

INFORMATION RETRIEVAL

信息检索

第4版

东南大学出版社

© 张厚生 主编

信 息 检 索

(第 4 版)

张厚生 主编

东南大学出版社

内 容 提 要

本书为“信息素养”课程提供教材,在阐述信息检索基本理论和方法的基础上,对信息检索领域最新进展作出介绍和分析,以 Web 信息检索、数据库检索内容为侧重,兼容对以印刷型文献为载体的知识信息检索的研究和讲授,并对数字化资源、网络资源及其检索工具、多媒体信息、全文检索、网上跟踪、搜索引擎和信息导航技术使用等分别进行了剖析,具有较强的科学性、系统性和实用性。

本书着眼于指导用户开展信息检索实践,适合作为大学生、研究生的教科书,也可作为科技人员、管理业者以及一切信息工作相关的读者大众的参考书。

图书在版编目(CIP)数据

信息检索 / 张厚生主编. —4 版. —南京:东南大学出版社, 2006. 8

ISBN 7-5641-0505-4

I. 信... II. 张... 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(2006)第 084734 号

东南大学出版社出版发行
(南京四牌楼 2 号 邮编 210096)

出版人:宋增民

江苏省新华书店经销 南京京新印刷厂印刷

开本:700mm×1000mm 1/16 印张:28 字数:565 千字

2006 年 8 月第 4 版 2006 年 8 月第 1 次印刷

印数:1~6 000 定价:33.00 元

(凡因印装质量问题,可直接向读者服务部调换。电话:025-83792328)

序

《信息检索》一书第1版(题名《情报检索》)自1987年1月问世以来,深受我国高等院校的本科生、研究生、青年教师以及图书馆、信息研究与产业机构的同行们和广大读者的欢迎。先后6次印刷,发行量超过6万册。1997年本书出第2版,2002年出第3版,先后数次重印,发行量达到10万册。我国自改革开放以来,随着社会主义建设事业的全面推进、高等院校培养适应信息化建设人才的目标的实施、科学研究向经济建设主战场的倾斜,人们对信息检索的需求与日俱增。这部书稿有助于人们更好地了解信息世界的发展动态,研究学术进展,指出读书门径,解答疑难问题,提供资料线索;有利于强化青年读者的信息和知识创新意识,提高科学研究的能力,并能使他们在浩如烟海的信息中,迅速而又准确地查找到自己所需要的知识与信息。

该书第1版于1989年10月获中国图书馆学会颁发的“中华人民共和国成立40周年、中国图书馆学会成立10周年图书馆学、情报学优秀著作奖”,又获“东南大学1993年度优秀教材奖”。1989年和1992年夏季,由国家教委发文委托东南大学张厚生同志牵头主办“全国高校‘文献检索与利用’暑期师资讲习班”,讲习班使用本书第1版作为教材,受到来自各地高校同行的好评。“八五”期间,以该书为教材的信息检索课程获江苏省教育委员会颁发的“普通高等学校教学质量奖”一等奖;第2版于2000年获“江苏省政府三等奖”。

随着时间的推移,由于科学技术的突飞猛进,新的检索内容、检索手段层出不穷。为了适应变化了的情况,他们对本书进行了大量的修订工作。如对Internet、多媒体、光盘检索、科技查新等都一一作了专门的介绍。从第1版到第2版,前后用了10年;从第2版到第3版出书,经过5年时间,从第3版到本次的第4版出书,经过了4年,加快了适应信息新技术的发展变化,适时地修订补充了新的内容,这在当前同类教科书中尚属首例。当然,此次修订不仅内容更新,而且适用面更宽,适合各种类型读者的阅读与使用。应该说,对一部书稿的增补、修订、再版,是使图书内容保持观点新颖、资料翔实、永葆高质量的必要措施。这充分说明了作者者的责任意识、大局意识与精品意识,是值得充分肯定的。

借此机会,特别感谢本书的主编——张厚生教授及其合作者们。张厚生同志长期以来,一直从事图书馆、信息服务和科学研究工作,并在管理、教学及图书馆信

息化方面颇有建树。1983年起他在东南大学策划、启动图书馆计算机管理目标。1984年又在学校支持下,建立了 DILOG 国际联机检索终端。1986年组织情报技术部研制的计算机模拟教学系统通过江苏省科委鉴定。1987年他参与的该校计算机中心同图书馆联合开发的项目“多用户计算机流通、检索管理系统”得到推行和使用,为该校图书馆管理手段的更新做了有益的尝试。“七五”以来,他先后牵头、担负省级和国家级科学研究立项项目多个,如江苏省“七五”规划项目“江苏地区文献资源建设、开发与利用”、国家社会科学“九五”规划项目“信息新技术在图书情报工作中的应用与评价的研究”、江苏省“九五”项目“信息服务与信息用户培养研究”、全国高等教育研究会立项项目“面向 21 世纪的教材建设的研究”、江苏省“十五”规划项目“信息资源数字化、网络化发展模式与管理模式研究”等。他主编及参加编著并已出版的专著有 10 多种。为此,本书仍然由他来主编与修订是值得信赖的。参加本书编写的其他成员是东南大学情报科学技术研究所和国家教育部科技查新中心南京工作站的一支有较高水平的信息管理与研究方面的中青年队伍。《信息检索》一书也是他们长期从事信息检索实践与教学研究的成果结晶,相信一定会受到广大读者的欢迎。

倪 波
于南京大学

注:倪波,教授,南京大学博士生导师。

初版钱序

近三十多年来,我国图书馆和情报事业有了很大发展,在图书馆学的研究方面成绩显著,培养图书馆学情报学专门人才的教育事业也有较大发展。据不完全统计,目前我国已有四十九所高等院校设置了图书情报学系或专修科;此外,夜大、电大和大学函授部也开设了图书馆学专业。但是把我国图书情报教育的一些统计数字,与美国和苏联等国的相比,还存在颇大差距,深感我国应奋起直追。

美国人口只有我国的四分之一,却有图书情报院、系 347 所,每年毕业的图书情报学士、研究生达 7 000 人左右;苏联人口只有我国的五分之一,至 1980 年就已设有图书情报学院、系 28 所,在校生达 33 000 人,每年毕业生也有 7 000 余人;日本人口只有我国的十分之一,每年图书情报专业毕业生竟达八九千人;印度人口是我国的一半,每年从设有图书情报专业的院校毕业的也有 2 000 多人;泰国人口是我国人口的二十分之一,现有图书情报系 7 个,在校本科生达 500 人。我国的台湾省人口仅占全国人口的百分之一,也有五所院校设有图书情报学系,在校学生 1 500 人。由此可知,国内图书情报专业教育显得非常落后。为适应“四化”和“四有”的建设,大力发展我国图书情报教育,实在已是迫在眉睫和当务之急的事了。南京工学院的领导和教师有鉴及此,于 1984 年 8 月创设了图书情报专业,《情报检索》就是该专业中至关重要的一门必修课。

☆

☆

本教材是配合国家教委规定的“文献检索与利用”课程教学的理工科的通用书,也是为工程技术人员所必备的参考书之一。本教材内容共有九章,第一、二章是情报检索的基本原理、方法与步骤,都属有关理论方面的材料;其后各章是各类型的情报检索刊物的介绍,主要对世界主要而著名的手工检索工具和电子计算机情报检索系统作简明扼要的介绍。

我国科技情报这一专业是从先进国家引进的,且已历有年所。但一般所编教材或写作论文,几乎都偏重于工作实践和工作方法,正如我国引进近代图书馆学一样,大都轻视理论基础。其实理论基础既是实践的结晶,又是实践的指针。如果对一门学科归纳不出一整套扎实的理论,就意味着尚未能把所有的实践上升成为理论。这种缺乏理论指导的实践,难免局限于经验甚至成为盲目的实践。长达八十多年来,我国图书馆学的基础理论何以未能建立完善,这是发人深思的一个大问题。

情报学与图书馆学有着密切关系,虽然,各自有其不同的范围、对象和方法,但是都必须重视理论,不要重蹈覆辙,这是历史的教训。因此本教材,对情报检索方

面的理论给予相当重视,虽然不能说异军突起,至少可以说是一良好的开端,有助于铺平进展的道路,对于《情报检索》这门课程的开拓具有一定贡献。

本教材具有以下四个特点:

一、适应性强。关于情报检索各科的教材,几年来虽已编撰不少,例如《医学文献检索与利用》、《农业文献检索与利用》等,但所讲授的内容多有一定的局限性。本教材取材比较广泛,具有通用性,适用面广。因此,本教材不仅可作理工科各专业的《情报检索》课教材,也可作其它有关各类学科以及从事情报工作人员的参考或指导书。

二、理论紧密联系实际。本教材内容结合情报检索语言介绍了典型检索工具的使用,从而使理论与方法密切相结合,有利于解决检索中的共性问题。与某些单列章节谈检索语言的书刊相比,不致感到枯燥、单调、乏味,更适合于教学。

三、内容新实。在选择教材内容上,不仅注重了扎实的基础,而且蒐集了不少新颖和实用的材料,例如在“事实、数据……”一章中,主要以词典、百科全书、年鉴、名录、指南等普通工具书为对象,使之成为从事情报检索的津梁,而《情报检索》一书才名实相符。又如在“专利文献”、“标准文献”、“科技报告”、“学位论文”、“会议文献”等有关章节中的内容,有利于理工科学生及广大工程技术人员的创造发明。

四、利于教学。本教材的初稿作者已经历了多次教学实践。据其学员一致反映,通过这一课程的学习,大开眼界,掌握了不少世界有名的主要检索刊物,通过学习,能初步解决若干咨询问题,这对于师生的教与学都大有裨益。本教材是经多次修改增补而成,内容更为翔实丰富,教学效果定能更上一层楼。

因此,具有这样四个特点的《情报检索》一书,可说是目前同类书中较为先进的,值得郑重推荐。

☆

☆

这本《情报检索》教材得以正式出版,对于广大师生和情报工作人员是一件十分可喜的事。但也确是来之不易,这是近几年来通过作者教学实践与探索,经历了不少考验获得优异成绩而达到的。当然在科学技术一日千里的今天,要使一本教材的内容不断更新而能适合时代前进的步伐,更非轻而易举之事。因此,本人殷切期望本书作者能够群策群力,不断关注有关的新问题、新材料、新方法和新动向,进行蒐集、整理、探讨、研究,而在二三年后再行修订版本。本人深信本书作者定能一如既往,不断努力,以自强不息的精神,使之再版、三版……在《情报检索》这一领域里,在培养我国情报检索人员的教学事业中,永放光彩。

钱亚新 于南京 1986年8月

注:钱亚新(1903~1990),研究员,著名图书馆学家、目录学家。曾任中国图书馆学会名誉理事,江苏图书馆学会名誉理事,江苏图书馆学报主编,南京大学、安徽大学图书馆学系兼职教授。

第4版前言

本书面向知识信息用户需求,为帮助读者提高信息素养,在增强知识创新中对信息处理、检索、分析、组织和利用的技能而编写。

信息检索、分析、组织和利用的能力,即为当前倍受国内外所关注的“信息素养”(Information literacy)。由美国大学与研究型图书馆协会指导委员会于2000年1月18日批准的《高等教育信息素养能力标准》认为,一个具备信息素养的人应达到的目标是:确定所需信息的范围;有效地获取所需的信息;鉴别信息及其来源;将检索出的信息融入自己的知识基础;有效地利用信息去完成一个具体任务;了解利用信息所涉及的经济、法律和社会问题,合理合法地检索、搜集和利用信息。

信息检索能力的培养为终身学习打下基础。

当前,对信息检索理论和方法的研究已成为国内外学者关注的视点。“开展信息素质教育,培养读者的信息意识和获取、利用文献信息的能力。”这是我国《普通高等学校图书馆规程(修订)》所规定的高校图书馆的任务之一。信息检索能力被认为是信息素养的最主要的组成要素。2001~2005年世界各国发表并被SCI收录的关于信息检索的理论与方法的研究方面的论文共2264篇。其中有Fidel R等的A multidimensional approach to the study of human-information interaction; A case study of collaborative information retrieval等,这些研究和Ricardo Baeza-Yates, Berthier Bibeiro-Nero著的《现代信息检索》等成果颇受学术界的关注。通过对维普中国科技期刊数据库的检索:2001~2005年国内关于信息检索理论和方法的论文共228篇。综观国内外的信息检索理论和方法的研究,主要关注的是检索原理、检索策略、网络信息检索技术、多媒体信息检索、智能信息检索等方面,但这些研究还不够系统、不够深入。

《信息检索》(Information Retrieval)在1987年1月初版时曾题名《情报检索》。1997年7月本书用现题名出版修订本(第2版),2002年5月东南大学百年校庆之际出第3版。自本书问世以来,深得社会各界读者的关注,并作为东南大学和多家兄弟院校理工科专业本科生、研究生的专用教材或专用参考书。仅东南大学每年数千名的毕业生,就几乎人手一册,作为工具书伴学子们迈出校门,走上新的征程。

随着社会的进步、科学的发展,信息新技术叠出,新的检索内容、检索手段不断产生,传统的媒体和检索工具、检索方式也在不断发生变化,人们的信息意识、信息观念在不断更新,经济技术和科学知识的创新能力在不断增强。随着Internet和网络技术的发展,数字化信息资源的涌现,网络环境给我们展现了一个五彩缤纷的

信息世界。更大数量级的信息,更多信息类型和表现形式,以及更复杂的时空关联关系,海量的网上信息资源并发增长,既为信息的开发与利用提供了便利条件,也为信息的发布与分享提供了外部环境。然而,信息产生和流动的随机性、信息时空关系和系统状态的不确定性,导致查找和使用上的困难;数字化、网络化信息分散、无序、动态变化等以及信息的庞杂同特定需求之间的矛盾,也给人们搜集与利用信息的运作增加了困难和不便。本书正是针对这些问题,增加了数据库查询、网络信息搜索和网络导航模式的内容。为适应用户获取信息的习惯和信息载体的转变,信息检索理论和方法也势必发生转变,这就需要研究并提供新的信息处理模式、检索模型和检索方法。搜索引擎是在浩瀚的 Web 中搜索各种有用信息最主要的一种检索工具,但往往忽略了信息的内容,且比较少地对信息资源做预处理,使得搜索引擎仍然存在不少的局限性,比如信息丢失、返回信息太多、信息无关等。这些问题有待于我们进一步研究解决。由此,还迫切需要研究提升对信息检索理论与方法的认识水平。相关研究应从多角度展开:从信息来源看,主要包括搜索引擎与结构化和半结构化数据库的整合以及对异构、跨语言和多媒体数据库等方面的研究;从信息获取看,主要包括检索语言、智能代理、检索模型和算法、域本体和语义、并行和分布式信息检索及信息可视化等方面的研究。以往的研究虽然对检索模型有着某些改进,但事实上仍然没有实质性的突破,这主要是由于以用户需求为中心的问题尚未作深入的研究,而这正是提高信息检索效率及准确度的关键。另外,有关针对中文信息检索的论述也给予关注,本书面对这些问题的解决作了分析和解答。

为了使信息检索向知识检索转变,让知识信息检索由经验走向科学,适应以印刷型文本线性检索向电子全文检索、数值检索、多媒体信息挖掘方向的变化和发展以及为了适应当前大学教科书内容不断更新的要求,自 2005 年 5 月,我们组织有关信息检索与利用课程的现任教师,在联系近几年教学实践和对第 3 版书稿进行全面审视、考量的基础上,再次修订、出版《信息检索》一书。面对社会信息化的趋势,为适应知识管理、知识创新中对信息检索能力的培养要求,本书采取兼容的态度,融合书本式和数字化、网络化信息检索方法与技能的研究成果于一炉。《信息检索》第 4 版的构架、写作目的、写作要求等由主编者策划并主持实施修订和编写。其修订的内容主要有:

(1) 将原书的第 9 章和第 10 章重新修订写作:第 9 章内容为基于文本的多媒体信息检索、基于内容的多媒体信息检索,对图像、音频、视频的信息检索的实现作了讲授;第 10 章讲述网络信息检索,中外文信息搜索引擎的方法,Google、百度、搜狗、天网等常用搜索引擎的使用,网上跟踪技术,网络导航和信息查询技巧。

(2) 重编第 7 章的“医学信息资源及其检索”、“国内外常用生物信息学服务器”,研究、介绍《中国生物医学文献数据库》单机版和网络版的检索, MEDLINE 光

盘数据库、互联网上的 PubMed 检索,其它如《化学文摘》、《数学评论》等都在分析了近几年新的变化后作出了新的阐述,借以充实信息检索基础和专业学科信息基本知识的内容。

(3) 对重要的检索工具,如美国《工程索引》、《科学引文索引》、《化学文摘》、英国《科学文摘》、英国德温特出版公司的“德温特世界专利检索体系”、国际标准化组织的 ISO 目录等著名的检索体系,首先着力于分析、介绍网络资源的检索技能;对于电子资源检索、纸印本查询步骤和方法也分别作了介绍;对于中文信息,增编了网络中文信息的搜索和中文数据库检索作为第4章列入本书正文。

(4) 第1章~第3章、第5章~第8章等各章内容,均按2002年本书第3版出版以后至2006年间的国内外近4年来有关信息检索发展和变化情况对原书或作增删,或作调整,或重新写作。

本书由张厚生主编。参加编写的有8位责任者:其中第1章~第3章由张厚生修订、编写;第4章由张厚生、陆美编写;第5章由陆美、张厚生、袁曦临修订、编写;第6章由陆美修订、编写;第7章由顾建新、朱佳鸣、崔冰洋、胡彬修订、编写;第8章由陆美修订、编写;第9章~第10章由钱鹏负责修订、编写。全书由张厚生负责加工、统编、定稿。在本书修订、再版过程中,得到东南大学图书馆李爱国、张明果,东南大学情报科学技术研究所王春梅、魏彬、项珍以及南京大学袁勤俭博士、杨海平博士的关心和支援;本书汲取了国内外许多专家学者的有关研究成果,同时,承蒙南京大学博士生导师倪波教授为本书作序,在此一并致谢。

限于水准和认知,书中可能存在舛误、疏漏或失当之处,尚祈学者时贤和识者先进不吝教正。谢谢诸位的阅读分享!

张厚生

2006年6月6日于东南大学

目 次

1 概 论	1
1.1 信息·知识·文献	1
1.1.1 信息	1
1.1.2 知识	2
1.1.3 文献	3
1.1.4 信息交流与文献	3
1.2 信息检索原理	5
1.2.1 信息来源	5
1.2.2 文献型信息源	5
1.2.3 零次信息源	10
1.2.4 检索工具	11
1.2.5 信息检索方法	14
1.2.6 文献特征	15
1.2.7 信息检索的步骤	16
1.2.8 检索工具中文献来源的识辨	19
1.3 知识内容检索与信息检索、科技查新	22
1.3.1 知识内容检索	22
1.3.2 经济建设和科学研究中的信息活动	23
1.3.3 科技查新工作	25
1.4 信息检索用户教育	29
1.4.1 培养学生信息检索与利用能力的意义	29
1.4.2 中国大学信息检索教育的发展	31
参考文献及信息来源	41
2 检索标识系统	42
2.1 检索标识系统概述	42
2.1.1 检索标识系统定义	42
2.1.2 检索标识系统的类型	43
2.1.3 检索标识系统的功用	43

2.2 主题法标识系统	44
2.2.1 标题词标识系统	44
2.2.2 关键词标识系统	45
2.2.3 单元词标识系统	46
2.2.4 叙词标识系统	47
2.2.5 《汉语主题词表》简介	48
2.2.6 国内编印的部分专科主题词表题录	48
2.3 分类法标识系统	49
2.3.1 分类法标识系统的概念	49
2.3.2 分类法标识系统的类型	50
2.3.3 体系分类法的构成原理和基本作用	51
2.3.4 体系分类表的基本结构	52
2.3.5 我国主要分类法介绍	53
2.3.6 国外主要分类法简介	62
附录 国外叙词表选介	65
参考文献及信息来源	72
3 书本式参考工具书和中文检索工具的使用	73
3.1 事实与数据检索	73
3.1.1 检索流程和工具书的使用	73
3.1.2 词典	76
3.1.3 百科全书	77
3.1.4 年鉴、手册	83
3.1.5 名录、指南	87
3.2 中文检索工具体系	91
3.2.1 中文科技检索刊物系列	91
3.2.2 中文检索刊物的结构体例	93
3.3 社会科学信息检索	95
3.3.1 社科咨询参考工作	95
3.3.2 社会科学检索工具书举要	100
附录 2006 年第 1 期《全国报刊索引》(自然科学技术版)著录条目格式示例	104
参考文献及信息来源	113
4 中文数据库检索	114
4.1 万方数据资源系统	123
4.1.1 概况	123
4.1.2 万方数据资源系统的检索	124
4.2 中国期刊网	130

4.2.1 概况	130
4.2.2 中国期刊网全文数据库检索方法	131
4.3 全文数据库——《中国学术期刊(光盘版)》	136
4.3.1 概况	136
4.3.2 检索方法	137
4.4 《中文社科报刊篇名数据库》	139
4.4.1 概况	139
4.4.2 检索方法	139
4.5 《复印报刊资料》索引数据库、《复印报刊资料》专题目录索引光盘 数据库和全文数据库	140
4.5.1 概况	140
4.5.2 检索方法	141
4.6 《中国科学引文数据库》和《中文社会科学引文索引》数据库	143
4.6.1 概况	143
4.6.2 检索方法	145
参考文献及信息来源	150
5 EI、SA、SCI 及其检索	151
5.1 EI 及其检索	151
5.1.1 《工程索引》概况及特点	151
5.1.2 Engineering Village 2 平台中的 Compendex	152
5.1.3 纸印本《工程索引》	159
5.2 SA 及其检索	164
5.2.1 纸印本 SA 的概况和特点	165
5.2.2 INSPEC 网络版	175
5.2.3 INSPEC Ondisc 光盘版	176
5.3 美国科学引文索引(SCI)及其检索	177
5.3.1 《科学引文索引》概况	177
5.3.2 SCI 结构与著录格式	179
5.3.3 SCI 的检索策略与方法	184
5.3.4 ISI Web of Science	187
参考文献及信息来源	196
6 专利文献及其检索	197
6.1 专利和专利文献概况	197
6.1.1 专利与专利制度	197
6.1.2 专利文献	198
6.2 《国际专利分类表》	199

6.2.1 《国际专利分类表》概况	199
6.2.1 《国际专利分类表》的分类及其使用	200
6.3 中国专利文献及其检索	201
6.3.1 中国专利说明书	201
6.3.2 中国专利文献网络检索系统	201
6.3.3 中国专利文献印刷型检索工具	206
6.4 世界专利及其检索	207
6.4.1 世界专利概况	207
6.4.2 网上世界专利检索	208
6.4.3 印刷版德温特世界专利检索体系	223
参考文献及信息来源	229
7 几种专业性信息系统及其检索	230
7.1 美国《化学文摘》及其检索	230
7.1.1 《化学文摘》概况	230
7.1.2 印刷版《化学文摘》及其检索	231
7.1.3 CA on CD(CA 光盘版)	235
7.1.4 SciFinder Scholar(CA 网络版)	237
7.2 美国《生物学文摘》及其检索	241
7.2.1 《生物学文摘》概况	241
7.2.2 印刷版 BA	242
7.2.3 OVID BIOSIS PREVIEWS 数据库	242
7.3 生物信息学数据库及其检索	252
7.3.1 生物信息学数据库类型	252
7.3.2 生物信息学数据库检索	258
7.4 医学信息资源及其检索	268
7.4.1 中国生物医学文献数据库(光盘版)检索	268
7.4.2 MEDLINE 光盘数据库检索	281
7.4.3 互联网上的 PubMed 检索	291
7.5 美国《数学评论》(Mathematical Reviews)	304
7.5.1 概况	304
7.5.2 印刷型 MR	305
7.5.3 电子型 MR 的检索	305
7.6 《金属文摘》	313
7.6.1 概况	313
7.6.2 MA 分类表	313
7.6.3 印刷版 MA 体系	314
7.6.4 METADEX 数据库	316
7.7 美国《应用力学评论》	319

7.7.1 《应用力学评论》概述	319
7.7.2 AMR 在线期刊	319
参考文献及信息来源	322
8 特种文献及其检索	323
8.1 学位论文及其检索	323
8.1.1 学位论文概述	323
8.1.2 学位论文数据库——万方系统资源介绍	324
8.1.3 CALIS 高校学位论文数据库	328
8.1.4 国际学位论文文摘数据库 PQDD	330
8.1.5 ProQuest 学位论文全文数据库	336
8.1.6 印刷型《国际学位论文文摘》及其检索	338
8.2 会议文献及其检索	340
8.2.1 会议文献概述	340
8.2.2 会议文献的检索工具	341
8.2.3 会议论文常用网络检索系统介绍	344
8.2.4 会议论文印刷型检索工具	346
8.3 科技报告及其检索	353
8.3.1 科技报告概述	353
8.3.2 美国政府科技报告	353
8.3.3 美国政府科技报告的检索工具及检索方法	354
8.3.4 其他各国科技报告的检索工具	361
8.4 标准文献及其检索	362
8.4.1 标准文献概述	362
8.4.2 国内标准情况及其印刷型检索工具	363
8.4.3 国内标准的网络检索	364
8.4.4 国际标准、美国标准及其印刷型检索工具	368
8.4.5 国际标准及国外主要标准的网络检索工具	373
参考文献及信息来源	376
9 多媒体信息检索	378
9.1 多媒体信息系统概述	378
9.1.1 多媒体系统的特点	378
9.1.2 多媒体系统的技术基础	379
9.2 多媒体信息检索的原理	384
9.2.1 基于文本的多媒体信息检索	385
9.2.2 基于内容的多媒体信息检索	385
9.2.3 多媒体信息检索的发展	391
9.3 多媒体信息检索的实现	394

9.3.1 图像信息检索	394
9.3.2 音频信息检索	396
9.3.3 视频信息检索	396
9.4 多媒体检索系统介绍	398
9.4.1 QBIC	398
9.4.2 CueVideo	399
9.4.3 Photobook	399
9.4.4 Visual SEEK	399
9.4.5 香港中央图书馆的多媒体信息系统(MMIS)	399
9.4.6 VideoQ	399
9.4.7 Cambridge 大学的 VMR 和 MDR	400
9.4.8 Compaq 的 SpeechBot	400
9.4.9 CORE	400
9.4.10 美国 NSF、ARPA 和 NASA 资助的项目	400
9.4.11 TV-FI	401
9.4.12 MIRES	401
参考文献及信息来源	402
10 网络信息检索	403
10.1 概述	403
10.1.1 网络信息检索的发展	403
10.1.2 搜索引擎的缘起	405
10.1.3 搜索引擎的工作原理	406
10.1.4 搜索引擎的检索	407
10.2 常用搜索引擎的使用	411
10.2.1 Google	411
10.2.2 百度	417
10.2.3 天网	422
10.2.4 FTP 星空搜索	424
10.2.5 Yahoo! 中文	425
10.2.6 搜狗	426
10.2.7 新浪爱问	427
10.2.8 Go/Infoseek	427
10.3 其他网络信息检索服务	428
10.3.1 P2P 对等网络	428
10.3.2 RSS 服务	429
参考文献及信息来源	431

信息检索(Information Retrieval),简单地讲,是指知识信息的有序化识别和查找的过程。广义的信息检索包括信息的汇集、存储与检索查找,而狭义的信息检索仅指有序化知识信息的检索查找。换句话说,就是在浩瀚的信息海洋中,有的放矢,查找所需要的有价值或能够帮助解决问题的知识,即是从众多信息中查询并获悉所需的相关数据、事实和知识信息的过程和方法。信息检索是经济生活和科学研究的一个重要组成部分,信息检索的方法与技能在人才培养、管理和科研、生产与建设、经济技术和科学知识的创新等活动中有着重要的作用。本书所讲的信息检索,包含了知识传播和知识检索,主要是指文献资料信息和相关知识的查找与利用。

1.1 信息·知识·文献

1.1.1 信 息

当今时代,信息如同空气一样,跟我们的生活息息相关。信息在包围着我们,人的五官、身体所感受到的可以说首先都是信息。同时,人们又在借助语言等手段向外传递、交流信息。人类社会,即使人们用非语言的交流符号,也在本能地用一些常见的形式向外界发布信息,如:用体态语言,即用手势、面部表情、体态表情;用眼光,所谓“含情脉脉”、“暗送秋波”等,向外传递自身的信息。人与人交流的空间距离,也无时不在表达一种信息:相距很近,为亲密区;相距1米左右,属于熟人区;相距1~3米左右,属于社交区;相距4米以上的,属于公众交际区。一个人同你谈话时,隔着很远的地方而不靠近的,这时他已向你传递了这样的信息:不愿和你过分的亲密。人们在交往中还会用些“喊”、“哭”或某些动作、手势等辅助语言传递信息。这些非语言传递的信息,有时比用语言传递的信息更快速、更真切、更直接。由此来看,信息联系着你和我,联系着大家,联系着世间万物。人类社会实践也在源源不断地产生出新的信息。但是,对于信息是什么,对于信息的概念和定义的解释,众说纷纭,莫衷一是。

控制论的创始人维纳说:“信息是人们在适应外部世界并且使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。”

信息论的创始人申农从研究通信理论出发,认为信息是关于环境事实的可通