

21

世纪 高职高专规划教材

现代信息检索

袁学松 宋雯斐 主 编 岳国庆 陈 暄 副主编 施风芹 主 审

21SHIJIGAOZHIGAOZHIGUANGUIHUAJIAOCA



中国水利水电出版社
www.waterpub.com.cn

21 世纪高职高专规划教材

现代信息检索

袁学松 宋雯斐 主 编

岳国庆 陈 暄 副主编

施风芹 主 审

中国水利水电出版社

内 容 提 要

本书共9章,讲述现代信息检索的必备知识。主要内容包括:信息检索基本概念、实体图书馆和数字图书馆、事实和数据检索、网络信息资源检索、国内著名数据库系统、国外著名数据库系统、特种文献检索、科技查新与科技论文写作、信息检索综合实例。每章后附实验安排和练习题。不仅为初学者提供了一个学习现代信息检索方法与技能的空间,同时也为人们学习利用现代信息检索技术、全方位获取有关信息提供了相关的知识和必备的技能。

全书结构合理,内容全面、系统,取材新颖,注重实用性,是一本通用性、实用性很强的学习信息检索方法和技能的教材。

本书可作为高职高专学校各专业信息检索课程的教材,亦可作为广大信息用户与企业的科研人员、管理人员、工程技术人员从事文献信息管理的实用参考书。

本书所配电子教案可以从中国水利水电出版社网站上免费下载,网址: <http://www.waterpub.com.cn/softdown/>。

图书在版编目(CIP)数据

现代信息检索 / 袁学松, 宋雯斐主编. —北京: 中国水利水电出版社, 2007
21 世纪高职高专规划教材
ISBN 978-7-5084-4299-0

I. 现... II. ①袁... ②宋... III. 情报检索—高等学校: 技术学校—教材 IV. G252.7

中国版本图书馆 CIP 数据核字 (2006) 第 149881 号

书 名	现代信息检索
作 者	袁学松 宋雯斐 主 编 岳国庆 陈 暄 副主编 施风芹 主 审
出版 发行	中国水利水电出版社 (北京市三里河路 6 号 100044) 网址: www.waterpub.com.cn E-mail: mchannel@263.net (万水) sales@waterpub.com.cn 电话: (010) 63202266 (总机)、68331835 (营销中心)、82562819 (万水)
经 售	全国各地新华书店和相关出版物销售网点
排 版	北京万水电了信息有限公司
印 刷	北京蓝空印刷厂
规 格	787mm×1092mm 16 开本 12 印张 290 千字
版 次	2007 年 1 月第 1 版 2007 年 1 月第 1 次印刷
印 数	0001—4000 册
定 价	18.00 元

凡购买我社图书,如有缺页、倒页、脱页的,本社营销中心负责调换

版权所有·侵权必究

前 言

我们面临的是一个崭新的信息社会,信息已成为当今社会发展的重要战略资源。信息检索作为专门研究信息存储与信息获取的学科,对于提高大学生的信息素养和信息获取能力具有重要的意义。

信息检索源于传统的文献检索。信息技术的进步,尤其是因特网的发展,带来了信息传播与存储方式的极大变革,对信息检索产生了深刻的影响。检索对象从传统的纸质文献扩展为数量庞大、类型多样的信息,其中,数字信息、在线信息、多媒体信息越来越多地受到关注;检索工具除了传统的字典、词典、索引、书目、百科全书、年鉴等纸本工具书外,Google、百度等搜索引擎的使用则更为普遍;检索语言突破了传统的主题法和分类法,自然语言应用显现出强劲的趋势;检索效果的评价也将面临新的挑战。

本书力求探究网络环境下信息检索的新变化,系统介绍信息检索的原理和方法以及各类型检索工具的使用,反映信息检索领域的新动态、新方法,推进信息检索教育。

与现有的同类教材相比,本书具有如下特色:

(1) 体系结构新颖。本书采用检索技术与检索方法相结合、理论与实训相结合、综合信息的检索与学科领域的信息检索相结合的体系结构,各章节又相对独立。

(2) 检索方法详细。本书用大量篇幅介绍了网络信息资源检索及著名的数据库系统检索,特别是对维普、万方和 CNKI 数据库系统的概述、方法、使用途径、使用步骤等都作了详细的讲述。

(3) 编写目的明确。本书侧重于用现代化手段进行信息检索,学生经过系统学习后,可强化信息意识,提高信息敏感性和信息素养,使学生既可成为信息的提供者,又可成为高效的信息接收者。

本书由袁学松、宋雯斐担任主编,岳国庆、陈暄担任副主编。第1章由袁学松编写,第2章和第7章由宋雯斐编写,第3章由陈暄编写,第4章由岳国庆编写,第5章由易秋香编写,第6章由朱玲编写,第8章由邢红刚、杜艳明编写,第9章由沈晓璐编写。全书由袁学松、宋雯斐统稿,施风芹老师审读了全书,并提出了许多宝贵意见。在本书的编写过程中曾得到沈才梁、孙兰兰、宣宾等的大力帮助,中国水利水电出版社计算机编辑室的编辑们也给予了热情的支持,在此一并表示感谢。

由于时间仓促及作者水平有限,书中难免有疏漏、不妥乃至错误之处,恳请广大读者批评指正。

编 者

2006年12月

目 录

前言

第1章 信息检索基础.....	1
1.1 有关信息检索的一些概念.....	1
1.1.1 信息、文献、知识	1
1.1.2 信息检索的定义和分类	6
1.1.3 信息检索的几个技术指标.....	7
1.2 信息检索的类型及特点.....	8
1.2.1 手工信息检索	8
1.2.2 脱机信息检索	8
1.2.3 联机信息检索	9
1.2.4 光盘信息检索	9
1.2.5 网络信息检索	10
1.3 信息检索的基本步骤	11
1.3.1 课题分析, 明确检索范围.....	11
1.3.2 检索系统和检索工具	11
1.3.3 确定检索点和检索途径	12
1.3.4 构造检索式	12
1.3.5 选定检索方法	12
1.3.6 索取原始文献	13
1.4 检索工具	13
1.4.1 检索工具的概念及作用	13
1.4.2 检索工具的基本类型	13
1.4.3 检索工具的构成	13
1.4.4 检索工具的功能	14
1.5 检索语言	14
1.5.1 检索语言概述	14
1.5.2 检索语言的类型	15
1.6 信息检索的发展趋势	16
1.6.1 信息检索的最新技术	16
1.6.2 新一代信息检索的特点	18
1.6.3 信息检索展望	18
本章实验安排	19
本章练习题	19
第2章 实体图书馆、数字图书馆.....	21
2.1 图书馆的类型和功能	21

2.1.1	图书馆的类型	21
2.1.2	图书馆的功能	23
2.2	实体图书馆的借阅方式及检索步骤	24
2.2.1	图书分类和图书馆目录	24
2.2.2	馆内阅览	27
2.2.3	外借服务	28
2.2.4	参考咨询	28
2.3	数字图书馆	29
2.3.1	数字图书馆的概念及特征	29
2.3.2	电子阅览室的使用	30
2.3.3	超星数字图书馆	30
2.3.4	方正 apabi 数字图书馆	33
	本章实验安排	35
	本章练习题	36
第 3 章	事实、数据检索	37
3.1	概述	37
3.2	字典、词典	37
3.2.1	印刷型字典、词典介绍	37
3.2.2	在线字典、词典	39
3.3	百科全书	44
3.3.1	综合性百科全书及其检索	45
3.3.2	专科性百科全书及其检索	49
3.4	名录	50
3.4.1	人名录	51
3.4.2	地名录	52
3.4.3	机构名录	53
3.5	年鉴	55
3.6	书目	55
3.7	文摘	56
3.8	索引	56
3.9	附录	58
	本章实验安排	58
	本章练习题	59
第 4 章	网络信息资源检索	60
4.1	概述	60
4.1.1	网络信息资源的定义和特点	60
4.1.2	网络信息资源的类型	61
4.1.3	网络信息检索的一般方法	62
4.2	网络检索工具——搜索引擎	62

4.2.1	搜索引擎的概念	62
4.2.2	搜索引擎的基本原理	63
4.2.3	搜索引擎的类型	64
4.3	综合性搜索引擎	66
4.3.1	Google 的使用	66
4.3.2	百度的使用	69
4.3.3	新浪爱问的使用	71
4.3.4	搜狐搜狗的使用	72
4.4	专业性搜索引擎	73
4.4.1	工作搜索引擎	74
4.4.2	论坛搜索引擎	75
4.4.3	读书搜索引擎	76
4.4.4	法律搜索引擎	77
4.4.5	软件搜索引擎	77
4.5	特色搜索引擎	78
4.5.1	中文视频搜索	78
4.5.2	游戏搜索引擎	79
4.5.3	购物搜索引擎	80
4.5.4	Web 2.0 搜索引擎	81
4.6	附录——搜索引擎网站	82
	本章实验安排	83
	本章练习题	83
第 5 章	国内著名网络数据库系统	84
5.1	维普数据库系统	84
5.1.1	维普数据库系统概述	84
5.1.2	维普数据库系统检索途径	86
5.1.3	维普数据库检索步骤	91
5.1.4	维普数据库检索实例	93
5.2	CNKI 中国期刊网数据库	94
5.2.1	CNKI 数据库概述	95
5.2.2	CNKI 数据库的具体使用方法	97
5.3	万方数据资源系统	104
5.3.1	万方数据库系统介绍	105
5.3.2	万方数据资源系统的具体使用	108
	本章实验安排	112
	本章练习题	113
第 6 章	国外的著名数据库系统	114
6.1	综合性数据库系统	114
6.1.1	DIALOG 数据库系统	114

6.1.2 OCLC 数据库系统.....	118
6.2 专业检索工具	124
6.2.1 美国《工程索引》	124
6.2.2 美国《科学引文索引》	128
6.2.3 美国《化学文摘》	133
本章实验安排	138
本章练习题	138
第7章 特种文献检索	139
7.1 专利文献及其检索	139
7.1.1 专利、专利文献基础知识.....	139
7.1.2 专利文献检索	142
7.2 标准文献及其检索	147
7.2.1 标准基础知识	147
7.2.2 标准文献检索	150
7.3 学位论文及其检索	153
7.3.1 学位论文基础知识	153
7.3.2 学位论文检索	153
7.4 会议文献及其检索	155
7.4.1 会议文献基础知识	155
7.4.2 会议文献检索	155
7.5 附录	156
本章实验安排	157
本章练习题	157
第8章 科技论文写作	158
8.1 科技论文写作国家标准.....	158
8.2 科技论文写作	160
8.2.1 论文选题	160
8.2.2 开题报告	161
8.2.3 科技论文的内容	161
8.2.4 论文的修改	168
8.3 科技查新	168
8.3.1 科技查新步骤	169
8.3.2 科技查新报告	170
8.3.3 附件	171
本章实验安排	171
本章练习题	171
第9章 信息检索综合实例	172
9.1 如何检索有关的就业信息.....	172
9.1.1 求职信息包含的内容	172

9.1.2	求职信息的传统获取途径.....	172
9.1.3	网络招聘网站类型	173
9.1.4	网络求职和招聘	173
9.1.5	网络招聘信息的搜集	174
9.1.6	网络求职陷阱的防御	175
9.1.7	求职准备	175
9.1.8	网上求职网站的类型及推荐.....	176
9.2	如何检索信息检索课程的相关资料.....	177
9.2.1	有关信息检索网站	178
9.2.2	使用搜索引擎搜索信息检索课程信息.....	179
9.2.3	使用图书馆和联机数据库系统搜索图书、论文信息.....	181
参考文献		183

第1章 信息检索基础

1.1 有关信息检索的一些概念

1.1.1 信息、文献、知识

1. 信息的定义、特征和分类

信息技术的发展突飞猛进,信息走进生产,走进生活,信息正在推动经济,改变社会。信息已经与物质和能源一起称为现代人类社会的三大资源。但由于信息的概念十分广泛,目前仍然是众说纷纭,并无严格定义,不同学者从不同角度对信息做出各种定义。例如,在文献信息学中,信息往往被理解为知识内容,而心理学家认为信息不是知识,知识存在于人们的大脑中,而信息是存在于人们的意识之外的东西,如自然界、印刷品、硬盘以及空气中。而在管理学界,信息则被理解为管理活动的特征及其发展情况的统称。图书情报学家则认为,信息可以定义事物和记录,记录所包含的信息是读者通过阅读而获得的。我国《辞海》载:“信息是指对消息接受者来说预先不知道的报道。”我国国家标准《情报与文献工作词汇基本术语》(GB4894-85)载:“信息是物质存在的一种方式、形态和运动状态,也是事物的一种普遍属性,一般指数据,消息中所包含的意义可以使消息中所描述事件的不定性减少。”信息论的奠基人申农:“信息是用来消除不定性的东西。”控制论专家维纳:“信息就是人与外界互相作用过程中相互交换的内容和名称。”我国学者周怀珍:“信息是物质和能量在空间和时间中分配的不均匀程度。”从上述种种说法可以看出,不同的使用者从实用角度对信息的理解可能出现很大的差异。信息作为一门学科,它涉及领域广,内容丰富,很难有一个统一的定义。为了便于理解,这里我们综合各个定义的合理的内核,将信息的定义概括为:信息是所有事物的存在方式和运动状态的反映,是频率变化的一种感知,并且总是借助于一定的物质载体传输和存储(或者说信息是指对消息接受者来说预先不知道的报道)。对于这个定义可以从两个方面理解:其一,客观世界中不同的事物是千差万别的,呈现出不同的状态和特征,从而在人们认识的过程中形成不同的信息;其二,人们感知外部世界通常是通过声频或视频的变化而产生的,正如我们能听到雷鸣、鼓乐,能看到闪电、图像……应该说,信息是有实质内容的,并不是虚无飘渺的东西,不是人们主观想象的。

分析研究信息的特性不仅有利于加深对信息概念的理解,而且更有助于对信息资源的利用。从信息内容及存在形式上看,信息具有以下一些主要特性:

(1) 传递性。传递性是信息的本质特性之一。所谓信息的运动过程,就是信息的传递与反馈过程,信息的获取必须依赖于信息的传递。信息的传递是与物质和能量的传递相关的,在传递的过程中必将伴有一定的物质运动的传递或变换。同时,信息的传递并不是被动地依赖于物质及其间的相互反映,人们进行信息交流和传递的时候,可以根据自己的目的选择适当的物质载体及其运动形式。客观存在的信息,通过人的智慧被感知以后,被转换成语言、文字、图

形、代码等可传递、可接收的形式,并且依据一定的物质进行存储和传输,所以信息具有传递性。信息可以全向传递,也可以定向传递。由于信息的传递性,使信息有可能在短时间内广泛扩散。

(2) 时效性。信息具有较强的时效性。信息的时效性是指如果信息不能及时反映事物存在的方式和运动状态,那么这一信息就会失去效用。客观事物总是不断地发生变化,因而信息的时效性也是不断变化的,如时效性很强的天气预报、股票信息、交易信息、科学信息等。如不能适时使用最新信息,信息就会失去其原有的价值。

(3) 共享性。信息不仅可以传递、转换,而且可以共享。信息的共享性是指同一内容的信息可以在同一时间被两个或两个以上的用户使用,并且信息的提供者并不因为将信息提供他人共用而失去原有信息的内容和信息量。我们对信息的收集、整理、加工、存储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。

(4) 客观性。信息是客观存在的,信息的产生源于物质,信息产生后又必须依附于物质。就世界整体而言,信息的根源是物质世界,信息是物质的基本属性,由于物质是客观存在的,所以信息的存在也是不以人们的意志为转移的。

其他的特性还有开发性、存储性、价值的不定性、变换性、可伪性等。

信息按照不同标准可作如下分类:

(1) 以存在环境为依据分为:自然信息、生物信息、机器信息(自动控制系统)和(人类)社会信息。

(2) 以认识主体为依据分为:

- 客观信息:是指对事物不加判断地如实和公正的报道,即关于认识对象的信息。
- 主观信息:一般是依据事实和分析阐明个人对论题的观点和见解,是经过思维主体加工的信息。

(3) 以信息的逻辑意义和利害关系划分:

- 有益信息:对社会发展有积极作用的、能够消除人们对未知事物不确定性的信息,它是人类社会的资源和财富。
- 无用信息:指对信息使用者所从事的某种活动没有作用的、多余的信息。衡量无用信息可以使用以下 3 个指标:一是德国文献学家瓦斯提出的废页率,二是美国信息科学家彭德尔伯里和加菲尔德等提出的不引率,三是信息冗余度,即信息中超出完整要求而在传递中属于多余的内容。信息的冗余度 $R=1-H/H_{\max}$ 。
- 有害信息:指对社会发展和信息用户有消极和阻碍作用的不真实或庸俗、媚俗的信息,主要有虚假信息和色情信息等。

(4) 按信息表现形式划分为:文字信息、图像信息、数值数据信息、语音信息。

(5) 按信息的媒体形式分为:

- 文本信息:按线性顺序排列,阅读时人们跟随文本的线性流向吸收其中的知识单元。
- 超文本信息:按知识单元及其关系建立的知识结构网络。其数据库由结点和链路组成,查阅超文本信息时,以知识片段及其关系作为追踪、检索的依据。
- 多媒体信息:多媒体是包括文本、图像和声音在内的各种信息表达或传播形式的总称。多媒体信息系统能针对用户的需求提供各种形式的信息。

- 超媒体信息：是指通过计算机控制，把各种文献载体和各种内容综合为一个整体，从而向读者提供各种形式和各种内容的资料。超媒体是超文本与多媒体两种技术的结合，即在信息浏览环境下超文本的信息管理方式与多媒体的信息表现方法结合在一起时就称为超媒体。

2. 知识的定义和分类

知识，是人类对于客观世界的认识。在这里，知识作广义的理解，包括消息、情况、事实、数据。知识可分为理性知识和感性知识。理性知识，是对客观事物的本质和规律性的认识，是经过思维、逻辑加工的知识，构成知识体系；感性知识，是对客观事物的描述和对现象、事实的感知，是未经逻辑加工的知识。

知识的存在必须有一定的物质形式。人脑、文献、实物这三种知识载体都是物质的。人们为了进行知识的传递和交流，还必须使知识具有能为感觉器官所感知的形式，即借助于文字、语言、符号、代码、电磁波、图像和实物等形式加以表现。这种表现形式也是信息。

人们获取知识的途径有两个：一是直接来源于产生信息的客观事物，二是通过信息载体或媒介（文献、电视、广播、他人等）的传递、交流而间接获得。然而，获得的信息能否转化为知识，转化得是否充分、完整，则取决于受主的认知能力。例如，人们对于卫星照片、气象云团的识读能力是不同的，所获得的信息的量和质将会有差别。

但是，今天人们感叹自身知识的亏空，原因不在于世界上缺乏所需的基本的知识和大量准确的信息，而在于知识的存储过于庞大和无序，其集中体现是出版物的泛滥。知识存储的无序化加剧了“信息污染”的程度，堵塞了通向知识之门的道路，耗费了大量探索自然和社会规律的精力，以至于人们不得不认真学习和研究获取信息的方法，掌握从大量无序知识中搜索对自己有用的、准确的知识的技能。因此，信息检索知识已经构成知识体系中一个不可缺少的部分。面临知识的无限性与个人能力的有限性、教育的时滞性与社会发展的多变性、书本的陈旧性与生活的现实性等矛盾，我们必须增强获取新知识的技能。

知识按不同标准可作如下分类：

- (1) 依照反映对象的深刻性和系统性程度，知识分为生活常识和科学知识。
- (2) 按知识的成熟程度，分为经验知识、理论知识。
- (3) 按领域：哲学知识、社会科学知识和自然科学知识。
- (4) 按用途：科学知识、技术知识、文化知识。
- (5) 按水平：低级知识、中级知识和高级知识。
- (6) 按性质：普通知识、专业知识。
- (7) 按载体：隐性知识、显性知识。

3. 文献的定义和分类

文献是记录有知识和信息的一切载体，是固化在载体（纸张、光盘）上的知识。文献是物化的精神产品，或者说，文献是知识信息的物化形态。其中，信息内容是文献的知识内核，载体材料是文献的存在形式和外壳，而符号系统和记录方式则是二者联系的桥梁和纽带。这4个要素相互联系就构成了文献的四维框架结构。

记录方式：是指将文献符号系统所代表的信息内容通过特定的人工记录手段和方法使其附着于一定的文献载体材料上。文献记录方式具体包括刻划、书写、印刷、拍摄、录制、复印和计算机录入等。

载体材料:是可供记录信息符号的物质材料,是全部信息载体中一个重要的子系统。文献载体反映了文献物质产品的性能,具有商品、保存和流通价值。文献载体大体经历了泥板、纸草、羊皮、蜡板、甲骨、金文、石头、简牍、缣帛等早期载体,到纸的出现,再到现代各种新兴文献的发展过程。

根据其内容性质和加工程度的不同,文献可分为以下4个级别:

(1) 零次文献:未经出版发行的文献,包括手稿、个人通信、原始记录等。

(2) 一次文献:首次出版的各种文献,也称原始文献。如期刊论文、科技报告、会议论文、专利说明书等。一次文献是以科研生产活动的第一手成果为依据而创作的文献,内容丰富、参考价值大,是我们利用的主要对象。

(3) 二次文献:报道和查找一次文献的检索书刊,如各种目录、题录和文摘等。二次文献是图书情报工作者在大量收集原始文献的基础上,经过分析、归纳、重组后出版的。二次文献是一次文献的集约化、有序化的再次出版,是储藏、利用一次文献的主要的、科学的途径。

(4) 三次文献:利用二次文献提供的线索,选用大量一次文献的内容,经综合、分析和评述再度出版的文献,如各种述评、进展报告、动态综述、手册、年鉴和百科全书等。我们使用的各种教科书也属三次文献。

上述的零次文献由于没有进入出版、发行和流通这些渠道,收集利用十分困难,一般不能作为我们利用的文献类型。而后3种文献是一个从分散的原始文献到系统化、密集化的过程。一般来说,一次文献是基础,是检索利用的对象;二次文献是检索一次文献的工具,故又称之为检索工具;三次文献是一次文献内容的高度浓缩,也是我们利用的一种重要情报源。

文献按存储载体可分为:

(1) 印刷型:通过铅印、油印和胶印等手段,将知识固化在纸张上的一类文献。例如,图书、期刊以及各种印刷资料。这是一种有着悠久历史的传统文献形式,至今仍广为应用。它的主要优点是便于阅读和流传;缺点是体积大、信息密度低。以纸张为存储介质,用印刷技术出版,如书、刊等。

(2) 缩微型:以印刷型文献为母本,采用光学摄影技术,把文献的影像固化在感光材料上的一类文献。常见的缩微型文献有缩微平片和缩微胶卷两种。这种文献的优点是体积小、信息密度高;缺点是阅读必须借助阅读机或利用缩微复印机。

(3) 声像型:这是一种非文字形式的文献。常见的有各种视听资料,如唱片、录音带、电影胶片、激光声视盘(CD-ROM)、幻灯片等。这类文献记录声音和图像,通过唱机、录音机、录像机、放影机和投影机等予以重现,可以使人闻其声,观其形。

(4) 机读型:这是近年来由于计算机的广泛应用而产生的一种新文献。常见的是各种已录有内容的磁带、磁盘和光盘。这种文献的存储、阅读和查找利用都必须通过计算机才能进行,所以既有信息量大、查找迅速的优点,又有设备昂贵、使用费用高的缺点。

按出版方式,文献可分为:

- **图书(Book):**对已有的研究成果或生产经验作概括论述。特点:带有总结性、成熟定型;出版周期长、信息传递慢;传授知识,而不是报道最新情报。包括教科书、专著、词典、百科全书、手册。
- **期刊(Periodical):**特点:名称固定;定期出版并计划无限期出版;出版周期短、报道快;数量大、内容丰富。

- 报纸 (Newspaper): 出版周期最短的定期连续出版物。特点: 内容新、涉及面广、读者最多、影响面广。
- 专利文献 (Patent Literature): 记录有关发明创造信息的文献。特点: 详尽、内容广泛。专利说明书既是技术文件又是法律文件。
- 标准文献 (Standards Literature): 对工农业产品和工程建设的质量、规格、检验方法及其技术要求等方面所做出的技术规定的文件形式。一般分为国际标准、地区标准、国家标准、行业标准、企业标准。
- 会议文献 (Conference Literature): 在各种会议上宣读和交流的论文、报告和其他有关资料。包括国际性会议、全国会议、地区会议和基层会议等。特点: 专业性强、内容新、学术水平高、出版发行较快。
- 科技报告 (Scientific & Technical Report): 对科学技术研究结果的报告或研究进展的记录。目的: 向上级主管部门汇报。特点: 每份单独成册, 有专门的编号; 内容具体, 有科研项目研究方案、实验记录、实验数据、图表等; 报道比期刊早; 保密或控制发行。

例如, 美国政府的四大科技报告:

- PB 报告: 美国商业部。1945 年 6 月, 美国成立商务部出版局 (Office of the Publication Board), 负责整理公布从战败国获取的科技资料, 并编号出版, 号码前统一冠以 PB 字样。20 世纪 40 年代的 PB 报告主要为战败国的科技资料; 20 世纪 50 年代起, 主要是美国政府科研机构及其有关合同机构的科技报告; 20 世纪 70 年代以后, PB 报告侧重于民用工程技术。
- AD 报告: 1951 年 5 月, 美国成立 Armed Service Technical Information Agency, 负责收集、整理、编辑、出版国防部所属海陆空三军军事系统科研机构及其与该部订有合同的工业企业、高等院校提出的军事科研报告, 在该部规定的范围内发行。AD 报告即是这个情报局出版的文献。
- NASA 报告: 美国宇航局。
- DOE 报告: 美国能源部。
- 学位论文 (Dissertation for Academic Degree): 作者为获取某种学位而撰写的科学论文。硕士博士论文具有较高的参考价值, 一般侧重于理论, 附有大量的参考文献, 借此可以看出有关专题的发展过程 and 方向。学位论文一般不出版发行, 而是保存在授予学位单位的图书馆里或从各高校或万方数据库系统获取。
- 政府出版物 (Government Publication): 各国政府部门及其所属机构颁布和出版的文献资料。一个国家的政府出版物的封面往往有其惯用的颜色。美国政府文件、英国下议院文件冠以白皮书, 英国国会文件冠以蓝皮书, 法国政府公布的报告冠以黄皮书, 意大利政府的报告和外交文件、英国政府各部发表的文件冠以绿皮书, 美国外交文书、英国中央统计处的国际收支表冠以红皮书, 而日本则将外务省的外交文书等冠以灰皮书。
- 产品样品资料 (Product Sample Book): 厂商为向用户宣传和推销其产品而印发的介绍产品情况的文献。如: 中国水利水电出版社教材目录与简介。

此外档案 Archives、Records、Music Score、Map 等都是文献家族中的重要成员。

4. 信息、文献、知识三者的关系

三者之间的关系如图 1-1 所示。通过图示可以看出信息>知识>文献。

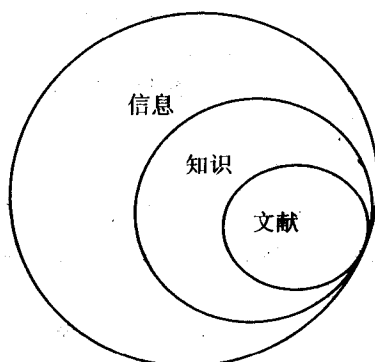


图 1-1 信息、知识和文献三者之间的关系

三者之间的联系与转换：

- 联系：知识是有组织的大量的信息，获得知识有赖于获得信息；信息是知识得以形成和传播的中介，而不是知识本身，知识是经过精心研究、领会后的有用信息，是人类对信息加工处理后的产物；并非所有信息都可成为知识，在信息时代，源于众多复杂客体的大量信息，只有借助于现代化的信息手段，并通过掌握现代信息科学技术的认知主体，才能真正转化为知识。
- 转换（如图 1-2 所示）：①物质运动发出信息；②信息经人脑加工变成知识；③知识被记录形成文献；④文献描述事物。

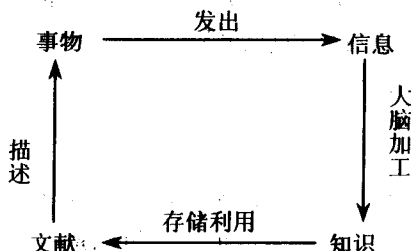


图 1-2 信息、知识、文献之间的转换

如果说搜集信息是人类赖以生存、发展的一种本能，那么信息检索则是每个大学生和科研人员必须具备的一种基本技能。要有效、快速地获取和利用最新信息，就必须掌握信息检索的技能。下面将为大家介绍有关信息检索的基础理论。

1.1.2 信息检索的定义和分类

信息检索是指将杂乱无序的信息有序化形成信息集合，并根据需要从信息集合中查找出特定信息的全过程，全称是信息存储与检索（Information Storage and Retrieval）。信息的存储主要是指对一定范围内的信息进行筛选、描述其特征，加工使之有序化形成信息集合，即建立数据库，这是检索的基础；信息的检索是指采用一定的方法与策略从数据库中查找出所需信息，这是检索的

目的,是存储的反过程。存储与检索是一个相辅相成的过程。为了迅速、准确地检索,就必须了解存储的原理。通常人们所说的信息检索是指后一过程,即信息查找过程,也就是狭义的信息检索(Information Search)。广义的信息检索分为:信息存储(信息检索的基础)和信息检索(信息存储的目的)。

信息检索(Information Retrieval, IR)又称情报检索、文献检索、资讯检索(台湾地区),简单地说就是将信息按照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信息的过程。

信息检索的类型,按检索手段分为:

(1) 手工检索(简称手检):例如以前到图书馆使用卡片目录找书。

(2) 计算机检索(简称机检):又分为脱机信息检索、联机信息检索、光盘信息检索、网络信息检索几个阶段。

按检索内容划分:

(1) 书目检索:是以文献线索为检索内容的信息检索。

(2) 数据检索:是以数据为检索内容的信息检索,要求从检索系统存储的大量原始调查数据和其他统计数据中查出所需的数字资料。

(3) 事实检索:是以具体事项为检索内容的信息检索,要求从检索系统存储的各种原始信息资源中查出专门的事实材料。

(4) 全文检索:即检索系统存储的是整篇文章或整本图书。

(5) 图像检索:即以图形、图像或图文信息为检索内容的信息检索。

(6) 多媒体检索:是以文字、图像、声音等多媒体信息为检索内容的信息检索。

按是否使用检索工具划分:

(1) 直接检索:是指利用一次文献进行检索,这是以前比较常用的一种查找方法。所花时间多、精力大,检出文献少。

(2) 间接检索:是指利用各种检索工具获得文献线索,再根据线索去查找原始文献线索的方法。

按信息检索手段划分:

(1) 传统信息检索:即手工信息检索,是利用各种印刷型检索工具来查找文献的一种方法。

(2) 现代信息检索:即计算机信息检索,是指利用计算机和网络来处理和查找文献信息的检索方式。

1.1.3 信息检索的几个技术指标

检索系统的质量影响着检索系统的使用效果,根据检索课题选择检索系统时需要考虑检索系统的质量。一般说来,可从以下几个角度对检索系统进行评价:

(1) 检索的方便性。检索系统使用起来是否方便可行,决定着检索系统的效用能否充分发挥。检索的方便性是由多方面因素决定的,如信息的编排组织是否科学合理;索引系统是否完备;索引深度是否适中;标识的专指度是否足够等。

(2) 收录的全面性。检索系统的全面性是指它收录的文献信息是否全面。在通常情况下,可以用覆盖面、摘储率、报导数3个指标来衡量。

(3) 检索效率。检索效率是指评价检索系统性能和质量的各种比率,包括查全率、查准率、新颖率、检索速度、检索方便性和成本效益比等。

检索效率主要用查全率和查准率两个指标来衡量。

查全率 (Recall Ratio) 是指从检索系统中检出的有关某课题的文献信息的数量与检索系统中实际与该课题有关的文献信息总量的比率。用公式可表示为:

$$R = \text{检索出的相关文献数量} / \text{系统中相关文献的总数量} \times 100\%$$

查准率 (Precision Ratio) 是指从检索系统中检出的有关某课题的文献信息数量与检索出的文献信息总量的比率。用公式可表示为:

$$P = \text{检索出的相关文献的数量} / \text{检索出的文献的总数量} \times 100\%$$

影响查全率和查准率的两个因素:

- 检索工具的质量
- 检索者的技巧及其对检索工具性能的了解程度

一般来说,查全率越高,查准率越低。

实例:某学生想查找“信息检索”方面的文献,已知在某系统中共存储有 250 篇“信息检索”方面的文献,某次操作共检索出 200 篇该方面的文献,其中符合条件的有 180 篇,求这次检索的查全率、查准率和漏查率。

(4) 报导及时性。检索系统的及时性是指它所报导新出现文献信息的速度,一般用“时差”来衡量,时差是指文献信息从其发表到检索系统提供给用户的这段时间间隔。

总之,收录范围、查全率、查准率、响应时间、用户负担和输出形式 6 项是评价检索效果和检索系统的主要指标。

1.2 信息检索的类型及特点

1.2.1 手工信息检索

是以人工方式查找和提供情报的系统。其特点是人直接参与检索过程。所使用的情报检索工具包括书本式目录、文摘、索引以及各种卡片(穿孔卡片、元词卡片)。手工信息检索具有操作简单、费用低廉、查准率高等优点,但耗时较多、效率低。随着计算机的普及,手工信息已逐渐被计算机信息检索所代替。

1.2.2 脱机信息检索

随着计算机的问世,20 世纪 50 年代相继产生了美国航空航天局的 NASA、美国国立医学图书馆的 MEDLARS 等一系列检索系统。这一阶段的计算机检索是以批处理方式进行的脱机检索,即检索只能在检索系统所在地进行,由检索人员定期将用户课题汇总,批量输入计算机,用户不能参与检索过程,不能即时浏览检索结果、修改检索方案,即不能人机对话。

脱机检索所存在的几点不足表现在:

(1) 地理上的障碍,指用户与检索人员距离较远时,不便于检索要求的表达,也不便于检索结果的获取。

(2) 时间上的迟滞,指检索人员定期检索,用户不能及时获取所需信息。