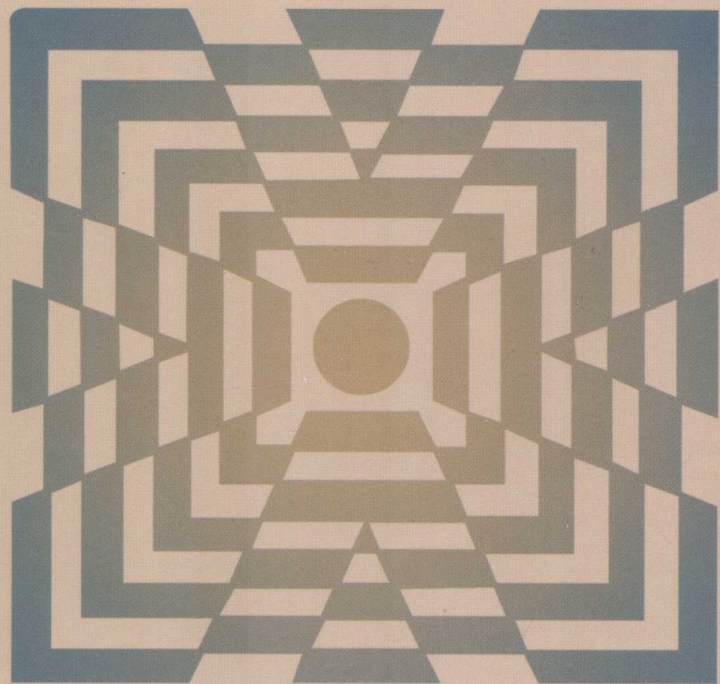


网络信息检索 原理与应用

编著 张英福 包和平 宛文红



 内蒙古大学出版社
Inner Mongolia University Press

大连民族学院教学改革项目(项目编号:YA2-004-04)

信息检索系列教程
网络信息检索原理与应用

编 著 张英福 包和平 宛文红
主 审 蔡明德

内蒙古大学出版社

2004 年·呼和浩特

图书在版编目(CIP)数据

网络信息检索原理与应用/张英福,包和平,宛文红
编著. —呼和浩特:内蒙古大学出版社,2005.7
(信息检索系列教程)

ISBN 7-81074-859-9

I. 网… II. ①张… ②包… ③宛… III. 计算机
网络—情报检索—教材 IV. G354.4

中国版本图书馆 CIP 数据核字(2005)第 088784 号

书 名	网络信息检索原理与应用
编 著	张英福 包和平 宛文红
责任编辑	石 斌 邓池君
封面设计	张燕红
出 版	内蒙古大学出版社 呼和浩特市大学西路 235 号(010021)
发 行	内蒙古新华书店
印 刷	赤峰地质宏达印刷有限责任公司
开 本	850×1168 1/32
印 张	11
字 数	276 千
版 期	2005 年 7 月第 1 版 2005 年 7 月第 1 次印刷
标准书号	ISBN 7-81074-859-9/Z·7
定 价	全套 230.00 元(本册 22.00 元)

本书如有印装质量问题,请直接与出版社联系

大连民族学院信息检索系列教程编审委员会

主任委员:黎树斌

副主任委员:包和平 王秀文

委	员:	马林	马春东	王立东	王培昌
		田森	刘向东	刘玉彬	刘江龙
		刘勇奎	闫严	江远	杜元虎
		杜智殊	宋敏	李光浩	李鸿
		徐国凯	南文渊	段晓东	徐辑彦
		袁晓宁	戴向平		

主 编:包和平

副 主 编:刘江龙 王学艳

编	委:	王波	刘敏	闫海新	陈新颜
		张敬荣	宛文红	郑巧红	金英玉
		金玲			

《信息检索系列教程》总序

由大连民族学院图书馆研究馆员包和平同志主持的教学改革项目《信息检索分类教学研究》课题,在东北电力学院图书馆、吉林省图书馆、东北财经大学图书馆、大连理工大学图书馆、大连大学图书馆和大连市图书馆的部分多年从事信息检索教学和研究的同志共同努力下,已取得初步研究成果。其研究成果《信息检索系列教程》正式出版了,这是大连民族学院教学改革项目取得突破性进展的一件大事、喜事,也是我国信息检索教学和研究的是一件大事、喜事,我代表大连民族学院教学工作指导委员会向勤奋耕耘并取得丰硕研究成果的编著者们表示热烈而诚挚的祝贺。

信息技术的高速发展,促进了社会的信息化,使信息资源成为国家的重要资源。物资、能源和信息,被称为现代社会生产力的三大支柱。信息资源的开发利用水平已成为一个国家综合国力的重要标志。

信息检索(Information Retrieval)是一门关于信息资源存储、整序和查找理论、方法的学问,其研究内容包括检索理论、检索语言、信息著录法、标引法和排检法、检索系统、检索策略、数据库、检索服务等,它是在传统的文献检索基础上,融合了最新信息存储与检索技术发展起来的。随着人类社会信息化程序的日益提高,人们对信息的渴求激增,信息的获取和利用成为人类赖以生存与发展的一种本能,信息检索技术成为每个大学生和科研人员必备的基本技能之一。

从20世纪80年代初期开始,我国高校对大学生进行信息技能教育主要是通过开设文献检索与利用课的形式来进行的。自1981年举行全国高校图书馆工作会议以后,教育部颁发了《中华人民共和国高等学校图书馆工作条例》,正式提出高等学校图书馆

应承担向在校学生“开展查阅文献方法的教育和辅导工作”的任务。1984年2月22日,教育部下发了《关于在高等学校开设文献检索与利用课的意见》。1985年9月,国家教委又下发了《关于改进和发展文献课教学的几点意见》。为了总结、交流文献信息检索教育的经验,全国高校图书情报工作委员会秘书处先后多次召开了文献检索课教学研讨会、文献检索课教学基本要求定稿会及工作会议。

1986年,经国家教委批准,全国高校图书情报工作委员会组织成立了文献检索与利用课系列教材编审委员会,负责规划、组织教材的编写和审订。推出了一套系列教材30种,从1987年陆续出版。这套书既有覆盖面较宽的通用教材,又有针对性强的专业性教材,而且,教材的编者都是有多年教学经验的老师,在同类书籍的基础上进行精编、补充、改革、修订后的结果。注重反映新成果,力求理论与实践的结合,注意基础理论的传授和检索能力的培养,有些还配有实习指导书。这一时期,一些院校又自编了一些专指性强的专业性教材,形成了百家争鸣的繁荣局面。

近年来,随着现代化信息技术、通讯技术的迅速发展,世界信息环境发生着巨大变化,伴随网络化、数字化而产生的大数量、多类型、多媒体、非规范、跨地域、跨行业、跨语种的网络信息资源,对原来相对集中和规范为标志的传统文献资源产生了突破性发展。信息用户、网络信息资源以及以计算机和现代通讯技术为核心的信息处理技术,构成了新的信息环境。在这新信息环境下,信息资源的组织、查询与利用方式随之发生了根本性的变革,培养学生获取和利用信息能力已成为高等教育教学活动的基本要求之一,信息检索课教学成为实现大学生获取与利用信息能力培养的重要教育环节。过去所编撰的相关教材已远远不能适应当前教学的需要。面对这一实际情况,包和平同志根据辽宁省教育厅《关于印发辽宁省普通高等学校〈信息检索与利用〉课教学大纲的通知》(辽

教办发[2000]87号)精神,结合我院实际需要,组织省内外信息检索课教师及部分专家率先开展信息检索分类教学研究,及时编撰出版了这套《信息检索系列教程》,可以说意义深远。该项研究不仅对现行的信息检索技能教育课程进行了全方位的改革和探讨,而且形成了多元化、多层次、集合式的教学模式。对拓宽信息检索教育的专业领域,推进并完善我国信息检索教育模式,实现大学生信息素养与创新能力的提高均具有深远的理论意义与实践意义。

信息检索教学的目标就是培养学生有效地管理、获取、评价和利用信息的一整套技术能力,是信息素质教育的重要组成部分。《信息检索分类教学研究》课题结合学院的学科建设和专业设置,制定了有专业特色的教学计划,并通过广泛调查、系统分析、综合研究等方法,将各学科信息资源的分布动态、管理模式、传播途径及检索技术等系统化、理论化,编制出各学科、各专业的适用性信息检索教程;同时,在学院教务管理机构的指导和配合下,实现分专业授课的教学模式,针对不同专业的教学对象,实施多样化教学方式,传授专业化教学内容,设计适用性实践方案。信息检索分类教学模式的研究和配套教材的编制,正是当前大力开展信息检索教学改革进程中最具有创新意识和发展潜力的举措。

作为《信息检索分类教学研究》课题的成果——《信息检索系列教程》共分10个分册,即《网络信息检索原理与应用》、《文史信息检索》、《经济文献信息检索》、《计算机科学与技术信息检索》、《机电工程信息检索》、《土木建筑信息检索》、《化学化工信息检索》、《生物医学信息检索》、《期刊管理与信息检索》、《民族文献信息检索》。其中《网络信息检索原理与应用》以信息化社会大力推进全民信息素质教育为时代背景,对网络信息检索技术的原理、运用与操作技巧等诸多理论与实践问题做了深入研究与探讨,其独到之处在于将深刻的理论内涵同最前沿的信息技术和网络技术融会贯通,开拓并延伸了网络环境下高效率地获取与利用海量信息

资源理论基地和实践体系,包括书中运用中英文信息数据库的检索实例来说明检索技术和操作技巧,都充分体现了信息检索的理论指导性和现代化检索技术的实用性。其他各分册分别针对不同层次,不同专业的大学生的信息需求,提供了一系列覆盖面广、且较精炼实用的教材,其特点主要包括以下几个方面:

一是重点突出。各分册在内容的取舍、结构的安排和检索实例的采用方面,密切结合各学科及其所需信息的特点,既有理论知识,又有实际事例;既有信息检索的内容,也有信息积累和利用的知识。重点突出,兼顾一般,互为补益,相得益彰。

二是内容新颖。该系列教程的编著者们密切关注信息及信息检索工具的发展动态和最新成果,尽可能使教程的内容立足于知识潮流的前沿,尽可能地反映最新的信息检索理论研究成果,追求具有时代特征的学术前瞻性和应用指导性,以满足新世纪信息检索课教学的需要。

三是短小精悍。信息检索课是一个实践性很强的课程,可以讲的内容比较多,但为了不加重学生的经济负担和课程压力,并兼顾信息检索语言和索引体系的系统性,重点选择了那些对学习最需要的内容,力求做到学以致用原则和信息检索所具有的方法论特征。

四是优化组合。信息检索所涉及的学科是极其广泛的,有关专业也较多,为使该系列教材有较广的覆盖面,在各分册内容结构的安排上,基本上本着便于按学科需要进行模块式组合的原则,并把重点放在那些对各学科有较大共性的检索工具书上,以利不同学科、专业按需要灵活进行组合。除阐述了各种信息资源、检索工具书以及大学生必须掌握的手检知识与技能外,对各类型数据库检索、联机检索与 Internet 网络检索作了重点介绍,并在手检与机检相结合方面进行了有益的探索。

总之,呈现在我们面前的这套《信息检索系列教程》表现出较

高的学术水平和实用价值,散发出浓浓的新意,可以给人以教益,因而是一套值得推荐的好教程。我相信该系列教程的出版,将有力推动我国现代信息检索理论研究和实践教学活动的发展,同时也相信每一个认真读过它的大学生,都会从中得到启迪。是为序。

蔡明德

甲申年中秋于民院红梅寓所

前 言

本书以信息化社会大力推进全民信息素质教育为时代背景,以飞速发展的 Internet 信息存储与检索技术为研究和探索对象,对网络信息检索的理论体系、研究对象、研究方法,网络信息检索技术的原理、运用与操作技巧等诸多理论与实践问题做了系统论述和探讨。

与目前国内出版的信息检索类教材相比,其独到之处在于将信息检索的理论内涵同最前沿的信息技术和网络技术融会贯通,开拓并延伸了网络环境下高效率地获取与利用海量信息资源的理论基地和实践体系。本书前四章系统论述了网络信息资源的概念、类型与服务方式,网络信息检索的基本原理等理论问题;后四章以此为理论基础,对搜索引擎等网络信息检索工具和常用中外文信息数据库的检索方法,各种网络信息资源的搜索途径进行了详细介绍。内容兼具理论性与实用性,充分体现了信息检索学的理论指导性和网络检索技术的实用性。

全书共分八章,编著人员分工如下:第一、二章为包和平,第三、六、七章和附录为宛文红,第四、五、八章为张英福。担任本教材主审的是蔡明德教授。

本书是大连民族学院教学改革项目《信息检索分类教学研究》课题的成果之一。在编著过程中得到大连民族学院蔡明德院长、黎树斌副院长以及教务处同志的大力支持,在此,对本书的所有支持和帮助者表示深深的谢意!同时,本书在写作过程中参阅并研究了大量网站、信息数据库以及相关文献,得到许多宝贵的经验和启发,在此对所有参考文献的作者和网站、信息数据库创建者表示

衷心感谢!

囿于作者能力、水平及所见资料数量,本书尚有许多不尽人意之处,讹误、错漏在所难免,敬请读者批评指正。

张英福

2004年10月20日

内容提要

本书是一部系统论述网络信息检索理论与技术的学术论著。全书共分八章,内容兼具理论性和实用性,理论部分主要阐述了网络信息资源的概念、类型与服务方式,网络信息检索基本原理与技术,网络信息检索系统与信息数据库检索方法;应用部分重点介绍了中外著名网络信息数据库检索方法,各种网络信息资源的搜索途径以及数字图书馆的应用与发展。本书对提高信息素质,掌握网络信息检索与利用技能具有全面性的指导作用和参考价值,适合科研人员、教师、学生等各类读者群。

目 录

第 1 章 信息与信息资源概论	1
1.1 信息	1
1.1.1 信息的定义	1
1.1.2 信息的特征	3
1.1.3 信息化社会	5
1.1.4 信息化社会的生存策略	7
1.2 信息资源	8
1.2.1 信息资源的涵义	8
1.2.2 信息资源的类型	9
1.2.3 文献信息资源	10
1.2.4 电子信息资源与网络信息资源	18
第 2 章 网络信息检索概论	25
2.1 信息检索的涵义	25
2.1.1 什么是信息检索	25
2.1.2 信息检索的功能和作用	26
2.2 信息检索的发展简史	28
2.2.1 手工信息检索阶段	28
2.2.2 计算机信息检索阶段	31
2.3 网络信息检索的现状与发展趋势	33
2.3.1 Internet 简介	33
2.3.2 什么是网络信息检索	47
2.3.3 网络信息检索的发展状况	49
2.3.4 网络信息检索的发展趋势	53
第 3 章 网络信息检索原理与方法	56
3.1 信息检索语言	56

3.1.1	信息检索语言的定义	56
3.1.2	信息检索语言的功能	57
3.1.3	信息检索语言的类型	57
3.1.4	等级体系分类语言	58
3.1.5	主题法语言	69
3.2	计算机信息检索技术	79
3.2.1	布尔逻辑检索	79
3.2.2	截词检索	81
3.2.3	字段检索	84
3.2.4	位置逻辑检索	86
3.2.5	加权检索	88
3.3	检索策略	89
3.3.1	检索策略的涵义	89
3.3.2	检索策略的制定	89
3.4	网络信息资源的著录	94
3.4.1	机读目录	95
3.4.2	元数据(Metadata)	113
第4章	网络信息检索系统	120
4.1	网络信息检索系统概述	120
4.2	网络信息检索系统的构成	122
4.2.1	物理结构	122
4.2.2	逻辑结构	126
4.3	网络信息检索系统的评价指标	128
第5章	Internet 信息检索工具及其使用	130
5.1	万维网(WWW)	130
5.1.1	相关概念	130
5.1.2	WWW 工作原理	132
5.2	WWW 信息检索工具及其类型	133

5.2.1	搜索引擎及其类型	133
5.2.2	常用中外文综合型搜索引擎	137
5.3	非万维网信息检索工具	167
5.3.1	文件检索工具 Archie	167
5.3.2	菜单式信息检索工具 Gopher	167
5.3.3	全文检索工具 WAIS	168
第 6 章	网络信息数据库的检索	169
6.1	网络信息数据库服务机构选介	169
6.1.1	中国高等教育文献保障系统(CALIS)	169
6.1.2	中国知识基础设施工程(CNKI)	171
6.1.3	万方数据资源系统	173
6.1.4	OCLC 及其 FirstSearch	179
6.1.5	美国工程信息公司	185
6.2	文献数据库	187
6.2.1	文摘索引数据库	188
6.2.2	全文数据库	205
6.3	事实型数据库	233
6.3.1	万方数据资源系统事实数据库	234
6.4	数据型数据库	243
第 7 章	各种网络信息资源的检索	249
7.1	电子期刊的网络检索	249
7.1.1	电子期刊概述	249
7.1.2	电子期刊的网络检索途径	252
7.2	电子报纸的网络检索	260
7.2.1	电子报纸概述	260
7.2.2	电子报纸的网络检索途径	263
7.3	电子图书的网络检索	265
7.3.1	电子图书概述	265

7.3.2	网络电子图书的文件格式及阅读方式	268
7.3.3	电子图书的网络检索方法	275
第8章	数字图书馆	283
8.1	数字图书馆概述	283
8.1.1	数字图书馆的演变、发展	283
8.1.2	数字图书馆的定义与特征	285
8.1.3	数字图书馆的类型	287
8.2	数字图书馆的信息服务模式	288
8.2.1	用户信息需求与数字图书馆信息 服务模式的关系	289
8.2.2	数字图书馆信息服务模式的定位	291
8.2.3	以用户为中心的数字图书馆信息服务模式	293
8.3	我国数字图书馆的发展及其前景	297
8.3.1	我国数字图书馆发展过程中主要的研究开发项目	297
8.3.2	我国数字图书馆的发展前景	302
8.4	数字图书馆的利用	303
附录1	中国六大主干网及主要教育、科研机构网址	309
附录2	中国图书馆网址	314
附录3	世界主要图书馆网址	321
附录4	中外文搜索引擎目录	323
附录5	中国期刊网址	328
附录6	世界主要期刊网址	330
附录7	中国报纸网址	335
附录8	世界主要报纸网址	339
附录9	国外100所著名大学网址	342
	参考文献	350

第 1 章 信息与信息资源概论

1.1 信 息

1.1.1 信息的定义

随着计算机网络技术和通信技术的飞速发展,信息的创建、加工、传播、存储及检索技术获得了前所未有的突破,使信息的可利用率显著提高,社会价值日趋突出,成为社会经济发展的支柱性资源,人们对信息的重视达到了空前状态,对信息科学涵义的理解也日渐深刻。

信息一词的拉丁词源是 Information,意思是通知、报道或消息。在中国历史资料中,唐代诗人李中在《碧云集》—《暮春怀故人》中曾有“梦断美人沉信息,目穿长路倚楼台”的诗句,其中的“信息”是音信、消息之意。信息的这一涵义一直沿用到 20 世纪中叶,此后伴随着信息论、控制论的产生,其诸多科学特性才被逐渐揭示出来。到目前为止,关于信息的定义已有百种之多,它们都从不同的侧面揭示了信息的某些属性和特征,但至今尚无一种定义,因其内涵的全面性、深刻性和科学性而被社会各界一致认同。

国外关于信息的定义以申农(C. E. Shannon)和维纳(N. Wiener)的界定最为著名。1948 年申农博士在其著名的《通信的数学理论》一文中提出信息是“两次不确定性的差异”,用以消除随机不确定性的东西,这一理论奠定了信息论的基础。1950 年美国著名科学家、控制论创始人维纳在《人有人的用处》中对信息的界定是“信息是我们用于适应外部世界,并且在使这种适应外部世界