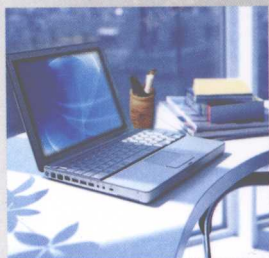


高职高专教材

计算机信息检索实用教程

张晓芳 主编



西安电子科技大学出版社
<http://www.xduph.com>

Computer

高职高专教材

计算机信息检索实用教程

张晓芳 主编

高明柯 王 倩 副主编

西安电子科技大学出版社

2007

图书在版编目(CIP)数据

计算机信息检索实用教程/张晓芳主编.

—西安:西安电子科技大学出版社,2007.8

ISBN 978-7-5606-1896-8

I. 计… II. 张… III. 机器检索:情报检索—教材 IV. G354.4

中国版本图书馆 CIP 数据核字(2007)第 117823 号

策 划 高维岳

责任编辑 曹 昉 高维岳

出版发行 西安电子科技大学出版社(西安市太白南路2号)

电 话 (029)88242885 88201467 **邮 编** 710071

http://www.xduph.com **E-mail:** xdupfxb@pub.xaonline.com

经 销 新华书店

印刷单位 陕西天意印务有限责任公司

版 次 2007年8月第1版 2007年8月第1次印刷

开 本 850毫米×1168毫米 1/32 **印 张** 7.25

字 数 175千字

印 数 1~4000册

定 价 15.00元

ISBN 978-7-5606-1896-8/TP·0982

XDUP 2188001-1

*** 如有印装问题可调换 ***

本社图书封面为激光防伪覆膜,谨防盗版。



内 容 简 介

本书结合高职教育的应用性、实践性和技能性的特点,从系统的角度出发,详细介绍了信息检索的原理与技术,主要讨论了如何迅速地检索到相关信息。主要内容包括:信息的基础知识、信息检索的原理与方法、搜索引擎与网上工具书的使用、电子期刊和报纸、数字图书馆和国外检索工具等。

本书可作为高职高专学生的教材或课外参考书,也可作为一般信息服务人员、咨询人员、管理人员的自学参考书。

★ 本书配有电子教案,需要者可与出版社联系,免费提供。

前 言

自 1951 年人们首次利用计算机进行文摘检索试验以来, 计算机信息检索系统历经脱机检索、联机检索阶段发展到了目前的网络化检索阶段。在此间 50 多年的发展历史中, 计算机信息检索在世界信息服务业中逐渐占据了主导地位, 并代表了当时世界信息服务的最高技术水平。进入 20 世纪 90 年代以后, 传统的计算机信息检索系统的生存受到了因特网(Internet)的竞争和威胁, 这是因为 Internet 在网络技术、网络化信息资源、网络服务应用等方面代表着现代信息服务的最新技术水平。进入到 21 世纪, 借鉴 Internet 的成功经验, 计算机信息检索系统逐步转向采用先进信息技术的新型网络化信息检索系统。

如果罗列当前网络上最流行的词汇, 相信“搜索引擎”一定位列其中, Google、百度这样的明星企业用它们的技术改变着我们的生活, 我们从来没有像今天这样能够如此方便、快捷地获取我们希望得到的信息。各学科的专业人员都想在培养信息意识的基础上, 掌握一些必备的信息检索知识, 熟悉信息获取的途径和技能, 完善自己的知识结构。

对于当代的高职高专学生来说, 掌握一门获取知识的技术比牢记许多知识显得更为重要。我们正处在一个网络时代, 知识在膨胀, 而人的记忆却是有限的, 所以, 必须要把注意力放在获取知识的技术上, 当我们需要知识的时候, 我们才有办法在我们的记忆之外来获取它。

“信息素质”教育是国外大学教育发展的一个重要趋势, 这种教育并不是单一的信息获取能力的培养, 而是认识和培养从各种

信息源中提取、评价和使用信息的能力,其具体内涵包括了解信息需求及问题所在、制定信息检索策略、掌握信息检索技术、评价信息并根据实际用途组织信息和使用信息,将新信息融会到现有知识结构中。虽然高职教育侧重于要求学生从各种信息源中学会获取,但是学生还是有必要了解信息的分析、评价和使用。

为此,本书在详细介绍信息检索的方法之外,还从信息的分析、评价和使用等方面作了较为全面的介绍,以便学生对未来向网络化、多媒体化、智能化方向发展的计算机信息检索系统有更好的适应能力。

本书在编写过程中参考了大量的文献资料,在此向参考文献的作者表示由衷的感谢。

由于作者能力有限,书中难免存在不妥之处,请读者给予批评指正。

编 者

2007年5月

目 录

第 1 章 概述	1
1.1 信息的基本概念	1
1.1.1 信息的定义	1
1.1.2 信息的类型	3
1.1.3 信息的主要特性与基本作用	6
1.1.4 信息、知识、文献的关系	10
1.2 知识创新与信息素质	11
1.3 文献知识	12
1.3.1 文献的载体类型	13
1.3.2 文献的出版类型	13
1.3.3 文献的级次	26
1.3.4 现代文献信息的特点	27
第 2 章 信息检索基本原理	31
2.1 信息检索的含义和类型	31
2.1.1 信息检索的概念	31
2.1.2 信息检索的基本类型	32
2.2 信息检索系统	34
2.3 信息检索工具	34
2.3.1 信息检索工具的概念	34
2.3.2 信息检索工具的特点	35
2.3.3 信息检索工具的基本类型	36
2.3.4 信息检索工具的一般结构	39
2.4 信息检索语言	41
2.4.1 语言分类	41
2.4.2 文献分类	43

2.4.3 分类法结构	44
2.4.4 常用分类法介绍	44
2.5 信息检索步骤	47
第3章 计算机及 Internet 信息检索	52
3.1 计算机信息检索基础知识	52
3.1.1 计算机信息检索发展概况	52
3.1.2 计算机信息检索系统	56
3.2 Internet 信息查询	58
3.2.1 中国 Internet 应用现状	58
3.2.2 Internet 信息常用获取方式	59
3.3 Internet 搜索引擎	60
3.3.1 搜索引擎概念	60
3.3.2 搜索引擎的分类	60
3.3.3 搜索引擎的使用技巧	63
3.3.4 常用搜索引擎选介	65
3.4 网上工具书	81
3.4.1 书目、索引	81
3.4.2 字(词)典	86
3.4.3 百科全书	88
3.4.4 年鉴	93
3.4.5 名录、表谱和手册	94
3.5 网上学位论文资源	101
第4章 电子期刊和报纸	115
4.1 电子期刊和报纸概述	115
4.2 电子期刊和报纸的特点	115
4.3 常用报刊信息检索与使用	116
4.3.1 中国期刊网(CNKI)	116
4.3.2 万方数据资源系统	121
4.3.3 维普中文科技期刊数据库	126
4.3.4 中文社科报刊篇名数据库	129

4.3.5 人大复印报刊资料数据库	132
4.3.6 人民数据库	135
第5章 数字图书馆	141
5.1 概况	141
5.1.1 数字图书馆的定义	141
5.1.2 国内数字图书馆的发展	143
5.2 数字图书馆的检索与使用	152
5.2.1 超星数字图书馆	152
5.2.2 北大方正 Apabi 数字图书馆	154
5.2.3 书生之家	160
第6章 国外常用检索工具介绍	162
6.1 SCI	162
6.2 EI	165
6.3 SSCI	167
6.4 A&HCI	167
6.5 ISSHP	168
第7章 特种信息检索工具介绍	170
7.1 专利及专利相关信息名词	170
7.2 中国专利信息检索	176
7.3 世界专利索引	194
7.4 专利信息网	197
第8章 综合实训	204
8.1 中国期刊网(CNKI)的检索使用	204
8.2 万方数据库的检索使用	210
8.3 超星数字图书馆的检索使用	214
附录	218
参考文献	222

第 1 章 概 述

1.1 信息的基本概念

1.1.1 信息的定义

信息作为名词术语,在中国最早出现在唐代诗人李中的诗句中:“梦断美人沉信息,目穿长路倚楼台”。但是,信息作为一个科学概念以及科学研究对象,却只有五六十年的历史。在人类社会的演变和发展过程中,信息一直在积极地发挥着人类已经意识或还没有意识到的重要作用。

自古以来,人们随时都在自觉不自觉地接收、传递、存储和利用信息。人类的信息活动也从来没有间断过。

20 世纪科学技术的空前进步,使人们对信息的认识和研究逐步建立在科学的基础上,形成了以信息为研究对象的理论——信息论,并迅速发展为一门具有丰富内涵的新学科——信息科学。

1948 年,美国科学家香农(C. Shannon)和维纳(N. Wiener)先后分别发表了《通信的数学理论》和《控制论:或动物和机器中控制和通信的科学》两篇著作,为信息学的建立奠定了理论基础。

然而,对于信息的定义,人们还没有一致的认识。香农这样描述信息:“信息是用以消除随机不确定性的东西”,从信息具有

减少人们认识的不确定性的功能上概括了信息的特征。维纳认为：“信息就是信息，不是物质，也不是能量”，把信息看成是与物质、能量具有同等重要性的客观世界三大要素之一。我国著名的科学家钱学森称信息是“激活了、活化了的的知识”，认为信息“就是为了解决一个特定的问题所需要的知识”，对信息进行了科学的概括。

可见，信息具有极其丰富的涵义，概括起来，信息的概念主要包含有以下三点。

(1) 为了一定目的而收集起来的正确的事实所构成的知识，也就是说，经过传递而又为人们所吸收的知识构成信息。

(2) 通过传递而进入人类社会交流体系的运动着的知识构成信息，即一方面信息可以通过一定的物质载体或其他交流形式进行传递，另一方面获取信息必须经过传递。

(3) 信息不仅是传递中的知识，而且必须是有效的知识。凡是人类社会的信息均构成一定的效用，或与接收者原来知识状态吻合而被认可；或有相异，从而发生知识的重构，生成新的知识结构。

因此，可以说在社会交流过程中，知识变为信息，信息转化为知识，它们总是处在相互作用之中。知识是静态的，其以实践为基础，通过抽象思维，对客观事物的规律性给予了概括。知识信息是人类社会中客观存在着的，以知识形态为主，同时包括数据、新闻、消息等非知识形态的社会信息。知识信息是知识的激活，知识是知识信息的条理化。

通常所说的原始数据并不能称之为信息。原始数据和信息的主要区别在于：一切信息虽然都是由数据组成的，但并非一切数据都能用于消除不确定性而产生有用的信息。原始数据需要经过加工处理才能转变为信息，经过交流和传递最终实现信息的应用。

因此,只有科学地了解和认识信息,才能更好地把握信息,才能使信息更好地为科学技术和社会经济的发展服务。

1.1.2 信息的类型

从产生信息的客体的性质的不同,信息可分为自然信息和社会信息。

(1) 自然信息。自然信息是指自然界发出的信息,它以纯自然物为载体,如天体信息、生物信息等。它是客观物质世界过程的原型。

(2) 社会信息。社会信息是经过人类利用语言、文字、符号、图形、图像等方式加工而成的信息,是人际间传播的信息,是人类活动的产物,也是我们这门课程的主要研究对象。

按照信息所包含的范围的不同,信息资源可分为广义信息资源和狭义信息资源。

(1) 广义信息资源。广义信息资源是指信息和与操作信息有关的物理设施、人力、机构、资金和运行机制等的总称。国内外许多学者认为信息资源主要包括四个方面,即具有与信息相关的技能的人;信息技术硬件与软件;信息设施,包括图书馆、文献信息中心、计算机中心和通信网络等设施;信息处理和加工的支持系统。

(2) 狭义信息资源。狭义信息资源是指信息本身及其载体的集合,也就是我们常说的信息源。作为一般信息的使用者,并不需要对广义信息资源进行研究,但应该了解狭义的信息资源的各种情况,以利于信息的检索。

按信息内容的表现形式的不同,信息可分为文献型、数据型、声像型和多媒体等信息源。

(1) 文献型信息源。文献型信息源的信息内容是以文字形式储存在各种不同载体上的。文献信息源是目前信息内容最丰富、

人们使用频率最高的信息源。

(2) 数据型信息源。数据型信息源的信息内容是以数据形式出现的,并存储在各种不同载体上的信息集合。

(3) 声像型信息源。声像型信息源的信息内容是以声音或图像形式出现的,如广播、电视、CD 光盘以及 VCD 光盘等。

(4) 多媒体信息源。多媒体信息源是随着现代科学技术的发展而出现的一种新的信息源形式,它集文字、声音、图像于一体,多以光盘或 Internet 网上资源形式出现。

按照信息的载体形式划分的不同,信息可分为印刷型、缩微型、声像型和电子型等信息源。

(1) 印刷型信息源。印刷型信息源是指以纸张为载体,用印刷方式记录知识的文献。印刷型文献方便实用,是人类交流和传播信息的主要形式。它的优点是阅读方便,利于流通;缺点为存储信息密度低、体积大、分量重、占用空间大、不易保存、难于实现自动化、不利于资源共享。

(2) 缩微型信息源。缩微型信息源是指以感光材料为载体,用缩微照相的方式将文字、图形、影像等信息符号按比例缩小后存储在感光材料上,并借助于专用阅读器而使用的文献。目前最常用的是缩微胶卷和缩微平片。它们的优点是信息密度高,且信息位序固定,不会散失弄乱,记录速度快,规格统一,体积小,重量轻,传递方便,拷贝、放大、还原容易,保存时间长,成本低。缺点是制作、保存和使用条件严格,文字图像小,不能直接阅读,必须借助于专用阅读机。

(3) 声像型信息源。声像型信息源是指以磁性材料、光学材料等为载体,利用特定设备,使用声、光、磁、电等技术设备记录声音信息和图像信息的文献,如唱片、录音带、录像带、电影片、激光唱盘等。声像型信息源形象、直观,有些用文字难以描述的事物可以录音、录像或拍成电视片,帮助人们观察和认识某些自

然现象,探索物质结构及运动机制。缺点是利用时也需要专门设备。

(4) 电子型信息源。电子型信息源主要指通过使用电子技术实现信息存储和传播的信息源,包括广播、电视、光盘、磁盘、磁带、网络信息源等。电子型信息源的前身称机读型信息源,是以软盘、磁带、光盘等磁性介质为载体,用键盘输入或光学扫描等手段记录,并通过计算机处理后生成的一类文献。包括计算机文档、光盘数据库、电子图书、电子期刊、电子报纸、电子邮件和电传文本等。具有内容的广泛性、使用的简便性、搜索的网络性和资源的动态性的特点,同时具有存储密度高,存取速度快,原有记录可以改变、更新等优点。电子型信息源除具有与印刷出版物相当的文献文本之外,还可以提供多维的、有序化的可操作的功能;可以对文本进行有目的的抽取、排序、重新组合,从而产生新的信息产品;可以对文献内容各个知识单元,甚至字频进行计算分析,使文献计量分析得以深化。无论从哪一方面看,网络信息资源几乎胜过了以往所有的传统信息资源,因此,通过网络来获取所需的信息已成为当今获取信息的最主要途径。

按信息内容的保密程度的不同,信息可分为公开信息源、半公开信息源和非公开信息源。

(1) 公开信息源。公开信息源的信息是可以共享的。公开信息源的信息量很大,如大量的科技信息、经济信息、生活信息等,这类信息可以作为信息商品进入流通领域。

(2) 半公开信息源。半公开信息源也就是我们常说的内部信息,它只能在一定范围内使用。

(3) 非公开信息源。非公开信息源即保密信息,一般是不能公开使用的。根据保密程度这种信息又分为各种保密级别,并有一定的保密时间,如常说的商业秘密、国家机密等。

按信息传承形式的不同,信息可分为口头信息源、实物型信

息源和文献型信息源。

(1) 口头信息源。口头信息源指以口头语言,如交谈、聊天、授课、讨论等方式获得的信息资源,它是没有记录下来的仅靠口口相传的信息,其特点是传递迅速、互动性强,但稍纵即逝,久传易出差错。

(2) 实物型信息源。实物型信息源指以实物,如文物、产品样本、模型、雕塑等形式表现的信息资源。例如:秦始皇兵马俑、维纳斯雕像、人造卫星等都包含着大量信息,其特点是直观性强、感觉实在、信息量大。

口头信息和实物信息目前还没有加工存储和提供检索的工具,使用者只能根据自己的需要去专门地、有目的地搜集、加工、分析和利用。

(3) 文献型信息源。文献型信息源指以文字、图形、符号等各种方式记录在各种载体上的知识和信息资源。包括图书、期刊、报纸以及学位论文、专利、标准、会议、政府出版物等各种文献。这类信息经过加工、整理,较为系统、准确、可靠,便于保存与利用。这类信息是当前数量最大、利用率最高的信息资源,是我们获取信息的重要来源。

1.1.3 信息的主要特性与基本作用

1. 信息的主要特性

信息之所以区别于物质与能量,并具有与物质、能量同等的重要作用,是源于信息所拥有的特性。物质在使用中是消耗的,能量就其个体而言在使用中也是消耗的,就其整体而言则是永恒的。而信息在其传递和使用过程中,可以重复使用,并可通过信息的加工处理而产生信息增值。信息作为一种资源,主要具有以下特性。

(1) 客观性。信息的存在是客观的，它源于客观存在的物质及其运动。信息的客观性还表现为信息是以物质的客观存在为前提的，即使是主观信息，如决策、判断、指令、计划等，也有它的客观实际背景，并以客观信息为“原料”，受客观实践的检验。

(2) 依附性。信息总是依附于一定的物质载体而存在的，如果不依附各种适当的载体，信息的含义和价值则不能传递和发挥。声音、语言、文字、颜色、图像、各种符号、光电磁、生物等各种信息信号以及纸张、胶片、磁带光盘、人的大脑等，无一不是信息的载体。

(3) 可传递性。信息的产生是同信息的传递联系在一起的，信息在传递过程中发挥它的作用。信息的传递和流通过程是一个重复使用的流通过程，在这一过程中，信息的占有者不会因传递信息而失掉信息，一般说来，也不会因多次使用而改变信息的自身价值。信息在时间上的传递通常被称之为信息的存储。

(4) 可塑性。信息可以加工处理，可以压缩、扩充和叠加，也可以变换形态。在流通和使用过程中，经过综合、分析、再加工，原始信息可以变成二次信息和三次信息，原有的信息价值也可以实现增值。为了有效地交流和传递，借助于先进的信息技术，文本、图像、数字、语言等各种形态的信息均可实现互相转换。

(5) 时效性。现代社会中，信息的使用周期迅速缩短，信息的价值实现取决于及时地把握和运用信息。信息是活跃的，不断变化的，及时获取有效的信息才能获得信息的最佳价值，如时效性很强的天气预报、经济信息、交易信息、科学信息等。不能及时地使用最新信息，信息的价值就会随其滞后使用的时差而贬值。

(6) 共享性。信息的共享性主要表现在同一内容的信息可以在同一时间由两个或两个以上的使用者使用，而信息的提供者并不失去所提供的信息内容和信息量，它是信息资源的广泛提供与

利用的基础。

2. 信息的基本作用

作为一种知识交流和社会交流，信息在人类社会和科技发展中起着重要作用。

(1) 信息是人类认识客观世界及其发展规律的基础。信息的基本功能主要表现为信息的认识功能。它是辩证唯物主义认识论的基础，是揭示客观事物发展规律的重要途径。信息是客观事物及其运动状态的反映，客观世界里到处充满着各种形式和内容的信息。人类的认识器官对各种渠道的信息进行接收，并通过思维器官将已收集到的大量信息进行鉴别、筛选、归纳、提炼、存储而形成不同层次的感性认识和理性认识。在这一认识过程中，人类是认识论的主体，信息是认识论的客体。

(2) 信息是科学研究的必要条件。人类的知识具有继承性和共享性。科学的大厦是千万个科学家在历史的进程中逐步建立起来的，任何一位科学家在从事科学研究时，都不能不借鉴前人的成果和依靠同时代其他人的帮助，这就是说，他同时需要在时间上和空间上的信息传递。在这个问题上，自然科学家和社会科学家各具特色。自然科学研究是一个比较严格的循序渐进的过程，很多成果都是在前人建立的基础上发展起来的。由于自然科学研究的抽象性、准确性和严格性，独立研究是非常困难的，因此，信息交流就十分必要，并且对信息的数量和质量等都有较高的要求。对于社会科学而言，其研究成果不具有像自然科学研究成果那样的准确性，但在多样性上远远超过前者，并更多地受社会政治和人为因素的影响，其不同观点和理论的统一，较之自然科学也往往更困难和更需要时间，因此，需要更多地占有信息资源，充分地认识和把握各种不同的信息观点。