据不完全统计,一个传统学科应用的技术50%是从别的学科来的,一个新兴学科应用的技术有70%~80%是从其他学科来的。同一专业的文献往往分散在许多不同的刊物上,举例来说,计算机方面的文献仅有40%刊登在本专业的刊物上,其余的60%则分布在机械、电子、电工技术、化工、物理、生物等学术刊物上。这种状况给搜集信息资料、查阅文献的工作带来了极大的困难,如果用传统的手工方法在浩如烟海的文献资料中查找所需要的科技信息,就好像在大海里捞针一样困难。

3. 更新速度加快,文献寿命缩短

在现代科学技术的发展中,时间和速度具有非常突出的意义。随着科学技术的发展,科学技术应用周期缩短,新技术、新产品过时的速度越来越快,科技信息的寿命也越来越短。随着科研周期的缩短,使得科技信息新陈代谢的速度加快了,文献的失效性严重,知识的更新速度一年比一年快,“信息爆炸”与信息利用之间的矛盾越来越尖锐。于是以科技信息为对象,以计算机为基础,研究科技信息的理论和方法的学科——计算机信息检索技术应运而生。

4. 对经济、社会的影响日益显著

人类社会的发展历史就是一部生产力进化史,生产力的每一次进步,都离不开人类的知识,特别是科学技术的发展。科学技术的每一次进步,都在深深地改变着世界的面貌。在信息化的社会中,科学技术起着核心和关键作用。科技信息直接体现了科学技术,反映了科学研究和生产技术领域的成果与经验等,成为科学技术发展的源泉和推动力。科技信息是各种社会资源中最重要、最活跃的因素,它面广量大、知识含量高、利用价值大、需求频次高、更新速度快,在促进社会发展和科技进步中,起着举足轻重的作用。

1.2 信息检索的作用和意义

随着科学技术的高速发展,产生了大量的、具有反复使用价值的科技信息。据统计,在科技信息的传播过程中,80%的时间用来获取信息,20%的时间用来加工和产生新的信息。在这种情况下,怎样有效地获取所需的科技信息,是每个科技工作人员的首要任务,这就要求掌握信息检索的方法和技能。

1.2.1 信息检索的定义

“信息检索”一词由莫尔斯(Calvin W. Mooers)于1949年首次提出,并且把它定义为一种“延时性通信形式”,“在时间上从一个时刻通往一个较晚的时刻,而在空间上可能还在同一点”。这种看法的出发点是从整体上把人类的信息传递看成是一种通信形式,即发生在当代人之间或是当代人与后代人之间的一种思想、文化、科学上的交流。而信息检索是实现信息传递的一个重要环节,它本身就意味着人类的“通信”。在某种程度上,信息检索可以和通信理论相比较,例如,在通信理论中有噪声干扰的情况下探测信息脉冲是否存在的问题与从范围广泛的知识海洋中找出适用信息的问题相类似。所以从本质上来看,信息检索就是一种通信。

也有人从信息处理的角度来理解信息检索。认为信息检索主要是围绕信息检索系统如何处理信息和信息结构这一基本问题,来探讨数据库管理系统、联机检索系统等所需的硬件和软件技术,以及信息检索在各个方面的应用等问题的。这种观点把信息检索看成是计算机科学的一个分支,认为信息检索已经超出了一般意义上的文献检索的范围,还包括银行、医院、国民经济等部门中的管理信息,甚至包括经过分析和重新组织的各学科专家的知识。

还有人仅从文献信息查找的角度来表述信息检索的定义。他们研究记录下来的知识(文献)如何进行组织,使之才能有针对性地被信息用户查询和利用,认为信息检索是从大量的文献中查询与信息提问中所指定的课题(对象)有关的文献的过程。在查找过程中,计算机信息处理只是实现机械化与自动化检索系统的一种手段。下面的定义就是从这种观点出发的:

信息检索(Information Retrieval)(信息检索是现在的称谓,以前称之为情报检索)是指将信息按一定的方式组织和存储起来,并根据信息用户的需要找出有关的信息的过程和技术,所以,它的全称应该为“信息存储与检索”(Information Storage and Retrieval)。信息存储与检索是广义的信息检索,它包括信息汇集、存储与信息检索两个过程,而狭义的信息检索则仅指有序化知识信息的检索查找,换句话说,就是在浩瀚的信息海洋中,有的放矢,查找所需要的、有价值的或能够帮助解决问题的知识,即是从众多科技信息中查询并获取所需资料信息的过程和方法,相当于人们常说的情报查询(Information Search)。

信息检索是经济生活和科学研究的一个组成部分,信息检索方法与技能在人才培养、管理和科研、生产与建设、经济技术和科学知识创新等活动中有着重要的作用。

1.2.2 信息检索的作用

信息检索是打开人类知识宝库的钥匙,它能帮助人们获取知识和利用知识,并使知识转

化为生产力,创造出更多社会财富和物质财富。人类的知识是一代接一代承接和积累起来的,历代流传下来的和目前正源源不断地涌现的信息资源是一个巨大的知识宝库。在德国柏林图书馆的大门上刻着这样一行字:“这里是人类知识的宝库,如果你掌握了它的钥匙,那么这里的全部知识都是你的。”这里所说的“钥匙”就是信息检索,掌握了这把钥匙,就能打开人类知识宝库,获取科研所需要的全部知识,可见,信息检索对于获取信息是多么的重要。信息检索的作用主要表现在如下几个方面。

1. 有利于减少重复劳动,提高科研成功率

科学技术史表明,科技发展的重要前提是积累、继承和借鉴前人的成果。没有继承和借鉴,就不可能有提高和创新。没有交流和综合,就没有发展。在当代物质条件下,科学上的继承、借鉴、交流和综合主要是通过信息检索所提供的途径来实现的。

任何一个科研项目,从选题立项、实际研究到成果鉴定,每一步都离不开信息。只有充分掌握有关信息,才能免除重复,少走弯路,保证科研的高起点、高水平,缩短研究周期,获得预期效果。任何科学研究和科学发明都必须利用现有的科技信息,否则,就会使科研工作重复而造成巨大的损失。根据有关资料报道,在20世纪五六十年代,美国、前苏联、日本等国家由于重复研究而造成的损失高达上百亿美元,仅美国的兵工系统每年损失多达10亿~13亿美元。在原欧共体的研究开发中,由于信息不灵,忽视已有的成果,使大约20%的研究工作重复或不能受益。我国也有类似情况,有约40%的科研项目与国外同行重复,每年造成的损失在十几亿人民币以上,这是一种人力、物力、财力的极大浪费。可见,由于信息意识薄弱、检索手段落后带来的损失是多么的严重。

2. 有利于节省科研时间,提高科研效率

随着科学技术的迅速发展,科技信息急剧增加。科技人员为了在前人已有的成果基础上开展研究工作,必须查找本领域内的已有文献,面对数量庞大的信息,依靠传统的手工方法,既难查得准又难查得全,耗费了大量的时间和精力。同时,由于现代科学技术交叉渗透,使得信息的专业性质不十分固定,这也增大了获取信息的难度。据调查,科技人员花在查阅文献上的时间要占总科研时间的40%~50%,如果能将这部分时间降到最低,使科技人员把主要的精力和时间用于构思和研究,就相当于增加了科技人员,或延长了科技人员的寿命。信息检索能够有效地缩短科技人员搜集信息的时间,进而缩短科研周期,提高科研效率,达到多出成果、快出成果的目的。

3. 有利于培养复合型、开拓型人才

由于现代科学技术的发展日新月异,随着时间的推移,旧知识不断被新知识所代替。与此同时,科研成果从发明到推广应用的周期大大缩短,知识的有效期不断缩短。据估计,科技人员所具有的科技知识12.5%是在大学获得的,87.5%是在工作岗位上学习积累的。另据估计,如果大学毕业后5年之内不补充新知识,原有知识的50%便会陈旧失效;10年之内不补充新知识,原有知识的100%便会陈旧失效。这表明,科技人员的知识绝大部分是在实践中学习积累的,如果能将学生从静态知识的学习引向动态知识的学习,使他们掌握一种获取新知识的方法和技能,随时补充、更新知识,就能适应科学技术和生产发展的需要。

4. 有利于为决策提供科学依据

虽然科技信息本身不能保证决策正确无误,但它是决策的基础。一个国家、地区或组织

要发展什么,限制什么,引进什么,都需要有准确、可靠和及时的科技信息作为依据,才能做出正确的决策。这里举个例子,在20世纪70年代末,荷兰飞利浦公司推出数码激光唱片,这项突破性的音响技术吸引了欧美诸多大公司纷纷投入巨资建厂生产。日本在得知这条信息后,经过仔细研究分析,做出了不放弃原已占领的磁带市场的决策。他们研制出效果更佳、功能更强的数码录音带及配套设备,使有些激光唱片公司刚刚投产或刚完成庞大的基建工程便面临严峻挑战。可见,如果决策者重视和善于利用科技信息,就有可能避免重大损失,还有可能先人一步,从中获益。事实证明,不仅科技人员需要科技信息,计划、管理、决策部门同样也需要科技信息。

1.2.3 信息检索的意义

科学技术的发展具有连续性和继承性特点,今日的社会文明由昨日的社会文明积淀、演变而来,而今日的社会文明又孕育着明日的社会文明。

无论是科学发现,还是技术发明,科技人员都必须在前人已有成果的基础上研究、再创造。据统计,科研工作中出现的各种问题,95%~99%是通过借助他人的经验解决的,只有1%~5%的内容是靠研究者本人的创造性劳动完成的。因此,及时、准确、系统地掌握与本研究领域有关的新理论、新方法、新技术、新工艺、新材料以及成功的经验和失败的教训等信息对任何一个科技人员来说都是至关重要的,而信息检索正是获取和掌握信息的有效途径。

现代科学的趋势是,学科越分越细,新学科不断涌现,研究领域越来越专、越来越窄。另一方面,学科综合化趋势日益突出,交叉学科、边缘学科、综合学科层出不穷,不同学科之间相互渗透、相互配合、相互促进、共同发展,已经成为现代科学技术发展的规律,没有哪一门科学技术可以脱离科学技术的整体水平去独立发展。这种新趋势给信息检索带来了更大的复杂性,在浩如烟海的信息面前,要想迅速、准确、全面地获取能够满足特定需要的有用信息,的确是越来越困难了。

信息的重要性在于它的获得和利用,也就是说,在未被获得和利用之前,信息是没有任何价值的。但是,信息的重要性和需求与信息的获取和利用之间又存在着严重的矛盾。随着大批研究成果的涌现,科技信息的数量急剧增长,如果不能合理解决数量庞大的信息与用户特定需求之间的矛盾,就无法避免重复劳动,就会走弯路,就必然造成人力、物力、才力和时间的浪费。另一方面,与迅速增长的信息相比,受到客观条件的限制,人们吸收和利用信息的能力并未得到相应的提高,面对大量的信息仍然不知所措,许多有价值的信息还没有被发现和利用就自生自灭了。在这种背景下,信息检索就显得更加重要了,因此,掌握、利用信息检索的方法和技术是每一个科技人员所必须的。

实际情况表明,需要科技信息的科技人员总是远比获得科技信息的科技人员为多,其原因是:需要科技信息的人员中相当大的比例缺少必须的科技信息检索的基本知识和技能。许多科技人员之所以在没有科技信息或信息不足的情况下开展工作,不是他们不知道科技信息的重要性,而是不知道如何获得所需要的科技信息。

信息检索对一个国家的建设和发展具有重大的意义,它可以提高一个国家利用现代科学技术成就的能力;能够及时利用适合自己国情的知识和情报,使一个国家在科学研究和生产建设等方面所做的努力更加合理化;可以为解决国计民生的重大问题提供广泛的知识基础;可以提供解决科学技术问题的新方法和新途径、拟定解决问题可选择的办法;可以提高

生产部门和服务部门的工作效果和工作效率;有助于各部门的各级领导做出正确的决策和进行科学管理。

思 考 题

1. 科技信息的特点为何?
2. 信息检索的作用和意义为何?

第2章 信息资源及其分类

2.1 基本概念

2.1.1 信息

信息(情报)是当今世界使用频率最高,也是最“时髦”的词语之一,例如,信息时代、信息资源、信息技术、信息产业、信息社会和信息爆炸等。信息无处不在,无时不有,它既存在于自然界,也存在于人类社会;既来自于物质世界,也来自于精神领域。

英文、法文、德文中的信息都来源于拉丁文 Information,译成中文时,作为广义的理解,被翻译成“信息”,20世纪,我国在科技领域多将之译为“情报”。那么,到底什么是信息呢?它的内涵和外延又是怎样的呢?半个多世纪以来,科学界一直对它的定义进行积极的探讨。关于它的定义不下百种,从不同的侧面反映了它的某些特征,但是至今尚无一种定义因其内涵的全面性和科学性而被社会各界一致所接受,下面是有关“信息”的几种较具代表性定义。

(1) 信息论的创始人香农(C. E. Shannon)从信息论的角度给“信息”下的定义是:“信息是在通信中任何可逆的重新编码或释译中那些保持不变的东西”,“信息就是用于消除随机不定的东西”。

(2) 英国的信息(情报)专家布鲁克斯(B. C. Brooks)则认为:“信息是使人原有的知识结构发生变化的那一小部分知识”。

(3) 前苏联信息(情报)专家米哈依洛夫给“信息”下的定义是:“信息(情报)是作为存储、传播和转换的对象的知识”。

(4) 控制论创始人维纳(N. Wiener)则认为:“信息是我们适应外部世界并且在使这种适应外部世界的过程中同外部世界进行交换的内容的名称”。这是从通信的角度,考虑机器和动物体内的通信,再转到动物体之间的通信给信息下的定义。

(5) 另一种较有广泛影响的定义为:“信息是指用文字、数据或信号等形式通过一定的传递和处理,来表现各种相互联系的客观事物在运动中所具有的特征内容的总称”。

尽管说法各异,但从以上的定义不难看出,信息与知识的概念是紧密地联系在一起的,知识性和可传播性是信息的基本属性。综上所述,可将信息的概念归纳为:信息是客观世界中各种事物的变化和特征的最新反映以及经过传递后的再现。信息是通过一定的物质载体形式反映出来的,是事物存在的状态、运动形式、运动规律及其相互联系、相互作用的表象。信息来源于世界上的一切事物,它是事物各种运动变化状态的客观显示;信息产生于事物在运动发展变化中的各种差异以及规律;信息是客观事物相互作用、相互联系的反映。一般说来,信号、消息、知识、情报、数据、资料、程序和指令等都可以统称为信息。

现在的信息概念不仅包括人与人之间的信息交换,而且还包括人与自动机之间、自动机

与自动机之间、动物世界与植物世界之间的信息交换,甚至细胞之间或机体之间的传递,也被看成是信息的传递。

在当今社会中,人们几乎无时无刻不与信息打交道,信息如同空气一样,跟我们的生活息息相关。人若离开了信息,就不会获得新的知识,社会离开了信息就会停滞不前。因此可以说,信息是社会发展的一项重要资源。

信息不是事件或物体本身,信息是事物、物质、事件某种属性的反映,是客观事物某些特性的表象和线索。信息和物质不同,它没有物质那样的实体,而是一种抽象的无所不在的东西,是一种能量形式。信息又往往以一定的载体形式传播,否则便会让原发事物自生自灭。信息是客观事物存在的方式或运动状态,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,经过提炼加工后,能成为有用的知识。数据、资料、语言、颜色、声音、光和光谱、自然界的信号等都可以传递信息。从类型上看,信息可分为来自自然界的信息与来自人类社会的信息两大类。依据其他标准,信息又可分为不同的类型,如有记录信息与无记录信息、动态信息与静态信息等。信息是已知的知识,又是未知知识的先兆,或者,也可把这种先导性的预兆现象称为信息。“成功之道,在于先知”,古人的这一名言早已说出了掌握信息的重要性和必要性。

2.1.2 知识

自有人类文明以来,知识作为人类认识自然和社会的阶段性成果,就一直受到了古今中外学者的重视。如英国著名学者培根有句流传非常广泛的名言:“知识就是力量”。

希腊时代的诡辩学认为,所有知识都是来自于个人的经验,而每个人对事物的感受不同,因此知识仅仅是相对于个人而言的,没有绝对意义的知识。

著名哲学家罗素在《人类的知识》一书中经分析得出结论:“知识是一个意义模糊的概念”。

诺贝尔物理学奖获得者彭亚斯认为,知识不是一堆杂乱无章的信息,而是有关世界怎样进行的信念。

彼得·德鲁克认为,知识是一种能够改变某些人或某些事物的信息,这既包括使信息成为行动基础的方式,也包括通过对信息的运用使某个个体有能力进行改变或者进行更有效行为的方式。

美国学者贝尔认为,知识是对事实或者思想的一套有系统的阐述,并提出合理的判断或经验性的结果。它通过某种交流手段,以某种系统的方式传播给其他人。

在以上观点中,“知识”不再是一个简单的、多元素的无序集合,而是被纳入一个动态的、与人或组织交互的系统。

1983年3月,国家科技领导小组在“关于知识与国家知识基础设施的研究报告”中,对“知识经济”中的“知识”定义为:经过人的思维整理过的信息、数据、形象、意向、价值标准以及社会的其他符号化产物,不仅包括科学技术知识,还包括人文社会科学的知识、商业活动、日常生活以及工作中的经验和知识、人们获取并运用和创造知识的信息,以及面临问题作出判断和提出解决方法的知识。这一定义基本上概括了国内专家和学者对知识概念的理解和认识。

另外,对知识概念通常有两种理解,即广义的知识和狭义的知识。广义的知识是指人们

通过学习、积累、发现、发明的各种知识的总和,包括普通知识和专业知识。狭义的知识是指知识经济研究的知识,通常是专业知识。

因此可以说,知识是建立在信息的基础之上,经过加工与编码后创造出来的新信息。信息常能导致深刻的见解,创造出新知识。知识是人们对客观事物存在和运动规律的认识。

知识和信息如果仅仅是拥有,未必能取得效益,只有将知识和信息应用到创新中去才能产生生命力。将信息转化为知识,再将知识转化为智慧,是一种动态的、开拓的、创新的过程,创新的结果是把经过信息到知识共享后得到的隐性知识上升为显性知识并付诸实践,成为智慧。反过来,智慧又常会转化为新的知识,新的知识又常会转化为新的信息,通过一定的手段和社会传递过程,借助于某些媒体传递给特定的使用者。

2.1.3 文献

信息有三种存在形式:一种存在于人的大脑中,为个人所掌握;一种存在于物质实体中,如样品、样机等实物;还有一种以文字、符号、图形等技术手段记录在特定载体上,如印制在纸上,摄制在胶片上,录制在磁带、磁盘、光盘上。人们通常把第三种信息存在形式称为文献。

“文献”一词源于中国,孔子是世界上最早提出“文献”一词的。孔子曰:“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”(《论语》)。近代以后,文献通常指有参考价值的图书资料。到了现代,文献的内涵更大了,文献就是用文字、图形、符号或声频、视频等技术手段记录知识信息的物质载体,或者说是“固化在物质载体上的知识信息”。《中华人民共和国国家标准·文献著录总则》(GB 3792.1—83)将“文献”定义为:“文献是记录有知识的一切载体。”换句话说,文献是用文字、符号、图形、声音、数据等方式,通过一定的技术手段(写、刻、印等)记录人类知识的一种信息载体。它记载的是人类世代认识自然、改造世界的知识和经验,汇集有无数的科学理论、方法、假说、数据和事实,它是人类智力劳动成功的表现形式。

国际化标准组织《文献信息术语国际标准》(ISO/DIS5217)对文献的定义是:“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”。我国现行的《文献著录总则》将文献定义为:“记录有知识的一切载体。”由此可见,文献是各种信息与知识的载体,不仅包括印刷品,也包括缩微资料、音像资料、机读资料等新型的信息载体。图书是文献,报纸、期刊是文献,录音带、录像带、光盘、电影片也是文献。

科技文献是记录有科学技术信息和知识的载体。它是人们从事科研、生产活动的历史记录,也是劳动成果的表现形式之一。它是千百万劳动人民和科研人员对客观事物认识的结晶,它积累了无数有用的事实、数据、理论、定义、定律、定理、技术、方法以及科学的构思和假想,记载着许多成功的经验和失败的教训。它能够反映当时人们对客观事物的认识程度和科学技术的发展水平,预示着科学技术发展的趋势和方向。

由此可见,文献的范围是非常广泛的。古代的甲骨文是文献,有文字和图形的碑刻是文献,竹简和帛书是文献,有铭文的青铜器是文献,现今的机读资料、缩微制品、电子出版物等也是文献。

纵观人类各个历史阶段的文献,虽然在内容、外在形式、记录方式等方面发生了深刻的