

21世纪高等学校规划教材 | 信息管理与信息系统



Web信息检索系统 的设计及应用

李志义 主编

清华大学出版社

21世纪高等学校规划教材 | 信息管理与信息系统



Web信息检索系统 的设计及应用

李志义 主编

清华大学出版社
北京

内 容 简 介

本书从互联网信息组织与检索的角度,较新颖地介绍了 Web 信息检索系统的原理、主要模块、相关理论、技术以及最新发展方向。同时结合实例详细阐述了 Web 查询系统设计的方法、步骤及编程实现的全过程。本书的最大特色是在传统《信息检索》有关教材的基础上,增加了基于 .NET 平台而设计的 Web 信息检索的完整案例。全书分三篇共九章内容,理论篇介绍了 Web 信息组织的基础理论、基于内容的多媒体信息检索、搜索引擎的原理与结构等;设计篇介绍了基于 ASP.NET 的开发环境配置及 VB.NET 编程基础、Web 查询系统和搜索引擎的设计,案例设计力求由浅入深、分步推进,每个案例均附有完整的源代码;应用篇引入检索案例介绍了常用搜索引擎与 Web 数据库系统的检索与利用,使检索理论与查询实践紧密结合起来。

本书每章的正文内容后附有“本章小结”和“思考与练习”,以便加深读者对本章知识点的理解,以及培养读者应用本章学到的知识来解决实际问题的能力。

本书层次分明,深入浅出;既有深入的理论分析,也有大量的设计和检索案例,具有学习和实用双重意义,可作为高等院校信息管理与信息系统、电子商务等专业的本科生或研究生的教学参考书和教材,也可供从事网站开发、Web 挖掘、数字图书馆等应用程序开发工作的科研人员参阅。

本书封面贴有清华大学出版社防伪标签,无标签者不得销售。

版权所有,侵权必究。侵权举报电话:010-62782989 13701121933

图书在版编目(CIP)数据

Web 信息检索系统的设计及应用/李志义主编. —北京:清华大学出版社,2011.10

(21 世纪高等学校规划教材·信息管理与信息系统)

ISBN 978-7-302-26481-1

I. ①W… II. ①李… III. ①机器检索—检索系统 IV. ①G354.4

中国版本图书馆 CIP 数据核字(2011)第 166438 号

责任编辑:高买花 薛 阳

责任校对:焦丽丽

责任印制:杨 艳

出版发行:清华大学出版社

地 址:北京清华大学学研大厦 A 座

<http://www.tup.com.cn>

邮 编:100084

社 总 机:010-62770175

邮 购:010-62786544

投稿与读者服务:010-62795954, jsjic@tup.tsinghua.edu.cn

质 量 反 馈:010-62772015, zhiliang@tup.tsinghua.edu.cn

印 装 者:北京鑫海金澳胶印有限公司

经 销:全国新华书店

开 本:185×260 印 张:20.75 字 数:515 千字

版 次:2011 年 10 月第 1 版 印 次:2011 年 10 月第 1 次印刷

印 数:1~3000

定 价:34.00 元

产品编号:039927-01

编审委员会成员

(按地区排序)

清华大学

周立柱 教授
覃 征 教授
王建民 教授
冯建华 教授
刘 强 副教授

北京大学

杨冬青 教授
陈 钟 教授
陈立军 副教授

北京航空航天大学

马殿富 教授
吴超英 副教授
姚淑珍 教授

中国人民大学

王 珊 教授
孟小峰 教授
陈 红 教授

北京师范大学

周明全 教授
阮秋琦 教授
赵 宏 副教授

北京交通大学

孟庆昌 教授
杨炳儒 教授
陈 明 教授

北京信息工程学院

艾德才 教授
吴立德 教授
吴百锋 教授

北京科技大学

杨卫东 副教授
苗夺谦 教授
徐 安 教授

石油大学

邵志清 教授
杨宗源 教授
应吉康 教授

天津大学

乐嘉锦 教授
孙 莉 副教授

复旦大学

同济大学

华东理工大学

华东师范大学

东华大学

浙江大学	吴朝晖	教授
	李善平	教授
扬州大学	李 云	教授
南京大学	骆 斌	教授
	黄 强	副教授
南京航空航天大学	黄志球	教授
	秦小麟	教授
南京理工大学	张功萱	教授
南京邮电学院	朱秀昌	教授
苏州大学	王宜怀	教授
	陈建明	副教授
江苏大学	鲍可进	教授
中国矿业大学	张 艳	教授
武汉大学	何炎祥	教授
华中科技大学	刘乐善	教授
中南财经政法大学	刘腾红	教授
华中师范大学	叶俊民	教授
	郑世珏	教授
	陈 利	教授
江汉大学	颜 彬	教授
国防科技大学	赵克佳	教授
	邹北骥	教授
中南大学	刘卫国	教授
湖南大学	林亚平	教授
西安交通大学	沈钧毅	教授
	齐 勇	教授
长安大学	巨永锋	教授
哈尔滨工业大学	郭茂祖	教授
吉林大学	徐一平	教授
	毕 强	教授
山东大学	孟祥旭	教授
	郝兴伟	教授
中山大学	潘小轰	教授
厦门大学	冯少荣	教授
仰恩大学	张思民	教授
云南大学	刘惟一	教授
电子科技大学	刘乃琦	教授
	罗 蕾	教授
成都理工大学	蔡 淮	教授
	于 春	副教授
西南交通大学	曾华燊	教授

出版说明

随着我国改革开放的进一步深化,高等教育也得到了快速发展,各地高校紧密结合地方经济建设发展需要,科学运用市场调节机制,加大了使用信息科学等现代科学技术提升、改造传统学科专业的投入力度,通过教育改革合理调整和配置了教育资源,优化了传统学科专业,积极为地方经济建设输送人才,为我国经济社会的快速、健康和可持续发展以及高等教育自身的改革发展做出了巨大贡献。但是,高等教育质量还需要进一步提高以适应经济社会发展的需要,不少高校的专业设置和结构不尽合理,教师队伍整体素质亟待提高,人才培养模式、教学内容和教学方法需要进一步转变,学生的实践能力和创新精神亟待加强。

教育部一直十分重视高等教育质量工作。2007年1月,教育部下发了《关于实施高等学校本科教学质量与教学改革工程的意见》,计划实施“高等学校本科教学质量与教学改革工程”(简称“质量工程”),通过专业结构调整、课程教材建设、实践教学改革、教学团队建设等多项内容,进一步深化高等学校教学改革,提高人才培养的能力和水平,更好地满足经济社会发展对高素质人才的需要。在贯彻和落实教育部“质量工程”的过程中,各地高校发挥师资力量强、办学经验丰富、教学资源充裕等优势,对其特色专业及特色课程(群)加以规划、整理和总结,更新教学内容、改革课程体系,建设了一大批内容新、体系新、方法新、手段新的特色课程。在此基础上,经教育部相关教学指导委员会专家的指导和建议,清华大学出版社在多个领域精选各高校的特色课程,分别规划出版系列教材,以配合“质量工程”的实施,满足各高校教学质量和教学改革的需要。

为了深入贯彻落实教育部《关于加强高等学校本科教学工作,提高教学质量的若干意见》精神,紧密配合教育部已经启动的“高等学校教学质量与教学改革工程精品课程建设工作”,在有关专家、教授的倡议和有关部门的大力支持下,我们组织并成立了“清华大学出版社教材编审委员会”(以下简称“编委会”),旨在配合教育部制定精品课程教材的出版规划,讨论并实施精品课程教材的编写与出版工作。“编委会”成员皆来自全国各类高等学校教学与科研第一线的骨干教师,其中许多教师为各校相关院、系主管教学的院长或系主任。

按照教育部的要求,“编委会”一致认为,精品课程的建设工作从开始就要坚持高标准、严要求,处于一个比较高的起点上。精品课程教材应该能够反映各高校教学改革与课程建设的需要,要有特色风格、有创新性(新体系、新内容、新手段、新思路,教材的内容体系有较高的科学创新、技术创新和理念创新的含量)、先进性(对原有的学科体系有实质性的改革和发展,顺应并符合21世纪教学发展的规律,代表并引领课程发展的趋势和方向)、示范性(教材所体现的课程体系具有较广泛的辐射性和示范性)和一定的前瞻性。教材由个人申报或各校推荐(通过所在高校的“编委会”成员推荐),经“编委会”认真评审,最后由清华大学出版

社审定出版。

目前,针对计算机类和电子信息类相关专业成立了两个“编委会”,即“清华大学出版社计算机教材编审委员会”和“清华大学出版社电子信息教材编审委员会”。推出的特色精品教材包括:

(1) 21 世纪高等学校规划教材·计算机应用——高等学校各类专业,特别是非计算机专业的计算机应用类教材。

(2) 21 世纪高等学校规划教材·计算机科学与技术——高等学校计算机相关专业的教材。

(3) 21 世纪高等学校规划教材·电子信息——高等学校电子信息相关专业的教材。

(4) 21 世纪高等学校规划教材·软件工程——高等学校软件工程相关专业的教材。

(5) 21 世纪高等学校规划教材·信息管理与信息系统。

(6) 21 世纪高等学校规划教材·财经管理与应用。

(7) 21 世纪高等学校规划教材·电子商务。

(8) 21 世纪高等学校规划教材·物联网。

清华大学出版社经过三十多年的努力,在教材尤其是计算机和电子信息类专业教材出版方面树立了权威品牌,为我国的高等教育事业做出了重要贡献。清华版教材形成了技术准确、内容严谨的独特风格,这种风格将延续并反映在特色精品教材的建设中。

清华大学出版社教材编审委员会

联系人:魏江江

E-mail:wei jj@tup.tsinghua.edu.cn

前言

在近半个世纪里,图书馆学、情报学、信息管理、计算机科学等方面的专家在信息检索领域展开了大量非常有价值的研究,包括图书分类、标引语言与组织、图书编目、计算机查询系统设计、检索模型、检索系统用户接口、系统可视化、信息过滤、查询与标引语言等。近年,随着互联网技术的迅猛发展,信息检索领域不断拓展与延伸,其研究的范围远远超出了最初的文本索引和文本文档搜索的范畴。这是因为网络技术、计算机通信技术、媒体数字化等各种技术的发展,使信息与数据的形式多种多样。除了文本之外,还有图像、视频、音频、超媒体等丰富多彩的信息载体,Web 已成为人类信息和知识传播的重要媒体,网络信息检索特别是搜索引擎应运而生。现代信息检索技术业已成为当前和未来研究的重点。本书正是基于这一研究的新方向和新的检索理论而撰写。同时遵照“十二五”规划教材的要求和规范,并结合多年从事信息检索有关课程的教学经验编写了这本教材。

相比而言,已有的信息检索方面的教材,主要偏向于传统的图书馆学、情报学领域的信息查询、信息组织和检索应用,注重介绍网络信息资源组织、检索的基础知识和基本原理,注重文献信息检索工具、网络数据库、搜索引擎的检索方法和技巧等有关内容。这类教材的缺点显而易见,主要表征为知识体系大同小异,内容重复,共性多,缺少特色,重检索应用、轻系统设计与案例分析。特别是在网络信息查询系统的设计方面,没有详细介绍其程序编写与实现过程,更没有完整的设计案例。对于信息管理与信息系统专业的学生而言,这些教材尤显肤浅和单薄。很多学生反映学完这些教材后,根本学不到专业知识,即学而不专。

因此,可以说到目前为止,系统介绍网络信息检索的原理与技术实现的专业书籍和教材极少,尤其是将网络检索技术与应用结合起来介绍的教材更少。而本书从计算机科学和信息技术的角度,为信息检索提供了一个整体而系统的理论框架,并从网络信息查询的设计与具体实现方面探讨了现代信息检索,特别是以搜索引擎为核心的 Web 检索的知识体系。可以说,本书的编写填补了图书馆学、情报学在信息检索案例设计方面的空白,并使得信息检索的理论介绍、系统设计与检索应用有机地结合起来。

本书的特色也在于此,且突出表现为从当今互联网信息组织与检索的角度,较新颖地介绍了 Web 信息检索系统的原理、主要模块、相关理论与技术以及最新发展方向。同时,结合设计实例详细讲解了 Web 查询系统设计的方法、步骤、编程的具体实现的全过程。因此,本书的最大特点是在传统《信息检索》有关教材的基础上,增加了基于 .NET 开发平台的 Web 查询系统设计的完整案例。从开发环境配置及 .NET 程序的创建、调试,到 VB.NET 程序设计基础,再到 Web 查询系统和搜索引擎的设计,案例设计由简单到复杂,深入浅出,并将计算机科学与图书馆学、情报学紧密地连接起来。

全书分三部分,包括理论篇、设计篇和应用篇。

理论篇包括第 1~4 章,主要介绍与 Web 信息检索系统设计有关的基础知识和相关理论,重点讨论 Web 信息检索的概念、原理、Web 检索系统所包含的主要模块、Web 信息组织

的模式、信息检索的流程、Web 信息资源的文档类型,索引、倒排文件、二叉树、查找的算法以及与网络查询密切相关的理论。其中,我的硕士研究生梁士金同学参与了“2.4 多媒体信息与文件格式”和“2.5 基于内容的多媒体信息检索”的编写以及“第4章 Web 搜索引擎的原理与结构”的资料收集。

本篇编写力求简练新颖,期望为网络信息检索提供理论依据并建立其理论体系,另一方面也为 Web 检索系统的设计与开发作理论铺垫。

设计篇包括第5~7章,主要介绍与 Web 信息检索系统设计有关的计算机编程基础和查询程序设计。本篇的编写思路非常明确,即围绕设计案例展开介绍,以.NET 为开发平台,分步推进,循序渐进地介绍 Web 查询系统的详细设计与实现的过程,特别介绍了较复杂的搜索引擎的设计。书中所涉及的每个案例均有完整的源代码,并调试正常。其中,我的硕士研究生肖炯恩同学参与了“7.7 基于.NET 的搜索引擎设计案例”的程序编写与设计,以及本节部分内容的撰写和资料收集。

应用篇包括第8章和第9章,从实操和应用的角度介绍了网络信息检索系统的使用方法和技巧,包括搜索引擎和网络数据库系统的搜索和利用。对其中具有代表性并经常使用的检索系统进行了较详细的介绍。从编写的思路而言,先概要性地介绍并列举 Web 检索工具和检索系统,然后通过检索实例探讨 Web 检索系统的查询技巧和搜索应用,力求每个检索工具与系统的介绍都配有相应的查询案例。因此,本篇注重查询案例分析,而并不穷尽罗列 Web 查询系统的操作方法和有关介绍。在学习本篇时,期望读者能通过已有检索系统的介绍而触类旁通、举一反三。本篇中的部分章节,梁士金同学参与了资料的收集与整理。

最后,要感谢对本书的编写给予鼎力支持与帮助的所有朋友、同事。感谢华师大经管学院领导在申报“国家十二五规划教材”给予的鼓励与支持;感谢清华大学出版社对本书申报“十二五规划教材”所付出的极大努力、支持和帮助;感谢本书中所引用和参考的前沿性研究成果和资料的作者们,特别是宁夏大学张冬教授,本书中的搜索引擎案例引用了他的开源软件 XunLong 0.7 版本,在此特致深深的敬意和谢忱;还要感谢清华大学出版社编审的辛勤工作和清华大学出版社给予本书出版的大力支持;感谢我的硕士研究生梁士金、肖炯恩、钟凤贵和沈之锐,他们在本书的资料收集、整理和部分章节的编写上付出了辛勤的汗水,在此谨致谢忱。此外,我还要感谢我的爱人董志云,是她给予我极大的动力,并不时督促我努力完成本书的撰写,深深感谢她在我背后所付出的辛劳和给予的莫大的支持,我将铭记于心;感谢我所有的家人,他们的支持令我力量倍增,在此我特向他们鞠躬并致谢。

由于作者水平所限,书中难免存在错误和不当之处,恳请读者不吝赐教。我的邮箱是:leeds@scnu.edu.cn,欢迎读者与我联系,以使本书更加完善。

李志义

于华南师范大学

2011年5月

上篇 理论篇

第 1 章 Web 信息检索概论	3
1.1 Web 信息检索及相关术语的解释	3
1.1.1 Web 信息处理系统	3
1.1.2 Web 信息组织与检索	4
1.2 Web 信息检索系统的逻辑结构	5
1.2.1 信息检索系统的逻辑结构	5
1.2.2 主要模块分析	6
1.3 信息检索的流程与步骤	8
1.3.1 Web 查询流程	8
1.3.2 Web 查找的一般步骤	9
1.4 Web 信息检索研究的历史、现状及进展	10
1.4.1 国外研究历史与现状	10
1.4.2 国内研究现状	11
1.4.3 Web 信息检索研究进展与趋势	12
1.5 本章小结	15
1.6 思考与练习	15
第 2 章 Web 信息资源的文档类型及基于内容的多媒体信息检索	17
2.1 HTML 文档	17
2.1.1 HTML 概述	17
2.1.2 HTML 的编写规则	17
2.2 XML 语言	20
2.2.1 XML 概述	20
2.2.2 XML 的编写规则	20
2.3 元数据	23
2.3.1 元数据概述	23
2.3.2 各专业领域中的元数据标准	24
2.3.3 元数据开发应用的标准化框架	25
2.4 多媒体信息与文件格式	27
2.4.1 多媒体信息的特点	27

2.4.2 多媒体信息的基本属性	27
2.5 基于内容的多媒体信息检索	33
2.5.1 基于内容的多媒体信息检索的特点	33
2.5.2 CMIR 系统的结构模型	33
2.5.3 基于内容的图像信息检索方法	36
2.5.4 基于内容的音频检索方法	37
2.5.5 基于内容的视频信息检索方法	40
2.6 本章小结	41
2.7 思考与练习	42
第3章 与 Web 查询技术相关的数据结构理论——索引、查找的算法、二叉树和图的遍历	43
3.1 索引和倒排文件	43
3.1.1 数据库索引概要	43
3.1.2 索引类型	44
3.1.3 倒排文件	48
3.2 树、二叉树和图的遍历	51
3.2.1 数据结构的概念	51
3.2.2 树和二叉树	52
3.2.3 图的遍历	56
3.3 查找的几种算法	57
3.3.1 查找的概念	57
3.3.2 查找算法	58
3.4 本章小结	66
3.5 思考与练习	67
第4章 Web 搜索引擎的原理与结构	68
4.1 搜索引擎概述	68
4.1.1 搜索引擎的概念	68
4.1.2 搜索引擎的分类	69
4.1.3 搜索引擎的基本构成	71
4.1.4 搜索引擎的主要商务模式	71
4.1.5 搜索引擎的发展趋势	72
4.2 搜索引擎的体系结构	74
4.2.1 搜索引擎的基本结构	74
4.2.2 搜索引擎的结构类型	76
4.3 搜索引擎对信息的抓取与采集	78
4.3.1 采集器的工作流程	78
4.3.2 采集器的遍历策略	80

4.3.3	网络爬虫的优化策略	82
4.4	搜索引擎对抓取信息的预处理	85
4.4.1	网页预处理的一般流程	86
4.4.2	索引网页库的创建	86
4.4.3	分析网页和构建倒排文件	88
4.4.4	常见的索引文件格式	91
4.5	搜索引擎的查询服务	92
4.5.1	查询方式及匹配	92
4.5.2	检索结果的形成及排序	93
4.6	本章小结	95
4.7	思考与练习	96

中篇 设计篇

第5章 Web 查询系统设计基础(一)——开发环境配置及.NET程序的创建、调试

5.1	Web 查询程序的开发平台——.NET 概述	99
5.1.1	VS.NET 与 .NET Framework	99
5.1.2	ASP.NET 的特点	100
5.1.3	ASP.NET 的执行过程	101
5.2	Web 查询程序的开发环境——.NET 的配置	102
5.2.1	IIS 的安装	102
5.2.2	Visual Studio 2008 的安装	103
5.3	利用 VS.NET 创建 ASP.NET 程序	105
5.3.1	使用 VS2008 创建 .NET 程序	105
5.3.2	.NET 程序的调试方法	109
5.4	ASP.NET 的程序结构及编写的几种方式	112
5.4.1	.NET 的程序结构	112
5.4.2	.NET 程序的几种编写方式	113
5.5	本章小结	116
5.6	思考与练习	116

第6章 Web 查询系统设计基础(二)——VB.NET 程序设计基础

6.1	VB.NET 编程的基本规则	117
6.1.1	程序元素命名的一般规则	117
6.1.2	VB.NET 的注释	118
6.1.3	程序的换行及其他	118
6.2	VB.NET 的数据类型及常量、变量	119
6.2.1	VB.NET 数据类型	119
6.2.2	VB.NET 的常量	120

6.2.3	VB.NET 的变量声明	122
6.2.4	数组	123
6.3	VB.NET 的运算符	126
6.3.1	算术与字符串连接运算符	126
6.3.2	比较运算符	127
6.3.3	逻辑运算符	127
6.3.4	赋值运算符	128
6.4	VB.NET 的条件与循环语句	129
6.4.1	If 语句	129
6.4.2	Select Case 语句	132
6.4.3	For/Next 循环语句	134
6.4.4	Do/While/Loop 和 Do/Until/Loop	136
6.4.5	While/EndWhile 循环语句	137
6.5	VB.NET 的过程和函数	139
6.5.1	Sub 过程	139
6.5.2	函数	140
6.5.3	变量的有效范围与参数的传递	142
6.5.4	系统函数	144
6.6	面向对象编程基础	147
6.6.1	对象概述	147
6.6.2	类	150
6.6.3	继承与封装	151
6.7	本章小结	152
6.8	思考与练习	152
第 7 章	Web 查询系统的设计	155
7.1	Access、SQL Server 2005 等数据库和表的创建与维护	155
7.1.1	Access 数据库的创建与表的操作	155
7.1.2	SQL Server 2005 数据库的创建与表的操作	157
7.1.3	SQL Server 2005 中的数据导入	161
7.2	ADO.NET 访问数据库概述	164
7.2.1	ADO.NET 的对象模型	164
7.2.2	ADO.NET 的命名空间	164
7.2.3	ADO.NET 的访问对象	165
7.2.4	对数据库的操作步骤	166
7.3	ADO.NET 操作数据库的方式	167
7.3.1	使用 SQL 命令对数据库进行访问	167
7.3.2	使用 DataAdapter 对数据库进行操作	171
7.3.3	对 SQL Server 数据库进行访问	173

7.4	基于.NET的查询程序设计实例	174
7.4.1	支持精确查询的实例	174
7.4.2	支持模糊查询的实例	177
7.4.3	利用数据源控件设计简单的查询系统	179
7.5	对查询结果的处理	182
7.5.1	分页显示结果	182
7.5.2	设置超链接以便浏览详情或相关资料	185
7.6	综合设计实例	188
7.6.1	创建母版页	188
7.6.2	使用母版页的几种方式	193
7.6.3	查询系统界面的设计及应用程序的编写	193
7.7	基于.NET的搜索引擎设计案例	197
7.7.1	案例简介	197
7.7.2	环境搭建与配置	200
7.7.3	系统设计的实现	205
7.8	本章小结	218
7.9	思考与练习	218

下篇 应用篇

第8章	搜索引擎的查找与利用	223
8.1	水平搜索引擎的查找与利用——百度搜索	223
8.1.1	百度概述	224
8.1.2	百度的基本搜索	225
8.1.3	百度的高级搜索	227
8.1.4	搜索结果说明	229
8.1.5	百度搜索引擎的搜索技巧和语法运用	229
8.2	水平搜索引擎的查找与利用——Google	234
8.2.1	Google的特色	234
8.2.2	Google的基本搜索	235
8.2.3	Google的高级搜索	236
8.2.4	Google的检索结果说明	239
8.3	水平搜索引擎的查找与利用——雅虎搜索	240
8.3.1	雅虎搜索的特色	240
8.3.2	雅虎的基本搜索	242
8.3.3	搜索案例	243
8.4	其他水平搜索引擎的查找与利用	244

8.4.1	搜狗	244
8.4.2	Bing(必应)	247
8.4.3	SOSO 搜搜	247
8.4.4	网易有道	248
8.4.5	其他搜索引擎	249
8.5	垂直搜索引擎	250
8.5.1	酷讯搜索	251
8.5.2	搜房网	253
8.5.3	其他垂直搜索引擎	255
8.6	元搜索引擎	260
8.6.1	Dogpile	261
8.6.2	其他英文元搜索引擎	263
8.6.3	国内的中文元搜索引擎	264
8.7	本章小结	265
8.8	思考与练习	265
第9章	Web 数据库系统的检索与应用	266
9.1	中文论文数据库的检索与利用——中国知识资源总库	266
9.1.1	中国知识资源总库简介	266
9.1.2	中国期刊网的检索与利用	268
9.2	中文论文数据库的检索与利用——其他论文数据库概要	275
9.2.1	《中文科技期刊数据库》	275
9.2.2	人大复印报刊资料数据库	277
9.2.3	万方数据的学术论文数据库	278
9.2.4	《全国报刊索引数据库》数据库	279
9.2.5	其他	280
9.3	外文论文数据库的检索与利用——ProQuest 数据库的检索与利用	284
9.3.1	ProQuest 简介	284
9.3.2	PQDT 概述	285
9.3.3	PQDT 的检索操作	285
9.3.4	ProQuest 旗下的 UMI 数据库检索	289
9.4	外文论文数据库的检索与利用——常用外文论文数据库检索概要	290
9.4.1	ISI Web of Knowledge 数据库	290
9.4.2	Elsevier ScienceDirect(学术期刊全文库)	293
9.4.3	Springer Online Journals(施普林格在线期刊)	294
9.4.4	EBSCOhost 数据库系统	296
9.4.5	OCLC FirstSearch 数据库检索	297

9.4.6 其他知名外文数据库检索概览.....	301
9.5 图书、报纸、会议文献、专利信息等有关资料 and 数据的检索.....	303
9.5.1 图书类数据库的检索概要.....	303
9.5.2 报纸类数据库的检索概要.....	304
9.5.3 会议文献数据库检索概要.....	305
9.5.4 专利信息数据库查找概要.....	307
9.6 本章小结	309
9.7 思考与练习	309
参考文献.....	311

上篇

理论篇

本篇主要介绍 Web 信息检索系统设计有关的基础知识和相关理论,重点讨论 Web 信息检索的原理、Web 检索系统所包含的主要模块、Web 信息组织的模式、信息检索的流程、Web 信息资源的文档类型,索引、倒排文件、二叉树、查找的算法以及与 Web 查询密切相关的理论。这样,一方面可为 Web 信息检索提供理论依据并建立其理论体系,另一方面也为 Web 检索系统的设计与开发作铺垫。