



普通高等教育“十一五”国家级规划教材

21世纪 信息资源管理系列教材

# 信息检索教程 (第二版)

王立清/主编

 中国人民大学出版社

普通高等教育“十一五”国家级规划教材  
21世纪信息资源管理系列教材

# 信息检索教程

(第二版)

王立清 主编

中国人民大学出版社

· 北京 ·

图书在版编目 (CIP) 数据

信息检索教程 (第二版) / 王立清主编. — 2 版.

北京: 中国人民大学出版社, 2008

(普通高等教育“十一五”国家级规划教材; 21 世纪信息资源管理系列教材)

ISBN 978-7-300-09671-1

I. 信…

II. 王…

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 (2008) 第 136035 号

普通高等教育“十一五”国家级规划教材

21 世纪信息资源管理系列教材

信息检索教程 (第二版)

王立清 主编

---

出版发行 中国人民大学出版社

社 址 北京中关村大街 31 号

电 话 010-62511242 (总编室)

010-82501766 (邮购部)

010-62515195 (发行公司)

网 址 <http://www.crup.com.cn>

<http://www.ttrnet.com>(人大教研网)

经 销 新华书店

印 刷 北京民族印务有限责任公司

规 格 170mm×228mm 16 开本

印 张 25.5

字 数 467 000

邮政编码 100080

010-62511398 (质管部)

010-62514148 (门市部)

010-62515275 (盗版举报)

版 次 2004 年 9 月第 1 版

2008 年 10 月第 2 版

印 次 2008 年 10 月第 1 次印刷

定 价 45.00 元

## 作者简介

王立清，女，1965年生于山西大同，汉族，管理学博士，中国人民大学信息资源管理学院副教授，情报学专业硕士生导师。主要研究方向：信息组织与信息检索、政府信息资源管理、信息资源数字化出版等。主持和参与国家社科基金及其他各类科研项目多项。主要著作有《政府网站的构建与运作》。主要论文有《澳大利亚、新西兰政府网站建设的元数据标准》、《新技术革命与图书馆业务流程重组》等。

## 内容简介

本书为普通高等教育“十一五”国家级规划教材。

主要内容包括信息检索概念和原理、信息检索系统、信息检索效果、检索语言、信息著录和标引、印刷型工具书检索、计算机检索概述、联机检索、网络信息检索、特种文献检索等。尤其对网络环境下检索语言的发展、网络信息检索的特点和方法、网络信息检索工具、网络数据库等进行了重点解析。修订后的本教材继续为读者提供完整的信息检索理论框架，同时补充丰富了信息检索领域新进展的相关内容，新增了信息检索的实践案例，努力突出基础性强、实用性好、新颖性明显、结构清晰、适用面广等特点。

本书可作为高等院校信息管理与信息系统专业及图书馆学、档案学等相关专业本科生信息检索课程核心教材，亦可作为各类信息管理机构工作者的培训教材和参考书，还可供广大信息检索爱好者阅读和参考。

## 第2版前言

信息检索教程(第二版)

P R E F A C E

信息检索是一门以信息及其相关检索系统的特点及使用方法为研究对象的课程,主要内容包括信息检索基本理论和原理、信息检索语言、信息检索策略、搜索引擎、网络数据库、印刷型工具书等,旨在培养和训练学生检索与检考、选择与鉴别、挖掘与处理文献信息资源的能力,使学生了解并掌握各种信息检索方法,学会运用各种信息检索系统从各类信息中获取所需要的知识。本课程作为信息资源管理领域一门重要的核心课程,是一门培养大学生的信息素养、提高自我知识更新能力和掌握信息检索技能的实践课程,是21世纪高校信息素养教育的重要课程之一,对于培养学生的信息素质与信息能力,掌握科学的方法进行文献信息的收集、整理、加工和利用,提高学生的自我学习能力和创新能力具有重要的意义。

《信息检索教程》一书初版于2004年9月,至今已近四年。此间,信息检索领域发生了很多的变化,信息资源数量有了极大的增长,全文检索技术得到了进一步发展和完善,搜索引擎不断推陈出新,网络数据库日益丰富,互联网平台成为用户获取信息的重要途径。另外,本书自出版后一直被用作中国人民大学信息管理信息系统、档案学专业的信息检索课程教材,也被国内其他一些院校所采用。在教学实践中,我们努力追踪信息检索的发展动态,积累信息检索的相关素材和案例,这些都为本次修订奠定了良好的基础。

本次修订对全书结构进行了调整,增加了第6章“计算机信息检索概述”和第12章“特种文献检索”,去掉了原来的第4章“信息检索策略和方法”,将其内容分散至相关章节;全面更新了

搜索引擎、网络资源目录、元搜索引擎、网络数据库等内容,补充介绍了新出现的网络信息检索系统;增加了一定的篇幅阐述搜索引擎、网络信息检索发展趋势;在每章之后增加了相应的案例。期望修订之后,本书能够更加全面、准确、及时地反映信息检索领域的最新发展成就,更好地满足信息检索教学和相关人员参考学习的需求。

本书修订后为12章。其中,第1、6章由王立清负责修订和编写,第2章由汤坚玉负责修订,第3章由王立清和宋薇负责修订,第4、5章由刘丽负责修订,第7、8、10章由汤坚玉和王立清负责修订,第9章由汪娟负责修订,第11章由杨凡和汤坚玉负责修订,第12章由王立清和汤坚玉负责编写。全书最后由王立清统稿。此外,汤坚玉参与了全书的校对工作,宋薇对本书配套的PPT讲稿制作付出了很多。

在网络环境下,信息检索的发展可谓日新月异,限于作者的学术水平,书中难免有疏漏和不妥之处,敬请业界内同行和广大读者批评指正。

王立清

2008年6月于中国人民大学

# 第 1 版前言

信息检索教程(第二版)

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,尤其是互联网的发展,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势头;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。

前言

1 8 至 100

书中内容分为 4 个部分:信息检索基础理论、印刷型工具书检索、光盘检索和联机检索、网络信息检索。主要内容包括信息检索的原理和方法与意义、检索评价、检索语言、信息著录与索引、检索策略、印刷型工具书的概况和使用、光盘检索概况及主要光盘数据库检索、联机检索系统的特点及利用、网络检索的特点、网络检索工具的原理和使用方法、网络数据库等。

与现有的同类教材相比,本书具有如下特色:

1. 全面与重点相结合,突出网络检索。本书试图构筑一个

科学的、全面的现代信息检索内容体系,融传统的中文工具书、英文工具书、社科文献检索和现代网络检索为一体,系统阐述信息存储和信息检索整个过程,展现信息检索的全貌。同时,伴随网络环境所引发的信息资源数字化、信息检索在线化趋势,互联网成为全球最大的信息资源宝库,网络检索已成为大学生获取信息的最重要方式,本书重点介绍了网络信息资源的检索,以体现网络环境下信息检索的新特点,吻合当前信息检索的实际。

2. 理论与实践相辅助,突出实用性和可操作性。信息检索重在培养学生获取信息资源的实际能力,本书在介绍检索工具时注重阐述其检索功能,对于重要的检索工具和数据库辅之以检索实例,而且,每章后附思考题。

3. 反映信息检索领域的新成果、新趋势,突出内容的新颖性。本书参考了近年来大量的国内外相关资料和研究成果,注重介绍信息检索在网络环境中的新变化和新发展,包括网络信息资源的特点、传统分类法和主题法的发展、自然语言的应用、网络分类法、网络信息检索标准、元数据、多媒体检索技术等。

4. 脚注详尽、规范,便于相关知识的扩展学习。本书对引用的主要观点、资料及相关内容进行了脚注,并对涉及的网站或网页资料注明了访问日期,在一定程度上起到了扩展本书内容的作用,增强了本书与相关信息资料的关联性,有利于获取有关资料和对问题的进一步研究。

本书第1章由冯惠玲、王立清、李林编写,第2章由冯惠玲、王立清编写,第3章、第4章由王立清编写,第5章、第6章由刘丽编写,第7章由王立清、唐宇萍编写,第8章、第9章、第10章由王立清编写,第11章由王立清、刘丽编写。全书由冯惠玲、王立清统稿。

由于时间仓促且水平有限,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。

编著者

2004年8月



# CONTENTS

信息检索教程(第二版)

## 目 录

<b>第1章 信息检索基础</b> .....	1
引子 .....	1
1.1 信息概述 .....	2
1.2 信息检索的概念和原理 .....	12
1.3 信息检索系统 .....	16
1.4 信息检索方法 .....	21
1.5 信息检索效果 .....	22
案例 .....	28
术语 .....	29
思考题 .....	29
<b>第2章 检索语言</b> .....	30
引子 .....	30
2.1 检索语言概述 .....	31
2.2 检索语言的理论基础 .....	35
2.3 分类检索语言 .....	39
2.4 主题检索语言 .....	53
2.5 分类主题一体化检索语言 .....	63
2.6 网络信息检索语言 .....	66
案例 .....	73
术语 .....	74
思考题 .....	75
<b>第3章 信息著录和标引</b> .....	76
引子 .....	76

3.1 信息著录的含义和标准 .....	77
3.2 机读目录与元数据 .....	82
3.3 信息标引的含义和步骤 .....	92
3.4 分类标引和主题标引 .....	96
3.5 自动标引 .....	102
案例 .....	107
关键术语 .....	111
思考题 .....	112
<b>第4章 参考工具书概述</b> .....	<b>113</b>
引子 .....	113
4.1 参考工具书的概念与特点 .....	114
4.2 参考工具书的种类与排检方法 .....	118
4.3 参考工具书的数字化 .....	126
案例 .....	134
关键术语 .....	135
思考题 .....	135
<b>第5章 参考工具书使用</b> .....	<b>136</b>
引子 .....	136
5.1 图书与知识型信息检索 .....	137
5.2 数据与事实型信息检索 .....	149
案例 .....	160
关键术语 .....	161
思考题 .....	161
<b>第6章 计算机信息检索概述</b> .....	<b>162</b>
引子 .....	162
6.1 计算机信息检索的含义和特点 .....	162
6.2 计算机信息检索策略 .....	167
6.3 信息检索技术 .....	176
案例 .....	181
关键术语 .....	182
思考题 .....	182

<b>第7章 联机检索</b> .....	183
引子 .....	183
7.1 联机检索系统概述 .....	184
7.2 主要联机检索系统简介 .....	192
案例 .....	205
关键术语 .....	207
思考题 .....	207
<b>第8章 光盘检索</b> .....	208
引子 .....	208
8.1 光盘检索系统 .....	209
8.2 主要光盘数据库选介 .....	213
案例 .....	219
关键术语 .....	222
思考题 .....	222
<b>第9章 网络信息检索概述</b> .....	223
引子 .....	223
9.1 网络信息资源分布 .....	224
9.2 网络信息检索原理与方法 .....	234
9.3 网络信息检索相关标准 .....	240
9.4 网络信息检索发展趋势 .....	250
案例 .....	259
关键术语 .....	260
思考题 .....	260
<b>第10章 网络信息检索工具</b> .....	261
引子 .....	261
10.1 网络信息检索工具的发展和类型 .....	262
10.2 搜索引擎 .....	264
10.3 网络资源目录 .....	291
10.4 元搜索引擎 .....	304
案例 .....	319
关键术语 .....	321
思考题 .....	322

<b>第 11 章 网络数据库检索</b> .....	323
引子 .....	323
11.1 网络数据库概述 .....	324
11.2 国外网络数据库检索示例 .....	330
11.3 中文网络数据库 .....	344
案例 .....	353
关键词 .....	354
思考题 .....	354
<b>第 12 章 特种文献检索</b> .....	356
引子 .....	356
12.1 科技报告检索 .....	357
12.2 会议文献检索 .....	363
12.3 学位论文检索 .....	367
12.4 专利文献检索 .....	371
12.5 标准文献检索 .....	375
12.6 档案文献检索 .....	382
案例 .....	384
关键词 .....	385
思考题 .....	385
<b>主要参考文献</b> .....	387

### 【本章要点】

- ◇ 解释信息的含义与特征
- ◇ 论述信息的分类
- ◇ 介绍信息检索的概念
- ◇ 阐述信息检索的原理
- ◇ 讨论信息检索系统
- ◇ 梳理信息检索方法
- ◇ 探讨信息检索效果评价

### 引子

奈比斯特说：“趋势就像一匹马一样，比较容易向着正在奔跑的方向奔去。”无疑，以知识经济为显著特征的信息社会，已经成为社会发展的一种无法抗拒的趋势。我们今天所处的时代被人们称为“信息社会”。什么是信息？信息是消息，人们在学习、工作、日常生活中随时随地都在接受和利用信息；信息是资源，它具有使用价值和价值；信息是财富，且是无价之财富；信息是生产力要素，更是一种不可估量的促进生产力发展的新动力。因此，人类社会的发展，科学技术的进步，都离不开信息资源的开发和利用。而且，信息资源的真实状况及开发利用程度，已经成为衡量一个国家经济、文化、科技以及综合国力的重要指标。我们

怎样快速地查找信息和有序地整理信息?信息检索是最快的途径。<sup>①</sup>

## 1.1 信息概述

### 1.1.1 信息的含义与特征

#### 1.1.1.1 信息的含义

信息(information)作为现今社会使用频率最高的词之一,显现着时尚,蕴涵着古老。人类很早就开始了信息活动,古波斯人塔顶设置的“喊话站”和古罗马人的“悬灯”讲述着古代信息活动的故事。我国远古的“结绳记事”和殷商的“烽火告警”,同样展现着信息存储和传递的方式。唐代诗人李中在《碧云集·暮春怀故人》诗中写道:“池馆寂寥三月尽,落花重叠盖莓苔。惜春眷恋不忍扫,感悟心情无计开。梦断美人沉信息,目穿长路倚楼台。琅轩绣珞安可得,流水浮云共不还。”这是汉语中较早使用“信息”一词的文字记录。南宋陈亮《梅花》诗曰:“一朵忽先变,百花皆后香,欲传春信息,不怕雪埋藏。”清朝的康熙是我国历史上第一个明确使用“信息”一词的皇帝,康熙三十四年(1695),他命大学士派人外出,“惟以侦探信息奏文为要”。当时,“信息”指音信、消息。<sup>②</sup>

人类历史推进到今天,信息引起了如此广泛、深入、持久的影响,信息化社会、信息高速公路、信息时代、信息系统、信息产业、信息技术……这一切都与信息紧密联系在一起。那么信息到底是什么呢?作为日常用语,信息就是消息,一切存在都有信息,如上课的铃声、网上发送和接收的E-mail、报纸杂志中的文章、电视播放的新闻等等。对人类而言,人的五官生来就是为了感受信息的,它们是信息的接收器,它们所感受到的一切都是信息。信息技术的发展,将会极大地帮助人类去感知和发现五官不能直接感受的大量信息。

对信息概念进行科学的探讨开始于20世纪20年代。1928年,哈特莱(R. V. L. Hartley)在《信息传输》中将信息理解为“选择通信符号的方式”。1948年,信息论创始人申农(C. E. Shannon)在《通信的数理理论》中将信息定义为“信息是用以消除随机不确定性的东西”。1950年,控制论创始人维纳

① 参见一凡的博客: [http://blog.163.com/yifan\\_998/blog/](http://blog.163.com/yifan_998/blog/), 2008-03-20。

② 参见孙凡:《信息的由来与发展》,载《情报科学》,1999(2)。

(N. Wiener) 将信息的概念引入了控制论, 在《人有人的用处——控制论与社会》中指出: “人是通过感觉器官感知外部世界”, “我们支配环境的命令就是给环境一种信息”, “信息这个名称的内容就是我们对外界世界进行调节, 并使我们的调节为外界所了解时而与外界交换来的东西”<sup>①</sup>。

关于信息的定义, 目前尚无统一的定论。不同领域的研究者从不同的角度出发, 对信息有不同的理解和认识。结合信息检索的特点, 我们认为信息的含义有广义和狭义之分。广义的信息指自然界和一切人类活动所传达出来的信号和消息, 是事物表现的一种普遍形式。从本质上说, 信息是事物自身(显示其存在方式或运动状态)的属性, 是客观存在的现象。狭义的信息指经过搜集、记录、处理和存储的可供检索的文献、数据和事实。它是人类对客观事物的认识, 是实践经验的总结, 是认识的结果, 是我们检索的对象。

### 1.1.1.2 信息的特征

信息的特征是指信息区别于其他事物的属性。信息的主要特征如下:

#### 1. 可存储性

信息可以存储, 存储和传递是信息的两种基本状态。大脑就是一个天然信息存储器。利用信息的可存储性, 人们可以有意识地将流动的信息以某种方式存储在物质媒介上, 使信息与物质媒介构成一种依附性很强的、相对稳定的关系。这种稳态的结构可以有效地避免信息的流失, 也使我们的信息检索有源可寻。

#### 2. 可传递性

可传递性是信息的本质特征之一, 指信息可以通过一定的传输工具和载体进行传递, 从而形成信息联系, 被人们感受和接收。信息的传递有空间传递和时间传递等不同类型, 需要依赖于一定的物质载体, 具有动态性和方向性的特征。正是信息的传递使人类掌握了更多的经验和知识, 推动了人类文明的进程。

#### 3. 可转换性

信息的可转换性表现在两个层面: 一是信息在一定的条件下可以转化成物质、能量、金钱、效益等其他东西, 这种转换主要依靠人类对信息的正确利用; 二是信息可以从一种形态转换为另一种形态, 自然信息可转换为语言、文字和图像等形态, 也可转换为电磁波信号或计算机代码。比如, 自然语言信息和机器语言信息的转换、不同语种信息的转换、不同载体信息的转换等等。

① 岳剑波:《信息管理基础》, 2页, 北京, 清华大学出版社, 2000。



#### 4. 可处理性

信息可以通过分类、整序、分析、综合、压缩、扩充等加工处理,而达到便于识别、效用更高的信息。人脑本身就是最佳的信息处理器,可以在感知信息的基础上,进行决策、研究、发明、创造等多种信息处理活动。计算机也同样具有信息处理功能,计算机可以输入各种数据文字等信息,进行相关的处理,以显示、打印、绘图等方式再生成信息。

#### 5. 可共享性

信息的共享性表现为同一种信息可以同时被许多人共同享用,这是信息不同于物质和能量的一个非常重要的特征。也就是说,数个接收者可以获得同一信源发出的同样信息,而在这一过程中,信息的内容不会减少或发生改变。信息可以广泛地扩散和传播,信息交换的双方不会失去原有的信息,而且还会增加新的信息。

#### 6. 可识别性

信息可采取直观识别、比较识别和间接识别等多种方式来感知和识别。信息作为表现事物特征的一种普遍形式,反映了事物的运动状态和存在方式,人类可以通过自身的器官去直接感觉和知觉信息,通过比较去认识信息,借助于先进的信息技术和手段去识别信息。

#### 7. 依附性

信息无法脱离物质而独立存在,在其存储和传递过程中必须依附于一定的物质载体,信息与物质载体构成一个整体。我们将这些信息赖以存储和传递的物质载体称为信息载体,信息载体泛指一切载有信息的物质媒体。

#### 8. 普遍性

信息是物质的基本属性,普遍存在于自然界和人类社会之中,也存在于人类的思维或精神领域之中。只要有物质存在的地方,就有信息的存在,物质普遍存在的属性导致了信息的普遍性。

### 1.1.2 信息的功能和类型

#### 1.1.2.1 信息的功能

信息具有重要意义,与物质和能源一起构成现代社会发展的三大支柱。在一般情况下,人们将物质和能源称为有形资源或第一资源,把信息称为无形资源或第二资源。物质和能源分别为生产提供材料和动力,信息则为生产提供智力。在当今信息经济社会里,信息为生产所提供智力的重要程度已远远超过第一资源。信息的功能是指信息的功效和作用,主要表现在以下方面:



### 1. 传承人类文明, 推进社会发展

人类所有的知识、故事都是信息, 这些信息记载着人类文明发展的轨迹。自古以来, 信息的积累和传播, 成为人类承上启下的纽带。信息作为人类了解自然及人类社会的凭据, 与物质和能源共同奠定了社会发展的基础。

### 2. 提供决策依据, 提高决策效益

现代管理学认为, 决策是对管理的未来行动目标作决定, 是在两个或两个以上的可行性方案中选择出一个满意方案的过程。决策过程无论是环境分析、准则制定、方案生成还是选择评价, 无一不与信息的采集、分析与利用有关。信息能够帮助人们减少应对决策时的不确定性和风险, 降低由于缺乏足够准备而造成的损失。社会活动的日趋复杂化, 增大了决策的难度, 大部分决策都是在面临着多种可能出现的结果中作出抉择的, 这就更需要全面、正确的信息作为决策的支持。

### 3. 保障有效控制, 保证系统秩序

控制是保障各个社会组织有利地和高效地获得和利用其资源的监控及实施调节行为的过程。控制的整个过程离不开信息, 信息是实现有效控制的灵魂。尤其在信息化水平不断提高的今天, 各种社会成员只有在信息交流通畅及时的情况下, 依据所获得的准确信息来行使各自的职责, 才会使整个系统处于有序状态。

### 4. 发挥参考作用, 推动知识创新

信息是知识的源泉和生产发展的催化剂, 信息中包含有大量人类实践活动的成果和教训, 人们对各种客观事实和社会现象的解释、论证和总结, 比较集中地反映了人类的研究成果。因而, 信息对我们的科学研究和社会实践均有着广泛的参考作用。通过相关的信息, 可以对所选项目是否具有创新性做出判断, 避免重复选题, 并参考他人的研究方法, 加快科研的进程。

## 1.1.2.2 信息的类型

从不同层面、不同角度、不同学科领域出发, 根据不同的分类标准, 可以对信息的类型进行不同的划分。对信息的分类可以从广义信息和狭义信息两个方面来进行理解。

### 1. 基于广义信息概念的信息分类

从信息的广义内涵来划分信息的类型, 即把信息理解为对客观事物存在方式和运动状态的反映。据此有以下分类:

#### (1) 依据信息的产生, 可以分为社会信息和自然信息。

社会信息指人类在社会实践活动中, 为生存、生产和社会发展而产生、处理和利用的信息, 是人类对外界事物的反映、人的思想和情感、人与人之间的联系