



面向 21 世纪 课程 教材

高等学校
信息管理类专业
信息核心课程教材

信息检索

焦玉英 符绍宏 何绍华 编著

武汉大学出版社

面向 21 世纪 课程 教材

面向 21 世纪课程教材

高等学校信息管理类专业核心课教材

信息检索

焦玉英 符绍宏 何绍华 编著

武汉大学出版社

图书在版编目(CIP)数据

信息检索/焦玉英,符绍宏,何绍华编著. —武汉:武汉大学出版社, 2001. 5

面向 21 世纪课程教材. 高等学校信息管理类专业核心课教材
ISBN 7-307-03150-7

I. 信… II. ①焦… ②符… ③何… III. 情报检索—高等学校—教材 IV. G252.7

中国版本图书馆 CIP 数据核字(2000)第 73190 号

责任编辑: 严 红

责任校对: 张 昕

版式设计: 支 笛

出版: 武汉大学出版社 (430072 武昌 珞珈山)

(电子邮件: wdp4@whu.edu.cn 网址: www.wdp.whu.edu.cn)

发行: 新华书店湖北发行所

出版: 武汉市新华印刷厂

开本: 880×1230 1/32 印张: 13.875 字数: 383 千字 插页: 2 插表: 1

版次: 2001 年 5 月第 1 版 2001 年 5 月第 1 次印刷

ISBN 7-307-03150-7/G · 534 定价: 18.90 元

版权所有, 不得翻印; 所购我社教材, 如有缺页、倒页、脱页等质量问题, 请与当地教材供应部门联系调换。

前 言

信息检索作为一个学科的历史可以追溯到 20 世纪中期。其有代表性者如 Granfield 所确立的标引语言及检索系统评价方案, Salton 提出的矢量空间模型与文献聚类技术, Roberson 及 Van Rijisbergen、Sparck Jones 等人研制的概率模型, Smeaton 在计算机语言学基础上的检索技术开发等的研究与试验, 都对直至 20 世纪 80 年代末期文献检索领域的研究对象、原则和方法产生过重要影响。尽管其研究成果和设计思想基本上是在一个模拟文献检索作出的人工或虚拟环境中进行的, 甚至还带有某种程度上的假说性, 但其研究成果却奠定了文献检索这门课程基于以“提问-检索”模式为核心的相关理论与方法的基础。传统的赋值标引、规范化的检索语言, 线性的书目数据库结构, 预定的检索策略, 以回答检索提问为主的服务方式等构成了该模式的主要概念体系、方法和原则。

20 世纪 90 年代以来, 以 Internet 为核心连接起来的全球计算机网络使传统的相对集中和规范为基准的文献数据库及其检索系统面临严峻的挑战。主要表现在: 信息资源内容之广泛涉及到各个学科领域乃至人类生活的方方面面; 信息类型以动态、静态、音频及超声频等多姿多彩的方式再现; 各类型、多种品牌的网络信息查询工具如雨后春笋般不断涌现; 网络化与数字化技术将分布在世界各地主机上的信息资源数据库联为一体, 为人们跨越时空、行业、地域快速、高效传递信息提供了国际化的知识信息平台, 极大地提高了获取信息的主动性。传统的“提问-检索”模式已逐步被网络

环境中“浏览-查询”模式取代。

尽管如此,网络信息检索在信息资源的组织与管理、系统工具的开发与设计思想、检索方法与策略等方面仍同传统的检索理论与方法体系有着相一致或是密切的联系。从理论与实践两个方面看,网络信息检索是传统文献检索理论与方法体系的拓宽与革新,是信息检索这门学科的一个重要分支。

本书共12章,由武汉大学传播与信息管理学院焦玉英、何绍华,北京师范大学信息技术与管理学系符绍宏共同完成。具体分工如下:第1、2、3、6、7、8、12章由焦玉英编写;第5、10、11章由符绍宏编写(研究生张萍、徐桂花分别参加了第5、10章的部分写作及资料搜集工作);第4、9章由何绍华编写。

本书的出版得到武汉大学信息管理学院院长马费成教授、北京师范大学康仲远教授的举荐、支持和北京师范大学袁名敦教授、倪晓健教授、武汉大学出版社副社长何皓的热情帮助,责任编辑严红为本书的出版付出了辛勤劳动,在此表示诚挚的谢意。

由于时间匆促且水平有限,对本书可能出现的缺点乃至错误,敬请广大读者批评、指正。

焦玉英

2000年9月

目 录

1 信息检索概述	1
1.1 信息的概念与特征	1
1.2 信息的类型与存储载体	4
1.3 信息的现代传输方式	9
1.4 信息检索的概念与类型	13
1.5 信息检索研究的核心问题	17
2 文献信息检索基础	21
2.1 文献信息类型演化及其结构形态	21
2.2 文献信息的揭示与组织	24
2.3 文献特征的描述	31
2.4 文献信息检索系统与检索工具	35
2.5 检索语言与索引系统	47
3 现代信息检索技术	63
3.1 全文检索	63
3.2 多媒体检索	68
3.3 超媒体及超文本检索	72
3.4 联机检索	74
3.5 光盘检索	77
3.6 网络信息检索	79

4 国内文献信息检索	87
4.1 国内文献信息检索发展概况	87
4.2 国内文献信息检索工具及其利用	91
4.3 国内检索刊物中各类文献的著录	96
4.4 中国主要信息网络系统	98
4.5 国内主要学术期刊数据库及检索方法	104
4.6 中国专利文献检索工具	115
4.7 中国学术会议文献检索工具	122
4.8 标准文献检索工具	128
5 国外综合型文献信息检索	157
5.1 国外文献信息检索系统概况	157
5.2 美国《化学文摘》	165
5.3 美国《生物学文摘》	176
5.4 英国《科学文摘》	182
5.5 美国《工程索引》	188
6 国外科技报告文献检索	194
6.1 科技报告文献概述	194
6.2 科技报告文献的检索工具	198
6.3 科技报告文献的检索方法	201
6.4 科技报告代号的检索	203
6.5 美国四大报告的网上检索	204
7 国外会议文献检索	207
7.1 会议文献概述	207
7.2 美国《会议论文索引》的检索	208
7.3 会议的网上检索	211

8 国外专利文献检索	214
8.1 专利文献概述	214
8.2 专利文献检索工具	217
8.3 美国专利文献检索方法	219
8.4 英国德温特专利文献检索方法	221
8.5 专利文献的网上检索	226
9 数据与事实参考工具	230
9.1 参考工具概述	230
9.2 百科全书	245
9.3 年鉴、手册	259
9.4 机构团体指南	264
9.5 传记资料	267
9.6 地理资料	270
9.7 统计资料	280
9.8 法规资料	285
10 计算机检索系统及其使用	290
10.1 重要国际联机检索系统	290
10.2 光盘检索系统	310
11 因特网信息检索	341
11.1 因特网信息资源	341
11.2 网络信息资源检索	353
11.3 各类网络信息检索工具	358
11.4 网络信息检索研究	403
12 信息检索与咨询服务及其效果评价	416
12.1 解答咨询服务	416

信息检索

12.2	定题服务·····	419
12.3	科技查新服务·····	422
12.4	网络信息资源查询辅导服务·····	424
12.5	检索效果评价·····	427
主要参考文献·····		432

1

信息检索概述

1.1 信息的概念与特征 ■

信息的概念十分广泛，围绕信息而出现的信息资源、信息技术、信息系统、信息产业、信息化社会和社会信息化等相关术语不胜枚举。可见，信息的观点、概念和方法已经被政治、经济、科技、文化、生产等各个领域所接受和应用。那么，究竟什么是信息？至今仍无确切、统一的定义，站在不同的角度就有不同的理解或解释。

从自然界角度看，客观事物的普遍属性通常可用它的运动性、时空性、能量、反映性、质料、系统性等表示。其普遍性既是事物千差万别的表现，又是事物之间相互联系的内容。信息概念的引入旨在表明，信息既不是物质，也不是能量，而是依附于自然界客观事物而存在，也就是说，只要有物质存在，就有表征其属性的信息。例如，地球昼夜的变化是一种信息，它反映了地球绕太阳自转的运动特性和状态；山的高度是一种信息，它反映出山的空间特性；树干的年轮是一种信息，它反映了树木成长的时间特性——树龄；闪电是一种信息，它反映了云层中所含能量的特性；花的香味也是一种信息，它反映了花分子结构的化学特性等等。因此，可以认为，信息的概念，实际上就是客观事物运动状态、时空特性、能量大小、质料、系统特征、相互联系方式等一切反映事物客观属性

的总称。从这种意义上看，信息比客观事物的属性更具一般性与普遍性。

从人的主观认识角度看，信息是储存在人脑中的思想、观念、知识等形态。它既是每个具体人这一特殊事物属性的一种表征，又是外部客观事物属性在人脑中留下的印记，是物质反映属性的高级形式。哲学家把人的这些直接接受客观事物信息的功能称为人的自然信息功能。人在自然界的活动即是不断排除信息传输过程中的噪声与干扰，选择、过滤和提取正确信息的过程，或者说，通过感觉、知觉等阶段对来自不同方位的信息进行综合，如将光、电、声、场、分子等各种介质的信息形态转换、传递，编码为大脑可以识别、存储的形式，并且不断借助各种信息技术来弥补人自身的自然信息功能缺陷，等等。

从技术角度看，信息的概念体现在一切人造工程、设备的技术和有关技术的特性之中。信息既是这些人造客观事物的表征，也是人的知识、文化及艺术水平等属性的反映。

从人工文化角度看，信息概念的实质在于它以某种编码形式储存或传输于某种介质之中，如存储在书本、纸张上的文字信息，记录在唱片、录音带上的声音信息，印制在画报、照片、录像带上的图像（形）信息，计算机系统中的各种数字、数据信息等。我们把存储在这些物理介质中的信息统称为文献信息。文献既是记录信息的物质载体，又是人类的精神文化产品。

图 1-1 标示出了客观信息与主观信息的相互关系。

图中，人从主观上认识世界以及与客观现实世界的信息交流活动是一种自人类社会产生以来在其整个历史过程中不断反复、发展的过程。这一过程不仅反映人类认识客观世界，逐渐逼近真理的过程，同时也推进着社会的信息化进程。

信息有如下的特征：

（1）客观性与普遍性

信息既不是物质，也不是能量，是客观事物普遍性的表征，信息是无处不在，无时不有的普遍社会现象。

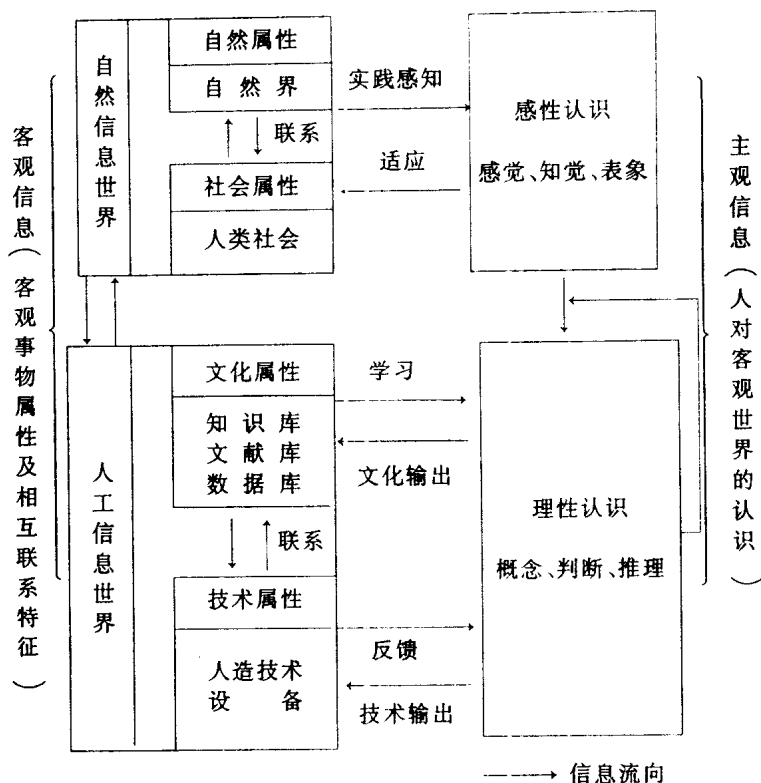


图1-1 客观信息与主观信息的相互关系

(2) 流动性与传递性

信息在事物之间的相互联系必定在信息的流动中发生。信息的传递性表现在人与人之间的消息交换，人与自动机、自动机与自动机之间的信息交换，动物界和植物界的信号交换，同时，人类进化过程中的细胞选择、遗传也被看做是信息的传递与交换。

(3) 多样性与综合性

信息在不同的领域具有多种不同的特性或表现形式，如客观事物中的各种自然属性；人工设备的技术特征；人类社会的各种社会特征；人脑中反映客观事物认识的思想、知识；人类交流信息过程

中的声音、文字、图像以及用各种编码形式记录下来的数据、新闻、情报、消息等。各种形式的信息又常常以综合的方式表现事物的特征，所谓“多媒体”正是信息多样性和综合性的集中表现。

(4) 相对性与有效性

从信息作为事物相互联系的反映角度看，信息源不确定的程度或者信息源接受信息量的多少，均与信宿的状态有关。这一特征在人作为信宿接受信息的过程中表现得尤为明显。同一信息对具有不同认知水平的人所产生的作用和有效性也不相同。

(5) 积累性与价值性

信息通过人脑思维或人工技术设备的综合、加工和处理，不断积累丰富，提高其质量和利用价值。信息的质量和值，实际上是对客观事物属性反映的深度和真实程度的认识。虽然信息是人类的一种重要资源，但信息只有被利用才会产生价值，否则，其价值或随时间的流逝而减少，或成为“信息垃圾”。

1.2 信息的类型与存储载体 ■

1.2.1 信息的类型 ●

信息与人类智能活动有关的知识、技术、科学、文化、社会等密切联系在一起，其涉及范围如此之广，以至于很难用统一的标准进行分类。

1.2.1.1 按信息表现形式划分

(1) 文字信息 文字是人们为了实现信息交流、通信联系所创造的一种约定的形象符号。广义的文字还包括各种编码，如ASCII码、汉字双字节代码、国际电报与单元代码以及计算机中的二进制数字编码等都是一些符号的约定。这些文字、符号、代码均是信息的表述形式，其内容再现于它们的结构属性之中。如基本笔画的不同组合，字和字母的不同组合，二进制码“0”和“1”的不同排列等，分别代表不同的信息内容。

(2) 图像信息 图像(形)是一种视角信息,它比文本信息直接,易于理解。人工创造的图像(形),如一张纸、一幅画、一部电影,大自然的客观景象等都是抽象或间接的图像信息。随着多媒体技术的发展,各类图像信息库将会极大地丰富人类生活。

(3) 数值数据信息 数值数据是“信息的数字形式”或“数字化的信息形式”。狭义的“数据”是指有一定数值特性的信息,如统计数据、气象数据、测量数据以及计算机中区别于程序的计算数据。广义的数据是指在计算机网络中存储、处理、传输的二进制数字符编码,文字信息、图像信息、语言信息以及从自然界直接采集的各种自然信息等均可转换为二进制数码,网络中的数据通信、数据处理和数据库等就是广义的数值数据信息。

(4) 语音信息 人讲话实际上是大脑的某种编码形式的信息转换成的语言信息的输出,是一种最普遍的信息表现形式。音乐也是一种信息形式,是一种特殊的声音信息,它是通过演奏方式表达丰富多彩的信息内容的。

1.2.1.2 按信息的出版类型划分

(1) 图书 包括专著、教科书、各种科普读物及各专业参考工具书等。图书经编著者精心选择,反复斟酌后写成,其内容系统、成熟、定型,信息经筛选,可靠性强,是人们从事学习、研究不可缺少的信息来源。不过,传统印刷业图书出版周期较长,体积大,更新速度慢,电子版图书的出现将弥补这一缺陷。

(2) 期刊 指按同一专业领域定期或不定期出版的连续性出版物。它出版数量大、周期短、内容新颖,能迅速反映国内外的各种学科专业的水平和动向,是人们获取一般基础理论研究知识的重要信息源。期刊按内容划分,有综合性的与专业性的;按性质划分,有学术性的、技术性的、消息性的、检索性的以及通报性的;按对原文压缩程度划分,有目录、索引、文献、快报、速评、文献指南、书目之书目等。

(3) 政府出版物 指各国各级政府部门及所属机构出版的文献信息资料,它主要包括社会科学与自然科学两大类。其中行政文

信息检索

件，如讨论会记录、各种法令、外交文件、统计数据占大多数，科技资料数量相对较少。

(4) 科技报告 指各学术团体、科研机构、大学研究所的研究报告及其研究过程中的记录。科技报告理论性强，是了解某一领域科研进展状况、发展动态的重要情报源。但科技报告保密性强，难以获取。

(5) 专利文献 指发明人向政府部门（专利局）递交的、说明自己的创造的技术文件，同时也是实现发明所有权的法律性文件。专利文献包括专利说明书、专利公报（摘要）、商标、设计公报以及检索专利的工具等。专利文献具有技术性、新颖性、独创性、实用性等特征，是重要的技术经济情报来源。

(6) 会议文献 指在国内外学术团体举行的专业会议上发表的论文与报告。与期刊相比，会议文献具有传播情报信息更迅速的功能。它反映了某学科、专业的最新成果和发展水平动向，是科研工作不可缺少的情报源。

会议文献一般分为会前、会中和会后三种形式。会前资料如会议通知、会议程序单、论文摘要；会中资料如开幕词与闭幕词、会议决议书等；会后资料如会议结束后经整理出版的专门会议丛刊、会议论文集等。

(7) 学位论文 指高等院校研究生（硕士或博士）攻读学位而撰写的毕业论文。它经专家评审、鉴定通过，一般来说具有学术性强的特点，往往有独到的见解。

(8) 技术标准和规范 主要指技术规范、技术标准、操作规程、建议、准则、术语、专门名词等在内的各种技术文件。在标准实践领域里，技术标准和规范在适用范围方面是有区别的。前者是一种得到管理机构认可，适用于一定专业领域的技术规范，具有法定性；后者是指对产品、材料、工艺流程或技术特点的说明书，它仅以满足买方或工业规定的要求为准则。

技术标准主要包括尺寸标准、材料标准、性能标准、方法标准、操作规程、术语和图形符号标准、文献标准等。

(9) 产品样本说明书 是制造厂家和产品销售者介绍其产品的宣传性出版物。它介绍的是已投产和行销的产品。通过产品样本说明书可以了解厂家的工艺水平、管理水平和产品发展趋势方面的情报。由于产品样本说明书附大量图表、产品特性曲线、方程等,因此具有直观的特点。同时,厂家为了推销产品,往往免费赠送,使产品样本说明具有易于获取的优点。

产品样本说明书除直接出版发行之外,还常常被包含在一些贸易刊物、企业介绍、数据手册之中。

(10) 技术档案 是在科技生产活动中形成的一系列以工程技术图纸、任务书、协议、合同、设计方案以及与此有关的调查统计数据等材料组成的文件。技术档案具有技术性、适用性、保密性等特征。

1.2.1.3 按信息的加工程度划分

(1) 一次信息 一次信息是人们研究或创造性活动成果的直接记录,一般指公开出版的图书、期刊论文、科技报告、会议文献、学位论文、发明专利等。不管其信息存储于何种物质载体及出版的版次,只要是原始资料就是一次信息。一次信息零碎、分散、无序,有的很难获取。

(2) 二次信息 二次信息是对一次信息加工、整理而成的。如目录、文献、索引等各种书目数据库是二次信息的核心。二次信息的形成是信息从分散、无序到集中、有序化的书目控制过程。

(3) 三次信息 三次信息是对一、二次信息综合、分析等深加工的产物。如评论、进展报告、述评、百科全书、年鉴、指南、期刊书目等。

1.2.2 信息的存储载体 ●

1.2.2.1 印刷型

它是以手写和印刷技术为手段,以纸张记录信息的载体形式。它的优点是可以直接阅读,携带方便,因此成为人类科研、生产、学习、文化交流等活动中最常用的工具。与现代信息载体相比,印

刷型信息载体存储信息密度小, 占用收藏空间大, 难以长期保存。

1.2.2.2 缩微型

缩微型主要指以感光材料记录文字及其相关信息的载体。常见的有缩微胶卷和缩微胶片。缩微型信息载体的优点是便于保存、转移和传递, 缩小文献的体积, 可节约书库面积达 95% 以上, 而其成本却只是印刷型的 1/10 左右, 但缩微型信息载体必须借助阅读机或阅读复印机才能使用。世界上许多大型文献信息中心都将学位论文、科技报告等文献制作成缩微品加以收藏和保存。

缩微型信息载体借助电子技术、计算机技术进一步增强了自身的功效。例如, 它可以制作成计算机存取载体的输入胶片 (Computer Input Microfilm—CIM) 和输出胶片 (Computer Output Microfilm—COM)。

1.2.2.3 声像型

它指记录声音、图像信号的信息载体, 如录音带、录像带、幻灯片、影视片以及近年来推出的高密度视、听光盘。声像型信息载体可以让人们通过自己的视觉、听觉感受到直观、形象、生动、逼真、丰富多彩的信息世界。

1.2.2.4 电子型

它是指采用电子手段并以电子形式存在, 利用计算机及现代通讯方式提供信息的一种新兴载体。它的前身是机读型。

电子型出版物内容丰富, 类型多。按信息存储介质划分, 主要有软磁盘与光盘两大类; 按出版物类型划分, 主要有电子期刊、电子图书、电子报纸、电子名录、电子地图、各种联机信息库和光盘数据库产品或磁带、软盘等产品; 按媒体的信息结构组织形式划分, 主要有以线性顺序组织知识单元的文本型出版物和以节点和链路组织知识的超文本出版物, 以及融文本、图像、声音信息于一体的多媒体出版物, 综合超文本与多媒体技术特点产生的超媒体出版物等。

电子出版物的问世是信息时代的重要标志, 它改变了书刊的物理形态, 开辟了一种新的信息分发渠道, 极大地提高了信息的传递