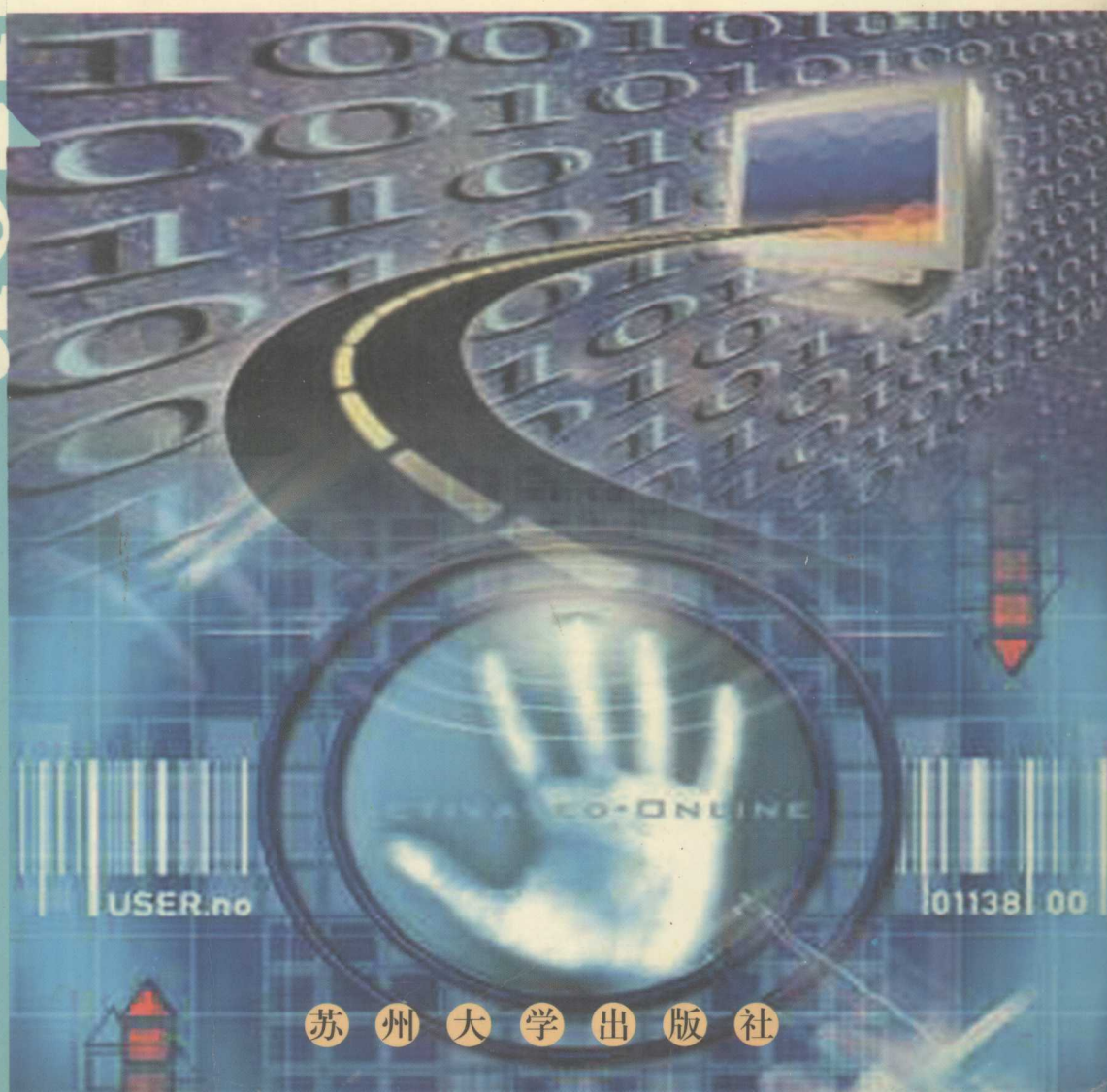


Computer

实用计算机信息检索

SHIYONGJISUANJI XINXIJIANSUO

★ 主 编 邹 忠 民
副 主 编 金 问 涛



苏 州 大 学 出 版 社

江南大学图书馆



90901894

实用计算机信息检索

主 编 邹忠民 副主编 金问涛
主要执笔 严大香 杨淑娟 袁 蓉



实用计算机信息检索

主 编 邹忠民

责任编辑 杨淑娟

苏州大学出版社出版

(地址: 苏州市十梓街1号 邮编: 215006)

苏州大学印刷厂印刷

(地址: 苏州市观前街 邮编: 215001)

开本 787×1092 1/16 印张 18.75 字数 468千

2000年12月第1版 2000年12月第1次印刷

ISBN 7-810-36000-1

苏州大学出版社

苏州大学图书馆

苏州大学图书馆

图书在版编目(CIP)数据

实用计算机信息检索/邹忠民主编. —苏州: 苏州
大学出版社, 2000. 12
ISBN 7-81037-761-2

I. 实… II. 邹… III. 计算机应用-情报检索-
高等学校-教材 IV. G354.4

中国版本图书馆 CIP 数据核字(2000)第 87184 号



实用计算机信息检索

邹忠民 主编

责任编辑 董张维

苏州大学出版社出版发行
(地址: 苏州市十梓街 1 号 邮编: 215006)
海安县塑料彩印二厂印装
(地址: 海安县双楼镇 邮编: 226671)

开本 787×1092 1/16 印张 18.75 字数 468 千
2000 年 12 月第 1 版 2000 年 12 月第 1 次印刷
印数 1—5500 册

ISBN 7-81037-761-2/TP·37(课) 定价: 26.00 元

苏州大学版图书若有印装错误, 本社负责调换
苏州大学出版社发行科 电话: 0512-5236943

前言

信息化浪潮正在席卷全球,信息用户教育已成为越来越重要的一项国民教育内容。苏州大学图书馆通过近两年来对本科生、研究生及青年教师的数据库检索培训和教学实践,深感编写一本新颖而实用的计算机信息检索指导用书的必要性。这是由于20世纪末计算机的应用已渗透到人类生活的方方面面,在文献检索工具上的反映则是增添了光盘版、联机版和Web网络版等多种新媒体;同时信息技术和信息高速公路的迅速发展,推动了虚拟化、数字化图书馆的建设,为广大读者开辟了一个网上资源利用的全新天地。

本书的编写力图体现两个特点:一是紧扣用计算机检索对工作和学习有价值的信息这个主题。为此,凡与机检有关的内容,包括基本知识、数据库描述及操作技巧等均有涉及,对机检系统的选介则涵盖光盘检索、联机检索和Internet网上查询等各个领域;二是强调实用性。以典型数据库的应用与操作技能为编写重点,文字表述力求通俗易懂,尽量引用实例说明问题,并配有众多照原稿复制的数据库操作页面插图,使读者能看得懂、学得会、用得上,具备初步计算机操作能力的读者则可以按图索骥、无师自通。

全书由三个板块构成:第1、第2章为与计算机信息检索有关的基本知识;第3至第10章对当前我国高校中广为使用的国内外大型数据库的内容、形态及其检索方法进行了具体描述;第11至第15章介绍了获取Internet信息服务的一些基本技能。各章编写的具体分工为:第1章由邹忠民编写,第2、第8章由杨淑娟编写,第3、第4、第6章由袁蓉编写,第5章由袁蓉、严大香编写,第7章由金问涛、袁蓉编写,第9章由杨淑娟、金问涛编写,第10章由杨淑娟、邹忠民编写,第11至第15章及其附录由严大香编写,全书由邹忠民负责统稿。

本书承教育部高教司文献检索课教学指导小组成员潘树广教授作序并审改部分章节,潘先生是国内知名文献学专家、高校文献检索课教学的开拓者之一,近年来又对计算机信息检索的理论与实践倍加关注并身体力行,有这样的大师指点,我们深感荣幸。本书的编写还得到徐记忠馆长的大力支持和具体指导,在此一并表示衷心感谢。鉴于时间仓促和学识有限,倘有不妥及错误之处,敬请读者斧正。

编者

2000年10月

序

潘树广

20多年前,我曾在《文献检索与语文研究》中说:“电子计算机的应用,使检索工具的编制和文献资料的查找实现了自动化。我国在这方面还比较落后,但可以预期,我们的子孙后代,将不会像我们这样,为了借一本书而在图书馆里等上个把小时,为了查核几条资料而到处奔走了。”(《辞书研究》1979年创刊号)真没想到,信息技术的发展速度比预料的快得多。不必等到“子孙后代”,就在我们这一代,已享受到光盘检索和 Internet 带给我们的方便,这是十分幸运的事。

如今,全球经济活动正由工业经济向知识经济逐步过渡。知识经济是以高科技产业为第一产业支柱,以智力资源为主要依托的经济。人们比以往任何时期都更加重视知识与信息的迅速获取和利用。教育部 1998 年颁布《普通高等学校本科专业目录和专业介绍》,对 87.55% 的专业明确规定了“掌握文献检索、资料查询的基本方法”或“运用现代信息技术获取相关信息的基本方法”的培养要求,各高校正积极加强文献检索或信息检索的课程建设和教材建设。我校邹忠民同志主编的《实用计算机信息检索》,为该课程的建设和发展作出了可贵的贡献。

本教材的一个鲜明特点,是定位准确,针对性强。作者经过多年教学实践,对教学对象的特点和信息需求有深入的了解。本课程是为文、理、工各专业本科生和研究生开设的公共选修课,而不是为信息管理专业开设的专业课。因此,教材应当简明实用,重在操作,而不必在理论上对信息科学作过多、过深的阐述。当然,任何操作技能的传授,都不能脱离一定的理论知识,但这种理论知识的阐述,应当紧紧围绕理论与实践的结合点进行。正是基于这样的定位,作者将全书设计为三大板块,即:计算机信息检索基础知识、中外重要数据库介绍和网络检索技术,重点鲜明,切于实用。

第二个特点,是实例丰富,图文并茂。作者介绍各类数据库和网上资源的利用方法,十分注意引用实例予以深入浅出的说明;所举之例,又多与学生的论文写作有关,令读者有亲切感。书中配以大量插图,形象直观,读者易于理解。

第三个特点,是推陈出新,信息及时。作为信息检索的教材,理应在信息意识方面作出示范,作者正是这样做了。细心的读者如果把这部教材和两三年前出版的同类教材作一比较,便可发现,一些陈旧的、实用性差的内容省略了,而新的数据库(或原有数据库的新发展)、新的网站则作了详细的介绍。例如大型数据库《中国学术期刊》,既介绍其光盘版(CAJ-CD)的内容在 1999 年发展为 9 个专辑的情况,又介绍其检索功能的改进,还详细介绍其最新的网络版,帮助读者不断跟踪最新信息。

当我读完这部教材稿时,想起一件往事。1996 年 1 月,我去哈尔滨参加全国高校文献检索课教学研讨会。南方冶金学院的杨老师对我说,他在《人民日报》上读到两位记者写的一篇文章。文章认为电子出版物的出现,使文献检索课失去了存在的价值。杨老师对两记

者的议论表示困惑,并在回校后将文章剪报寄给我。我深感,这篇文章反映了认识上的一些误区,颇有代表性,很有必要借此机会予以澄清。

两位记者说,他们上大学时曾经学过数十小时的文献检索课,很费劲。然而,“如今具备光盘检索条件的高等院校里,所谓‘文献检索’课的全部内容,也就是花一个小时的时间回顾文献检索的历史,然后用10分钟的时间讲解光盘数据库的用法。电子出版物的出现,将一门复杂的课程简化得让它最终失去了存在的价值”(《人民日报》1995年11月23日)。这些议论幼稚得可爱。第一,两位记者把光盘检索技术看得太“简化”了。读者看了面前这部沉甸甸的教材就可知晓,光盘数据库的用法岂能讲10分钟就使学生了解?它和其他学问一样,不下苦功是学不好的。第二,两位记者将光盘检索视为万能,以为它可以取代手工检索。其实,且不说仅有光盘检索,即使再加上网络检索,也不能完全取代手工检索。理由很简单,计算机检索的前提条件是将文献信息进行数字化处理,而目前尚有浩如烟海的文献信息(尤其是汉语文献)尚未制成光盘或尚未上网,还得靠书本式检索工具进行检查。因而,今天的大学生,应同时学好“手检”和“机检”,才能如鱼得水、左右逢源。第三,两位记者概念混乱,把“文献检索课”与“光盘检索”对立起来,认为有了光盘检索,文献检索课便“失去了存在的价值”。殊不知,光盘检索本来就是文献检索课的一个重要组成部分,它在90年代初陆续走进各高校的课堂,使文献检索课的内容更为丰富,更有吸引力。90年代后期,文献检索课又增加了Internet的内容,更增强了现代感和生命力。所不同者,部分院校将“文献检索课”改称“信息检索课”,是考虑到“信息”的涵义比“文献”更广。“信息检索”是“文献检索”的拓展,而不是取消。

本教材的作者,是我校图书馆几位中青年同志。密切为教学服务,是苏州大学图书馆的优良传统。记得在70年代末,当我初次开设文献检索课时,图书馆的同志就给予主动的配合。他们把上课要用的图书送进课堂,把学生的实习用书集中陈列在阅览室专架上,并承担辅导工作,使我深受感动。我当时写了篇《图书馆为教学服务的生动事例》,发表在江苏省图书馆学会的会刊上,说的就是我校图书馆热情支持文献课的事迹。从80年代中期开始,图书馆从协助开课转为独立开课。他们针对各院系不同的专业特点讲课,取得良好的教学效果。尤其是最近几年,图书馆文献检索室大量更新设备,并对本校师生进行计算机检索的培训,深受欢迎。这本教材,就是他们教学经验的总结和理论升华。古人云:“功崇惟志,业广惟勤。”我深信,文献检索室的同志们必将以不懈的努力取得更大的成绩。是为序。

2000年10月于苏州大学

目 录

第 1 章 引论

- 1.1 信息 1
- 1.2 信息科学 5
- 1.3 信息技术 6
- 1.4 信息检索 7
- 1.5 信息检索教育 8

第 2 章 计算机信息检索基本知识

- 2.1 检索的概念 10
- 2.2 检索系统的类型 11
- 2.3 检索系统的构成 13
- 2.4 检索语言 14
- 2.5 检索方法 18
- 2.6 检索技术 19
- 2.7 检索策略 22

第 3 章 《中国学术期刊》

- 3.1 《中国学术期刊》(光盘版) 27
- 3.2 《中国学术期刊》(网络版) 32

第 4 章 社科类数据库

- 4.1 《中文社科报刊篇名数据库》 37
- 4.2 《中国人民大学复印资料》 41
- 4.3 《中国法律法规大典》 45

第 5 章 科技类数据库

- 5.1 万方数据资源中心数据库 47
- 5.2 重庆维普资讯公司数据库 62
- 5.3 中国专利信息检索系统 71

第 6 章 《Chemical Abstracts》光盘数据库

- 6.1 概况 79
- 6.2 CA on CD 的检索 79
- 6.3 CA 的检索结果 85

第 7 章 OCLC 参考和资源共享服务系统

- 7.1 OCLC 简介 88
- 7.2 FirstSearch 服务数据库指南 89
- 7.3 New FirstSearch 的检索方法 97

第 8 章 Ei Village 工程信息村

- 8.1 Ei Village 的由来 108
- 8.2 Ei Village 的基本功能 108
- 8.3 进入 Ei Village 108
- 8.4 Ei Village 的网络导航 110
- 8.5 Ei Village 的独家服务 112
- 8.6 Ei Village 的社区服务 113
- 8.7 Ei CompendexWeb 114

第 9 章 Web of Science

- 9.1 概况 123
- 9.2 Web of Science 的检索方法 124

第 10 章 Dialog 国际联机检索系统

- 10.1 Dialog 简介 141
- 10.2 Dialog 系统数据库指南 142
- 10.3 Dialog 系统的检索方法 149
- 10.4 Dialog 系统的特色服务 152
- 10.5 DIALOG@SITE 154

第 11 章 Internet 概述

- 11.1 什么是 Internet 162
- 11.2 Internet 能干什么 168
- 11.3 如何获取 Internet 信息服务 168

第 12 章 WWW 信息查询

- 12.1 关于 WWW 171
- 12.2 浏览器的使用 171
- 12.3 实用网站介绍 185
- 12.4 搜索引擎的使用 190

第 13 章 网上免费资源择要

- 13.1 中国科学数据库 207
- 13.2 联合西文期刊篇名目次库 211
- 13.3 人民日报网络版全文检索系统 214
- 13.4 MEDLINE 生物医学数据库 218

第 14 章 电子邮件(E-mail)的使用

- 14.1 电子邮件(E-mail)是什么 227
- 14.2 电子邮件的收发和管理 229
- 14.3 如何申请免费电子信箱 240

第 15 章 使用 FTP 下载网上资源

- 15.1 什么是 FTP 244
- 15.2 FTP 下载网络资源的方法 244
- 15.3 压缩与解压缩 252

附录 1 部分国家和地区域名代码表 256

附录2 Internet 网址选 258

附录3 网上免费数据库选 274

主要参考文献 289

第 1 章

引 论

随着计算机的迅速普及,人类正步入“信息化”社会。在高度信息化的社会里,人们可在办公室和家庭中,使用计算机通过四通八达的计算机通信网络,去浏览、存储、处理和查找自己需要的数据和信息,去分享人类知识宝库中取之不尽、用之不竭的信息资源。目前,因特网的迅速发展,计算机信息检索网络的迅速完善,已使人们看到了信息化社会的美好前景。

计算机信息检索的产生和发展,不仅是科技情报战线具有划时代意义的事件,也是人类社会的重大事件之一,在某种意义上说,计算机信息检索与数据处理奠定了信息化社会的技术基础。

计算机信息检索,顾名思义,就是利用计算机对信息资源进行检索,从中获取自己所需的信息。这里,信息是对象,计算机是工具,检索是一种手段。要了解计算机信息检索,首先应了解信息是什么,以及围绕信息的信息科学、信息技术、信息检索等有关问题,并从中了解开展信息检索教育的目的和意义。

1.1 信息

信息一词是从英文 Information 翻译而得,它来源于拉丁文“*Informatio*”,这是作为一种广义的理解,如作为狭义的理解通常被翻译成“情报”。

1.1.1 信息的定义

信息是什么?这是一个貌似简单而又很难说明白的问题。围绕其定义所出现的流行说法不下有百种,例如哲学家认为信息是物质的普遍属性,是认识世界的依据;数学家认为信息“是使概率分布发生变动的东西”(Tribes, etc., 1971);物理学家说“信息是负熵”(Brillouin, 1956);而通信专家则把信息看作是“不定度”的描述。另外,“信息是选择的自由度”(Hartley, 1928);“信息是人与外界相互作用的过程中所交换的内容的名称”(Wiener, 1950);“信息是用以消除随机不定性的东西”(Shannon, 1948);“信息是事物之间的差异”(Longo, 1975)以及信息就是消息、信号、数据、情报、知识等,不同的学者从不同的角度对信息作出各种定义,可谓众说纷纭,难以有统一的定论。显然信息与整个人类的生存和发展有着极其密切的关系。不过,最早把信息作为一种科学研究的对象来加以探讨的却是通信科学技术工作者,当时他们只把信息看作是消息的同义语。这种认识有很长的历史渊源,例如,早在一千多年前,我国唐代就曾有“梦断美人沉消息,目穿长路倚楼台”的诗句,其中“消息”一词就是音信、消息的意思。同样,在西方出版的许多文献著作中“消息”(Message)和信息(Information)两词也是互相通用的。出现电信技术后又引申出“信息就是信号”的说法;出现计算机技术后又进一步派生出“信息就是数据”的说法。此外还有把信息看成是情报的。这些说法都只抓住

了一些表面上的共同点,即消息、信号、数据、情报和信息一样,都是在通信系统中传送的东西。实际上,认真分析一下就知道,这些概念之间是有原则区别的。不难理解,消息是信息的外壳,信息则是消息的内核。所以,同样是一分钟的消息,有的消息所包含的信息量可能很大,有的则可能很小。类似地,信号也不等同于信息,信号只是信息的载体,信息是信号所载荷的内容。同样一个信息既可以用这种信号来载荷,也可以用那种信号作载体。比如一个简单的“是否”信息既可以用“0-1”数字来代替,也可以用“正-负”电流来载荷,还可用“有-无”光通量来表示等。至于数据,它只是记录信息的一种形式,而且不是唯一的形式,因此不能把它等同于信息本身。情报一词在日语中的确就是信息,但是在汉语中,情报只是一类专门的信息;是信息的一个子集,是狭义的“信息”,以信息的形式而流动。至于知识则是由信息加工出来的产物,是一种认识主体所表述的事物运动状态和方式的认识论层次的信息,是人们对客观对象的规律性的认识(客观对象是多种多样的,有具体对象,有抽象对象。例如,在程序设计语言中,常量是对象,变量也是对象。程序中出现的常量与变量是客观问题中常量与变量的抽象,彼此之间相互对应。规律性反映客观对象的运动规律,而这些规律在程序中反映为“抽象对象的处理规则”)。它具有普遍和概括性质,一方面是人们认识世界的结果,另一方面又是人们改造世界的方法。信息具有知识的本性,但它又比知识的内涵更广泛、更松弛。因此,从科学概念和客观对象的规律性来考虑,信息、情报和知识之间的逻辑关系应为包含与被包含的关系。

“信息是人与外界相互作用的过程中所交换的内容的名称”是美国数学家、控制论的主要奠基人维纳(N. Wiener)在1950年出版的《控制论与社会》一书中提出的。他写道:“人通过感觉器官感知周围世界”,“我们支配环境的命令就是给环境的一种信息”,因此,“信息就是我们在适应外部世界、并把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名称”。“接收信息和使用信息的过程,就是我们适应外界环境的偶然性的过程,也是我们在这个环境中有效地生活的过程”。在这里,Wiener把人与外部环境交换信息的过程看作是一种广义的通信的过程。这当然是没有问题的。因为广义的通信本来就可以泛指人与人、机器与机器、机器与自然物、人与自然物之间的信息传递与交换。但人与环境之间互相交换的内容中不仅有信息,也有物质和能量,把它们统称为信息,岂不是把信息与物质及能量混为一谈。因此,这种理解是不够确切的。

“信息是选择的自由度”,这是哈尔特雷(R. V. L. Hartley)1928年在《贝尔系统电话杂志》上发表的“信息传输”论文中提出的。他把信息理解为选择通信符号的方式,并用选择的自由度来计量这种信息的大小。他认为,发信者所发出的信息就是他在通信符号表中选择符号的具体方式。例如,假定他从符号表中选择了“I am well.”这样一些符号,他就发出了“我平安”的信息;如果他选择了“I am sick.”这些符号,他就发出“我病了”的信息。发信者选择的自由度越大,他所能发出的信息量也就越大。比如,假定发信者只能从二元符号表(如只含有0和1两个符号的符号表)中选择符号,而且规定他发出的每个“字”只能由一个符号组成。显然,在这个限制下,他的选择自由度很小,他所发出的不同的“字”只有0和1两个。如果放松限制,规定每个“字”可以由两个符号组成,那么,他可发出00、01、10、11四个不同的“字”。因此,它们所能载荷的信息量就比原来增加了。如果进一步放松限制,使符号表的符号数也可增加,由原来的0、1增加为0、1、2,那么他的选择自由度就更大,他可发出00、01、02、10、11、12、20、21、22九个不同的“字”。此外,Hartley还注意到,选择的具体物理内容是无

关紧要的,重要的是选择的方式。不管符号代表的意义是什么,只要符号表的符号数目一定,“字”的长度一定,那么,发信者所能发出信息的数量就被限定了。Hartley 的这种理解虽能在一定程度上解释通信工程中的一些问题,但存在着一些严重的局限性:(1) 他所定义的信息不涉及内容和价值,只考虑选择的方式,因而只属于语法信息的范畴,丢弃了语义信息和语用信息;(2) 即使在语法信息范畴内,也没有考虑到信源的统计性质;(3) 把信息理解为选择的方式就必须有一个选择的主体作为限制条件。这些缺点大大限制了它的适用范围。

“信息是用以消除随机不定性的东西”,这是另一位美国数学家申农(C. E. Shannon)1948年在《贝尔系统电话杂志》上发表的一篇题为“通信的数学理论”中提出的。这篇论文以概率论和统计数学为工具,深刻阐述了通信工程的一系列基本理论问题,给出了计算信源信息量和信道容量的方法和一般公式,得到了一组表征信息传递的编码定理。虽然文中并没有直接叙述信息的定义,但 Shannon 在进行信息的定量计算的时候明确地把信息量定义为随机不定性之差,这就表明了他对信息的理解:信息是用来减少随机不定性的东西。这种认识比仅仅把信息看作消息或通信内容要深刻得多,而且在数学上很容易证明,但还存在着一些缺陷:(1) Shannon 理论中的不定性纯粹是波形形式上的不定性,与此相应的信息概念也是纯粹的形式化的概念(称为语法信息),因此,它的适用范围受到限制;(2) 它只考虑了随机型的不定性,不能解释与其他型式的不定性(如模糊不定性)有关的信息问题;(3) 这种信息定义只从功能角度上来表述,还是没有从根本上回答信息是什么的问题。

“信息是事物之间的差异”,这是意大利学者朗固(G. Longo)在1975年出版的《信息论:新的趋势与未决问题》一书序言中的说法。他认为:信息是反映事物的形式、关系和差别的东西,它包含在事物的差异之中,而不在事物本身。我国已故电子学家冯秉铨先生也赞同“信息就是差异”的理解,他还用一个十分形象的例子来说明这个观点。他说,如果某人发出一个恒定持续的声音“啊——”,人们是不能从中得到什么信息的,因为它没有差异;但如发出的声音出现抑扬起伏,就会表达出某种信息,比如,“啊↘”就表示出感叹的信息;而“啊↗”则表达出疑惑的信息;“啊~”又表示出恍然大悟的信息。总之,差异论的实质就是“有差异就有信息,没有差异就没有信息”。我们赞成“有差异就有信息”的观点,但反过来说“没有差异就没有信息”就不够确切。

近年来对信息的论述又趋向具体化。美国图书馆学会的信息专家认为:信息是一切思想、事实和富有想像力的作品,它们以各种方式进行了记录和传播。我国大多数学者从认识论出发认为:信息具有物质属性,是物质的一种存在形式,它以物质的属性或运动状态为内容,并且总是借助于一定的物质载体传输或存储。

总起来说,在现有的各种理解中,Shannon 的定义比较深刻,而且这种定义还导出了相应的算法。但是正如前面所述,Shannon 定义也还存在不少缺陷,我们还必须进一步去寻求更合理、更科学和更有用的信息定义。

那么,信息的定义究竟是什么?我们认为,信息是一种独立的研究对象,是物质运动的一种特征形式,反映了各种运动的相互联系。从认识论的意义上说,它是认识主体所感受或所表述的事物运动的状态和方式;从内容和形式两方面来认识,自然界与人类社会存在的事实与变化及人类对它们形成的认识和创造是信息的内容组成,载体记录和媒体传播是信息的存在形式,它们构成了信息的基本属性。这里所说的“事物”,既可以是外部世界的实在客体,也可以是主观世界的精神现象;而“运动”既可以是物体在空间中的位移,也可以是一切

意义上的变化;“运动的状态”是指事物在特定时空中的性状和态势,“运动的方式”是指事物运动状态随时空的变化而改变的式样和规律。

1.1.2 信息的特征和性质

信息无处不在,无时不有,存在于人类社会和自然界中,虽然难以统一其定义,但从上述的讨论中可以导出信息的一些重要的特征和性质。

1. 信息的基本特征

(1) 信息不同于物质。物质是信息的载体,物质的运动是信息的源泉,但信息只是事物运动的状态和方式。它来源于物质,又不是物质本身,它从物质的运动中产生出来,又可以脱离源物质而相对独立地存在。

(2) 信息不限于精神。由于客观世界的物质客体和精神世界的事物都可以产生信息,因此,信息也来源于精神世界,但是又限于精神的领域。

(3) 信息不同于能量。信息是事物运动的状态和方式,它提供的是知识和智慧,而能量是事物做工的本领,它提供的是动力。传递信息需要能量,驾驭能量则需要信息,它们都与事物的运动相关联。因此,信息与能量息息相关,但是又与能量有本质的区别。

(4) 信息具有知识本性。信息可以改变人们的知识状态,使人们由“不知”变为“知”,或由“知之较少”变为“知之较多”。人们也可以把信息加工成为知识。因此,信息具有知识的本性,但它本身不一定就能够被称为知识,它比知识的内涵更广泛、更松弛。

(5) 信息是具体存在的。信息不是虚无缥缈的东西,也不是可以随意想象和“创造”的事物。它是现实世界各种事物运动的状态和状态改变的方式,具有非常具体和真实的品格。它可以被人(生物、机器等)所感知、提取、识别,可以被传播、存储、变换、处理、显示、检索和利用。

(6) 信息可以共享。由于信息可以脱离源事物相对独立地存在并负载于其他载体,因此它可以被无限制地进行复制、传播或分配给众多的用户,为大家所共享。

2. 信息的主要性质

(1) 普遍性。信息是普遍存在的。只要有事物存在,只要有事物的运动,就会有它们运动的状态和方式,就存在着信息。信息存在于自然界,存在于人类社会,也存在于思维领域。

(2) 无限性。一切事物运动的状态和方式都是信息,而宇宙时空中的事物是无限丰富的,因而它们所产生的信息也必然是无限的。即使是在有限的空间中,比如在地球上,事物也是无限多样的,而在无限的时间长河中,事物的发展变化更是无限的,因而信息自然也是无限的。

(3) 相对性。对于同一个事物,不同的观察者由于其观察能力、理解能力和目的性的不同,他所获得的信息量就有可能各不相同。

(4) 转移性。由于信息具有脱离母体而相对独立的能力,因而它可以通过一定的方法使之在时间上或在空间中进行转移。在时间上的转移称为存储;在空间中的转移称为通信。

(5) 变换性。信息是可替代变换的,它可以由不同的载体和不同的方法来载荷。

(6) 时效性。信息是事物运动的状态和状态改变的方式,而事物本身是在不断发展变化的,因此,信息也会随之变化。一切活的信息,包括信息发出、接收到利用的时间间隔和效率以及信息本身更新的速度,都随时间而变化,因此,信息也是有时效、有“寿命”的。

(7) 传递性。信息借助于一定的物质载体可以进行传送,其传递过程是:信源→编码→信息→信息传递渠道(载体)→译码→信宿。编码和译码是把要传递的信息用语言、文字、图形、代码、符号等形成可传递和可接受的信息,是信息被准确接收的关键。

(8) 转化性。信息在一定的条件下可以转化为物质、能量、时间及其他。正确而有效地利用信息,就可能在同样条件下创造更多更好的物质财富,开发或节约更多的能量,节省更多的时间。

(9) 共享性。在信息的扩散和用户分享信息的过程中,信息载体本身的信息量并不因此过程而减少,各用户分享的信息份额不因分享的人的多少而受影响。

(10) 可扩散性。信息的传递性决定了信息的可扩散性。可扩散性是指信息能通过各种渠道、媒介进行传播。随着信息网络覆盖面的不断扩大,信息的扩散是没有穷尽的。

(11) 可扩充性。人们对世界的认识是无限的,因此信息资源的扩充与累积也是无限的。

1.2 信息科学

1.2.1 信息科学的定义

信息科学是一门新兴的科学,在文献中还没有给出过正式的定义,也没有建立起明确的边界。我们认为,像定义其他科学一样,我们可以把信息科学定义为“研究信息及其运动规律的科学”。而更为精确的定义则是:所谓信息科学,就是以信息作为主要研究对象、以信息的运动规律作为主要研究内容、以现代科学方法论作为主要研究方法、以扩展人的信息功能(特别是其中的智力功能)作为主要研究目标的一门科学。这里,以信息作为主要研究对象就是信息科学区别于其他传统科学的最根本的特点之一,也是信息科学之所以能够成为一门独立学科的最根本的前提。信息科学的研究内容则由其自身所特有的新的独立的研究对象和研究领域所决定的一种全新的研究内容——研究信息的运动规律,即信息的产生、提取、再生和施效的原理和规律。信息科学的研究方法是一种与新对象新内容相适应的全新的科学方法。它主要包括一个方法、两个准则,即信息方法和功能准则、整体准则。信息方法是整个方法论体系的灵魂,它有信息分析和信息综合两层含义,前者解决高级复杂事物的工作机制的认识问题,后者解决高级复杂事物的工作机制的实现问题,从认识到实践(实现),信息贯穿始终。两个准则则是保证信息方法能够正确实施的法则。功能准则是指在利用信息方法来分析和实现高级复杂的信息系统时,主要应当着眼于系统的功能,而不必关心它们的具体结构;整体准则则是指在利用信息方法来分析和实现高级复杂的信息系统时,主要应当着眼于整体的功能的优化,而不必关心个别局部功能的最优。信息科学的研究目标是扩展人类的智力功能。

1.2.2 信息科学的意义

十分明显,信息科学具有极为重要而深远的科学意义和社会意义。

1. 提出了新的研究对象,开辟了新的重要的研究领域

如上所述,信息是一种区别于物质和能量的独立的研究对象,而且是一种极为重要的研

研究对象。从一定的意义上来说,物质、能量、信息都是人类生存和社会发展所不可缺少的资源,其中,物质和能量是更为基本的资源,信息则是一种较高级的资源,信息资源提供给人类的是知识和智慧。当人们能够透彻阐明信息的本质和利用信息资源的基本规律时,它将会引起时代观念的深刻变革,并由此导致对社会资源的全面科学的利用。

2. 孕育了现代科学方法论,为现代科学的研究开拓出崭新的局面

和材料分析及能量分析的方法不同,信息方法关心的主要是对象所涉及的信息过程。尤其是那些高级的运动形式,它不仅重视“事物运动的状态和方法”的形式方面(即语法信息),而且还特别重视“事物运动的状态和方法”的含义(即语义信息)以及“事物运动的状态和方法”的价值效用(即语用信息)。正是对事物的语义和语用信息的充分利用,才有可能揭示诸如生命、思维、智慧这样一些高级的活动和复杂的过程的奥妙。

3. 提供了强有力的为自然科学和社会科学研究的科学方法

在传统科学深入研究物质与能量的运动规律的基础上,信息科学着重探讨信息的运动规律,把信息、能量和物质看作是一个有机的统一体。同时,信息科学又把信息过程、控制过程与系统的优化过程统一起来,得到了“利用信息、提高控制、优化系统”这样一种高级运动的重要模式。运用这种模式,信息科学就不仅能够揭示出自然科学领域内许多高级复杂事物(如高等生物、智能机器等)的运动本质和规律,而且也可以解释大量社会现象的活动本质和规律,从而在自然科学与社会科学之间架起了联系的桥梁。可以肯定,利用信息科学的原理和方法,必将有力地推动自然科学与社会科学研究的发展和进步。

4. 推进了信息科学技术的进步和广泛应用,大大促进社会的信息化过程

由于信息科学技术的进步,各种先进的智能化工具将不断问世,并越来越普遍,越来越有效地被社会所利用,大大地促进了社会的信息化过程,其结果便是全面地增强人类认识世界和改造世界的能力,把社会进步推进到一个空前的水平。

总之,信息科学与以往一切传统科学不同,它破天荒地把信息推上了科学的舞台,使它与物质和能量鼎足而立,并以崭新的思想和方法大大丰富了科学的宝库。它的兴起极大地改变了整个科学的结构、内容和方向,改变了科学发展的图景和科学的思维方式,从而将彻底地而且不可逆转地改变科学的面貌,因而也将改变整个社会的面貌。

1.3 信息技术

技术并不是从古就有的,也不是永恒不变的。科学技术是在人类与自然的斗争中为了增强自己的力量,赢得更多更好的生存发展机会而发生发展起来的。因此,科学技术的天职就是辅助人类。具体地说,科学技术是通过加强或延长人的某种或某些器官的功能来辅助人的。例如,飞机、舰船、车辆等是用来加强或延长人的行走器官的功能,起重机、水压机等各种动力工具是用来加强或延长人的体力即力量功能。这样,从技术的本质意义上我们就可以给出信息技术的明确定义:信息技术就是能够扩展人的信息器官功能的一类技术。人们通过感觉器官(眼、耳、口、鼻和皮肤等)来获取信息;通过传导神经(导入神经网络和导出神经网络等)来传递信息;通过思维器官(记忆系统、联想系统、分析推理和决策系统等)来加工和再生信息;通过效应器官(操作器官手、行走器官脚和语言器官口等)来施用信息。人类的这四类信息器官和它们的信息功能是有机地联系在一起的,这种有机的联系使它们能够

执行一种整体性的高级功能——认识世界和改造世界,使世界协调发展所需要的智力功能。

根据上述给出的信息技术的定义和相应的分析我们可以明确信息技术的基本内容(内部结构)有四项,通常称为“信息技术四基元”。

- 感测技术 包括传感技术和测量技术,如遥感、遥测技术等,它是感觉器官功能的延长,使人们能更好地从外部世界获得各种有用的信息。

- 通信技术 它是传导神经功能的延长,其作用是传递信息。

- 计算机技术 包括硬件和软件技术以及人工智能技术等,它是思维器官功能的延长,使人们更好地加工和再生信息。

- 控制技术 它是效应器官功能的延长,其作用是根据输入的指令(决策信息)对外部事物的运动状态实施干预,即信息施效。

在信息技术的四项基本内容中,通信技术和计算机技术是核心,感测技术和控制技术则是核心与外部世界之间的接口。没有通信和计算机技术,信息技术就失去了基本意义。而没有感测技术和控制技术,信息技术就失去了基本的作用:一方面没有信息的来源,另一方面也失去了信息的归宿。可见,信息技术的四项基本内容是一个完整的体系,它们和谐有机地合作,共同完成扩展人的智力功能的任务。

由此,我们还可以引申出信息技术的两个比较具体的定义:

其一,信息技术是指能够完成信息的获取、传递、加工、再生和施用等功能的一类技术。

其二,信息技术是指感测、通信、计算机(包括人工智能)和控制等技术整体。

由此可见,把信息技术单纯理解为“一个C”(computer)、“两个C”(computer and communication)和“三个C”(computer, communication and control)是不够全面的。

1.4 信息检索

信息检索(Information Retrieval)一词是1950年C.N. Moore在一次国际数学家会议上提出来的,迄今为止,该词并没有公认一致的明确定义,虽众说不一,但大部分说法相差不多。例如,原苏联切尔内认为,信息检索是从大量文献中查找含有用户提问中指定特征的文献或含有用户所需事实的文献的过程;英国维克利解释说,信息检索是指从汇集的文献信息中找出特定用户在特定时间所需信息的操作过程。我们认为,信息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。它包括信息存储和检索两个方面。所以,信息检索的全称又叫信息存储与检索(Information Storage and Retrieval)。这是广义的“信息检索”。狭义的信息检索则仅指该过程的后半部分,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寻(Information Search)。这对信息用户而言,更为重要。这里说到的存储就是在检索之前必须把大量杂乱无序的文献信息收集起来加以科学地组织,使之有序化并按一定规律存储起来形成书本式检索工具或计算机可读数据库;而检索则是运用已经组织好了的各种检索工具,通过手工方法或使用计算机,按照存储信息所依据的规律,将用户所需的信息取出来。存储是为了检索,要检索就必须存储,而人则是存储与检索的主人。因此,人、检索(工具)和信息资源便成为信息检索系统的三个基本要素。

信息检索有多种类型,可以从不同的角度去分类。按信息检索内容来分,由于作为检索对象的信息的形式不同,通常分为目录检索和文献检索(信息以文献形式出现)、事实检索和

数据检索(信息以事实和数据形式出现);按被检索信息的年代来分,有定题检索(选择性信息提供)和追溯检索(一次检索可获数年甚至十几年内与某一课题有关的信息);按检索手段来分,有手工检索和计算机检索。这些检索的涵义将在下面的检索基础篇中有较为详细的介绍,这里不作展开。

1.5 信息检索教育

在高等院校开展信息检索教育意义十分重大。在大学教育中,对大学生来说,五种能力(即自学能力、研究能力、思维能力、表达能力和组织管理能力)的培养是最主要的。信息检索教育则是培养学生能力的基本技能和方法之一。对于硕士生来说,在学习阶段中,主要是学会如何独立地从事研究工作。通过信息检索教学,学会多快好省地查找文献信息,并在此基础上,善于总结前人经验,在前人成就的基础上有所创新。对于博士生来说,主要是通过信息检索,培养在总结前人经验的基础上选择具有创造性的研究方向的能力,能够开辟新的研究领域。对于科研人员来说,在开始研究课题之前,首先要利用信息检索,并进行大量的调研工作,以便了解和继承前人的成果,摸清该学科或课题的现有水平和发展趋势,以及与其相关学科之间的关系。

当今,我们正处在一个知识、信息量激增的时代。随着科技人员数量的增加和强有力的信息系统的出现,一方面,人类的智力在更广阔的时间和空间拼接着、繁衍着,使得科技成果以前所未有的速度增长着;另一方面,知识的更新在加快,信息情报的有效期在缩短。因此,当今知识、信息的发展已使科技人员检索适用的文献信息变得越来越困难,人们所面临的形势对传统教育提出了挑战。人的大脑已无法容纳如此巨大的知识,人们的知识结构需要不断地更新、充实。国际教育发展委员会主席埃德加·富尔说过:“我们再也不能刻苦地、一劳永逸地获取知识了,而需要终身学习如何去建立一个不断演进的知识体系——‘学会生存’。”教会学生去建立一个不断演进的知识体系,培养学生的适应能力和独立科研能力,正是现在高等教育中存在的薄弱环节,亟需加强。在对学生进行传授基本知识的同时,要着重培养他们独立吸收和运用信息的能力。

要掌握具有信息吸收能力的知识必须克服三个方面的语言障碍,即自然语言障碍、学科专业语言障碍和图书信息系统中的检索语言障碍。自然语言障碍,主要是指外语水平低,迫使科技人员无法阅读用他们不懂的语言发表的文献信息,从而可能失去真正有价值的信息。学科专业语言障碍,主要是指由于学科越分越细,每门学科应用的术语、符号也随之专深,使科技人员相互之间有“隔行如隔山”之感,使人们难以广泛地涉猎相邻领域的新成就、新信息。检索语言障碍,主要表现为不懂信息检索语言,不掌握信息检索的方法和技能,以致不能有效地利用图书信息系统去检索有关文献资料,以调整知识结构,解决实际问题。如果学生仅仅具备一定的外语和学科专业知识,而没有一定的信息检索与利用的能力,就会面对汪洋大海般的文献资料陷入找不到、读不完的困境。因此,在高校中开设“信息检索”课,有目的地培养学生的信息意识和信息吸收能力,使其具备自我知识更新和获取最新信息的能力,其重要性就在于此。掌握信息检索的方法与技能的重要意义,至少可以从以下几个方面说明。

- 轻松获知 掌握了信息检索的方法和技能,就可以使人们的大脑从繁重的记忆活动