

现代信息检索技术

许增福 吴春玉 主编

石油工业出版社

内 容 提 要

本书的内容包括信息及其信息检索基本概念和信息检索基础理论,国内外主要手工检索工具的检索及其相关的数据库和网络资源,计算机检索基础知识、信息检索技术、检索策略,联机检索、光盘检索和网络检索方法和过程,以及信息的利用。

本书可作为大专院校学生文献检索课的教材,也可供科研人员及普通信息检索用户学习使用。

图书在版编目(CIP)数据

现代信息检索技术/许增福,吴春玉主编

北京:石油工业出版社,2002.7

ISBN 7-5021-3826-9

I. 现…

II. ①许…②吴…

III. 情报检索

IV. G 252.7

中国版本图书馆 CIP 数据核字:(2002)第 045489 号

石油工业出版社出版

(100011 北京安定门外安华里二区一号楼)

河北省地质局测绘院印刷厂排版印刷

新华书店北京发行所发行

*

850×1168 毫米 32 开本 8.5 印张 228 千字 印 1—3000

2002 年 7 月北京第 1 版 2002 年 7 月河北第 1 次印刷

ISBN 7-5021-3826-9/TP·74

定价:17.00 元

《现代信息检索技术》编委会

主 编：许增福 吴春玉

副主编：金秋颖 韩 颖 李瑞斌

前 言

随着知识经济时代的到来和现代信息技术的迅猛发展,信息已成为人类社会发展的—种驱动力,人们越来越重视对信息资源的有效开发和利用,然而,当今信息的无限增长、庞杂和人们对信息的特定需求之间的矛盾也越来越突出。因此信息素质教育已成为当代大学生素质教育的主要内容之一。

早在 20 世纪 80 年代,教育部在高等院校中大力推广的文献检索与利用课,对培养大学生的信息意识、获取信息的能力、自学能力方面发挥了很好的作用,受到了高校的重视和大学学生的普遍欢迎。

我们在多年开设文献检索课的基础上组织人力编写该书,目的在于向高校的本、专科生和研究生提供一个信息检索与利用的理论和实际相结合的入门性书籍,使学生通过学习本教材,能够掌握现代信息资源检索的基础知识、基本原理和实际检索的基本技能,为提高学生的信息素养尽—份力量。

随着信息技术的发展,检索手段在不断更新,尤其是 Internet 上的检索工具更是名目繁多,花样迭出,因此提醒读者在使用本书的过程中,注意浏览和借鉴其他资料,以求全面、系统地掌握这一领域的知识。

本书的主要内容有信息及信息检索基础知识、手工检索工具、数据库和网络信息资源检索以及文献信息的利用。

本书编写分工如下:第—、四、五、八章由许增福编写;第二章第—、二、四、五节,第三章第二~四节由吴春玉编写;第

六、七章由金秋颖编写；第九章由韩颖编写；第二章第三节、第三章第一节由李瑞斌编写。全书的整体构思、结构设计、修改和定稿由许增福、吴春玉完成。

本书虽经著者努力，但错误、缺点在所难免，希望广大读者批评指正。

编 者

2002 年 6 月

目 录

第一章 绪 论	(1)
第一节 信息及其相关知识	(1)
第二节 知识创新、信息意识、信息素质	(5)
第二章 信息检索基础知识	(9)
第一节 信息检索概述	(9)
第二节 信息源及其种类	(17)
第三节 信息检索工具	(29)
第四节 信息检索语言	(36)
第五节 信息检索程序、策略及效果评价	(52)
第三章 计算机信息检索概述	(57)
第一节 计算机信息检索发展过程	(57)
第二节 计算机信息检索技术	(58)
第三节 数据库及其种类	(65)
第四节 计算机检索策略	(72)
第四章 综合性科技信息检索	(74)
第一节 科技知识与事实数据的检索	(74)
第二节 中文科技信息检索	(76)
第三节 外文综合性科技信息检索工具	(83)
第四节 外文综合性科技信息数据库	(99)
第五章 专业性科技信息检索	(106)
第一节 专业性科技信息检索工具	(106)
第二节 专业性科技信息数据库	(126)
第六章 特种文献检索	(132)
第一节 专利文献及其检索	(132)
第二节 会议文献及其检索	(140)

第三节	科技报告及其检索·····	(144)
第四节	标准文献及其检索·····	(158)
第五节	学位论文及其检索·····	(151)
第七章	联机检索及光盘检索·····	(157)
第一节	联机检索系统·····	(157)
第二节	光盘信息检索·····	(171)
第八章	Internet 信息检索 ·····	(195)
第一节	Internet 基础知识 ·····	(195)
第二节	网络信息检索概述·····	(197)
第三节	中外著名搜索引擎的使用·····	(204)
第九章	文献信息的利用·····	(211)
第一节	文献信息的收集、整理和分析·····	(211)
第二节	科技文献写作的意义及步骤·····	(221)
第三节	几种重要科技文献及其写作方法·····	(225)
附录	·····	(241)
参考文献	·····	(264)

第一章 绪 论

第一节 信息及其相关知识

一、信息的涵义和种类

1. 信息的涵义

当今社会是信息社会，信息无处不存在，无时不有，无人不用，无孔不入。随着人类的进步，信息悄悄地渗透到人们的衣食住行，信息的利用程度关系到人们的生活质量，并且其影响程度呈上升趋势。提起信息似乎人人皆知，然而，学术界还没有对信息统一的定义。信息的英文词汇是 Information，目前对这一词汇的解释有 100 余种之多。不同领域，从不同角度对信息这个概念有不同的解释。

哲学家从产生信息的客体来定义信息，认为能被其他事物感知的、表征该事物特征的信号内容即该事物向其他事物传递的信号。在新闻界普遍认为，信息是对事物运动状态的陈述，是物与物、物与人、人与人之间的特征传输。心理学者认为，信息是存在于意识之外，它存在于自然界、印刷品、硬盘以及空气之中。而经济学家则认为，信息是与物质、能量相并列的客观世界的三大要素之一，是为管理和决策提供依据的有效数据。

信息论的创始人申农 (C.E.Shannon) 从通信系统理论的角度把信息定义为“用来减少随机不确定性的东西”。他主要研究信息的本质，用数学的方法研究信息的计量、传递、变化、储存，其核心是信息熵的概念。信息熵就是衡量对某种事物认识的不确定度，即信息的有序度。申农从这一思想出发推导出了信息测度数学公式，为信息认识做出了重大的贡献，也标志着信息科

学迈入了定量研究阶段。控制论专家 N. 维纳对信息的定义是“信息就是我们适合外部世界进行交换的内容的名称”。他的信息概念是从信息发送、传输、接收的过程中客体与认识主体之间相互作用的角度来定义的。

信息科学是一门新兴学科，它的许多分支学科仍在随着社会、经济和科学技术的发展而发展，其内涵和外延不很确切；人们研究的角度和使用目的不同，出现了以上诸多的解释。但国内情报领域比较认同的看法是信息是事物存在的方式和运动状态的表征，信息并非事物本身，而是事物发出的体现它存在和运动状态的信号和消息。

信息普遍存在于自然界、社会各界和思维界。因此，信息可以理解为通过信号传来的消息。信息是被传递的数据、事实、符号的集合，客观事物存在反映的物质质量。

信息具有传递性、存储性、可识别性、并消除不确定性、替代性和可共享性等特性。

2. 信息的种类

从不同角度，信息可划分出很多种类。

1) 按信息产生的客体性质来划分，可分为自然信息和社会信息。

自然信息是指宇宙间自然界客观存在或随机发生的各种生命信息、动植物界信息、物质物理信息等。

社会信息是指人类和社会维系生存、生产和发展过程中所产生、传递和利用的信息。这一类信息就是人们平常所说的信息。

2) 按信息的社会属性划分，可分为政治信息、军事信息、经济信息、科技信息、管理信息、生活信息等。

3) 按信息的传递范围划分，可分为公开信息、半公开信息和非公开信息。

4) 按信息的加工程度划分，可分为一次信息、二次信息、三次信息。

5) 按价值观念划分，可分为有价值信息和无价值信息。

6) 按运动状态划分, 可分为动态信息和静态信息。

动态信息是指像新闻、气象等具有较强的实效性、随时都在变化着的信息。而静态信息是指历史的、不再变化的信息, 比如: 各类文献、历史资料等。

7) 按载体形式划分, 可分为文献信息、口头信息、电子信息。

文献信息是指文献所表达的内在信息, 是以文字、符号、声音、图像为编码, 记录在各种载体上的信息。口头信息是指存在于人脑记忆中的、通过交谈、讨论、报告等方式交流传播的信息。电子信息是近几年随着计算机技术、通信技术、多媒体技术和高密度存储技术迅速发展为基础而发展起来的一种新兴的信息传播形式, 它将文字、声音、图像融为一体, 提供给人们更加完善的信息。

二、知识、情报、文献、资讯及其信息的关系

1. 知识及其信息的关系

《现代汉语词典》对知识的解释是: “知识是人们在改造世界的实践中所获得的认识和经验的总和。”在《辞海》中把知识定义为: “人们在实践中积累起来的经验, 从本质上说, 知识属认识范畴。”这两种解释均把知识作为实践或实践活动的认知成果来看待, 这是正确的。然而, 知识的这种定义不足之处在于未能揭示认知活动中的主体与客体的动态关系, 而这种关系正是知识的本质。

1996 年国际经合组织 (OECD) 发表的《以知识为基础的经济》的报告中系统地提出了知识的 4 个 W 概念: ①知道是什么 (Know-what): 指关于事实方面的知识; ②知道为什么 (Know-why): 指原理和规律方面的知识; ③知道怎么做 (Know-how): 指操作的能力, 包括技术、技能、技巧和诀窍等等; ④知道是谁 (Know-who): 包括了特定关系的形成, 以便可能接触有关专家并有效地利用他们的知识, 也就是关于管理的知识和能力。

世界经合组织的报告把第一、二类知识称为编码知识即显性知识，把第三、四类知识称为意会知识也叫隐性知识。所谓意会知识就是不显露的、难于用语言表述的知识，它是不脱离认知主体的。

意会知识的获得主要靠“实践”。意会知识包括认知、情感、信仰、经验、技能等，一般来说，这些知识是不能通过正式交流与传播渠道转移的，必须通过行动来表现出来。而编码知识是指以文字、图像、符号来表述，以印刷或电子方式记载，可供人们交流的有形的、结构化的知识，编码知识比较容易获得、理解和交流，它具有公共性，可以存储在图书馆、局域网或数据库中传播和获取，成本低。

知识与信息密切相关，人们为了进行知识的传递和交流，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、图像和实物等加以表现。这种表现形式就是信息。因此，知识与信息具有内在联系与较强的互动性，但两者实质上是内在与形式的关系，是两个不同的概念，不存在简单的包含关系或交叉关系，但一定条件下可以相互转化。

2. 情报及其信息的关系

情报在《辞海》中解释为：①以侦探手段或其他方法获得的有关敌人军事、政治、经济等各方面的情况，以及对这些情况进行分析研究的成果，是军事行动的重要依据之一。②泛指一切最新的情况报道。在人们的印象中，情报一词带有军事行为，具有保密性的含义。随着时代的进步，情报这一词已无法适应时代的发展。1992年国家科委决定将科技情报改为科技信息，许多情报所已改为信息所，这样更具有时代的特点。

3. 文献及其信息的关系

随着生产力的发展，出现了各种不同的存储和表达人们思想的载体材料，发明了各种各样记录知识信息的方式，文献大量涌现，文献概念的外延也在不断扩大。因此，不能把文献仅仅看作我们习惯的纸质形式的载体。目前有关文献的较权威的定义主要

有两种：一是《文献情报术语国际标准（草案）》（ISO/DIS5127）的定义，“为了把人类知识传播开来和继承下去，人们用文字、图形、符号、声频、视频等手段将其记录下来，或写在纸上，或晒在蓝图上，或摄制在感光片上，或录到唱片上，或存储在磁盘上。这种附着在各种载体上的记录统称为文献”；另一个定义为，“文献是记录有知识的一切载体”（GB 3792.1—83，“文献著录总则”）。我们在这里文献可以理解为固化信息的载体。因此，信息与文献的关系是内含与载体的关系。

4. 资讯

资讯是海外学者对 Information 的译名，与信息有相同的含义。

第二节 知识创新、信息意识、信息素质

一、知识创新与信息意识

美国学者 D.M.Amidor（1993）提出：知识创新是指科学家和工程师进行跨学科、跨行业、跨国家合作，研究共同感兴趣的问题，其研究结果加速了新思想创造、流动和应用，加速了这些新思想应用于产品和服务，以造福于人类社会的过程。狭义的知识创新指通过学习研究获得和创造新知识的过程，它发生在知识生产、传播应用的全过程；广义的知识创新指为了经济和社会利益创造、传播和利用新知识，使其转变成市场化的产品或服务，包括科研获得的新知识、新知识的传播和应用、新知识的商业化等。

知识创新需要知识积累为前提，我们首先阅读大量的科学文献，掌握有关的思想、事实、理论和方法等信息，在此基础上进一步分析、综合和研究，才能有所创新。在知识积累的过程中人们经常感到由于知识的存储过于庞大和无序，堵塞了通向知识之门的道路，耗费了大量的精力，因此人们需要认真学习和研究获取信息的方法，掌握大量无序知识中搜索有用的、准确的、全面

的知识的技能。能否进行知识和技术的创新，与有没有能力获取创新所需要的信息有关。因此，信息检索知识已经构成知识体系中不可缺少的一部分。

信息意识是信息在人脑中的集中反映，即社会成员在信息活动中产生的认识、观点和理论的总和，是人们凭借对信息与信息价值所特有的敏感性和亲和力，主动利用现代信息技术捕捉、判断、整理、利用信息的意识。同样的信息，有的人善于抓住，有的人却默然视之。这是由于个人的信息意识强弱不同造成的。有无信息意识决定着人们捕捉、判断和利用信息的自觉程度，而信息意识的强弱对能否挖掘出有价值的信息、对文献获取能力的提高起着关键的作用。信息意识分感性阶段信息意识和理性阶段信息意识。感性意识是指主体由于某种随机的、临时的信息需求而接触和感知文献信息的外部世界及表面特征，是大脑类似于感性认识的反映，是信息意识的低级阶段。理性意识是指主体对一系列感性意识的抽象和综合过程，是综合了许多文献信息的感性材料，经过创造性思维，去粗取精，去伪存真，从中抽取文献信息的共同本质及其运动的规律，是信息意识的高级阶段。感性意识与理性意识是统一的、不可分割的两个阶段，理性意识依赖于感性意识，没有丰富的感性意识就不可能发展到正确的理性意识；而感性意识又有待于发展到理性意识，只有发展到理性意识才能全面地反映信息的内部关系与本质。

信息意识从信息认知、信息情感和信息行为倾向三个层面体现出来，信息认知是指对信息和信息活动的了解和看法，其中最重要的是评价性的认知；信息情感是指人们在感受信息的过程中，逐渐形成的反映需求关系的内心体验，这种体验相对持久、稳定，而不是即时产生的情绪；信息行为倾向是指个人在信息活动中表现出来的行为趋势，是信息行为的心理准备状态。人们的信息搜集活动是受信息需求驱使的，而影响需求的力量大小主要就是需求被意识的清晰程度，即意识越明确，行动目标越清楚，则信息活动的动机就越稳定、持久、强烈，努力程度也就越高。

如果这种意识仅停留在感性阶段，那么接受信息总是处于被动状态；反之，信息意识经常在觉醒的、活跃的状态，就会促使人们主动制定信息活动计划，预见到各种变化，并做出积极的选择。

二、信息素质

培养信息意识的重要途径是重视信息素质教育。

信息素质（Information Literacy）一词最早是由美国信息产业协会主席 Paul Zurkowski 在 1974 年给美国政府报告中提出来的。他认为信息素质是人们在工作中运用信息、学习信息技术、利用信息解决问题的能力。目前有关信息素质的定义较多。美国图书馆协会认为：信息素质是人们知道什么时候需要信息并找到、评价及有效地利用所需信息的能力。信息素质能力较强的人知道如何学习，因为他们了解知识是怎样组织的，知道如何找到信息，他们能够终生学习，因为他们能够发现所有与自己职责相关的或决策所需的信息。美国大学与研究图书馆协会认为：信息素质是一系列有关个人能意识到信息需要并能找到、评价和有效利用所需信息的能力。从狭义上看来，信息素质包括各种有效地使用信息技术和信息资源的技能；从广义上看，信息素质是一种自由的艺术，它包括了社会、文化和哲学等内容。信息素质可分为：①工具素质：能使用印刷和电子信息资源的有关工具，如计算机、软件等；②资源素质：了解信息资源的种类、形式和查找、检索方法；③社会结构素质：掌握信息的社会地位、作用、影响；④研究素质：掌握利用信息进行有关学科研究的技能；⑤出版和传播信息素质：能利用文本或多媒体报告研究结果。

总之，信息素质是指从各种信息源中检索、评价和使用信息的能力，是信息社会劳动者必须掌握的终身技能。信息素质的内涵具体包括能意识到准确和完整的信息是决策的基础；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；能评价信息；能根据实际用途组织信息；使用信息，将新信息融会到现有知识结构中。

生活在信息社会的人们，尤其是当代大学生应掌握一些图书

馆学、情报学的理论知识和方法，对于了解和掌握信息的组织、检索和分析是非常必要的。随着网络的普及，网上信息素质教育资源日渐丰富。如：北京大学图书馆（<http://www.lib.pku.edu.cn>）、清华大学图书馆（<http://www.lib.tsinghua.edu.cn>）、上海交通大学图书馆（<http://www.lib.stju.edu.cn>）、华中理工大学图书馆（<http://www.lib.hust.edu.cn/index.asp>）、中国国家图书馆（<http://www.nlc.gov.cn>）等各大学图书馆及一些公共图书馆的主页下，介绍了各种信息资源及其利用方法、用多媒体形式介绍了文献检索方法等，内容从简单的“怎样利用图书馆”到介绍国内外各大数据库及网上信息的获取方法，应有尽有。

第二章 信息检索基础知识

第一节 信息检索概述

一、信息检索

1. 信息检索的含义

信息检索是指将信息按一定的方式进行加工、整理、组织并存储起来,再根据信息用户的需要找出有关信息的过程。它的全过程又叫信息存储与检索。这是广义的信息检索的含义,主要是对信息工作者而言的。狭义的信息检索则仅指后半部分,即用户根据需要,借助检索工具,从信息集合中找出所需要的信息的过程。

2. 信息检索的意义

信息检索是查找信息的方法和手段,它能使人们在浩如烟海的信息海洋中迅速地、准确地、全面地查找所需的信息。可以说信息检索对人们的学习、生活和工作等方面都有非常大的作用。

(1) 信息检索在科学研究中的作用

人类的科学技术发展与进步都是在前人已有的知识和成果之上产生的,正如著名科学家牛顿所说的那样:“如果我发现的,我所发现的要比笛卡尔和培根远大一点的话,那就是因为我站在巨人的肩膀上的缘故。”科学研究的每一项成果都是在总结归纳了大量信息的基础上完成的,因此信息检索对科技人员来说是十分重要的。具体表现在以下几方面:

1) 充分地开发和利用信息资源。

2) 节省科研人员的时间,提高效率,减少人力、物力等方面的费用。

- 3) 防止自身知识老化, 再学习的工具。
 - 4) 继承和借鉴前人的研究成果, 避免重复研究和少走弯路。
- (2) 信息检索在生活中的作用

有人认为: “我既不搞科研, 也不搞管理, 信息检索可能与我无关。” 在过去可能是这样, 但由于网络的出现, 改变了我们的生活。现在有些发达国家和地区已实现了衣食住行等诸多问题都可以在信息网络上解决, 因此离开了信息检索将很难生活。

二、信息检索原理及手段

1. 信息检索原理

人类的信息检索行为总是从特定的信息需求开始, 并在特定环境和信息检索系统中完成, 这里所说的环境包括产生需求的环境、信息检索系统的运行环境和其他制约因素。特定的检索系统包括完成检索过程所需的一定设施和工具, 它可以是图书馆、信息中心或信息经济人, 也可以是某种工具书(如: 文摘索引、目录、资料集、手册、词典等)或机读信息源(如: 各种机读数据库)。

人类的信息需求千差万别, 获取信息的方法也各种各样, 但信息检索的基本原理却是相同的, 可以把它最本质的部分概括为一句话: 对信息集合与需求集合的匹配与选择(见图 2-1)。

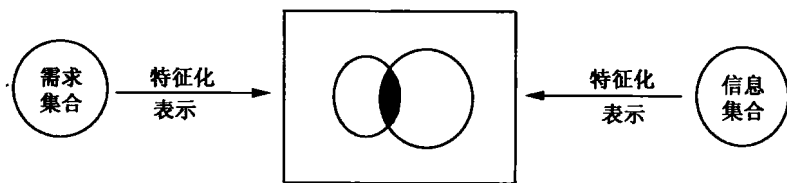


图 2-1 信息检索基本原理示意图

2. 信息检索手段

根据信息检索的基本原理, 实现信息检索的基本方式可分为传统信息检索和现代信息检索。传统信息检索简称“手检”, 现