

<<Web信息检索系统的设计及应用>>

图书基本信息

书名：<<Web信息检索系统的设计及应用>>

13位ISBN编号：9787302264811

10位ISBN编号：7302264813

出版时间：2011-10

出版时间：李志义 清华大学出版社 (2011-10出版)

作者：李志义 编

页数：313

版权说明：本站所提供下载的PDF图书仅提供预览和简介，请支持正版图书。

更多资源请访问：<http://www.tushu007.com>

<<Web信息检索系统的设计及应用>>

内容概要

《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》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。

《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》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。

《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》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。

《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》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。

书籍目录

上篇理论篇第1章 web信息检索概论1.1 web信息检索及相关术语的解释1.2 web信息检索系统的逻辑结构1.3 信息检索的流程与步骤1.4 web信息检索研究的历史、现状及进展1.5 本章小结1.6 思考与练习第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2.1 html文档2.2 xml语言2.3 元数据2.4 多媒体信息与文件格式2.5 基于内容的多媒体信息检索2.6 本章小结2.7 思考与练习第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3.1 索引和倒排文件3.2 树、二叉树和图的遍历3.3 查找的几种算法3.4 本章小结3.5 思考与练习第4章 web搜索引擎的原理与结构4.1 搜索引擎概述4.2 搜索引擎的体系结构4.3 搜索引擎对信息的抓取与采集4.4 搜索引擎对抓取信息的预处理4.5 搜索引擎的查询服务4.6 本章小结4.7 思考与练习中篇设计篇第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5.2 web查询程序的开发环境——.net的配置5.3 利用vs.net创建asp.net程序5.4 asp.net的程序结构及编写的几种方式5.5 本章小结5.6 思考与练习第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6.2 vb.net的数据类型及常量、变量6.3 vb.net的运算符6.4 vb.net的条件与循环语句6.5 vb.net的过程和函数6.6 面向对象编程基础6.7 本章小结6.8 思考与练习第7章 web查询系统的设计7.1 access、sqlserver2005等数据库和表的创建与维护7.2 ado.net访问数据库概述7.3 ado.net操作数据库的方式7.4 基于.net的查询程序设计实例7.5 对查询结果的处理7.6 综合设计实例7.7 基于.net的搜索引擎设计案例7.8 本章小结7.9 思考与练习下篇篇应用篇第8章 搜索引擎的查找与利用8.1 水平搜索引擎的查找与利用——百度搜索8.2 水平搜索引擎的查找与利用——google8.3 水平搜索引擎的查找与利用——雅虎搜索8.4 其他水平搜索引擎的查找与利用8.5 垂直搜索引擎8.6 元搜索引擎8.7 本章小结8.8 思考与练习第9章 web数据库系统的检索与应用9.1 中文论文数据库的检索与利用——中国知识资源总库9.2 中文论文数据库的检索与利用——其他论文数据库概要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9.5 图书、报纸、会议文献、专利信息等有关资料和数据检索9.6 本章小结9.7 思考与练习参考文献

章节摘录

版权页：插图：进一步而言，分布式搜索引擎的主要特点还有：（1）各检索服务器之间协同工作，每个服务器只搜索自身区域内的信息资源，彼此之间只传递搜索结果信息，加快了检索速度，减轻网络及主搜索引擎服务器的负担。

（2）与网络资源本身的分布式特性相适应，增加搜索服务器方便，具有可扩展性。

（3）索引信息划分到各个数据库中，使得各索引数据库规模小，易于管理，缩短查询响应时间。

此外，分布式搜索引擎的分布性主要体现在三个方面：Web信息搜集的分布性，主要是指多个采集器的分布协同工作；索引器的分布性，即对搜集到的Web信息以分布式的方式进行索引文件的构建、存储、组织和管理；检索的分布性，是指搜索引擎能借助相关的并发控制方法和技术快速响应分布在不同区域内的大量网络用户的同时访问。

搜索引擎对信息的抓取与采集在宏观上讨论搜索引擎体系结构的基础上，本节开始重点介绍搜索引擎对Web资源的采集算法和与此相关的基本知识。

通常，搜索引擎可以采用批量搜集和增量搜集两种策略获取Web信息。

所谓批量搜集是指搜索引擎每隔一段时间对Web信息进行一次全面的搜索，并用所搜集到的信息覆盖前一次搜集的信息。

而增量搜集基于网络中的大部分信息往往不经常更新的考虑，在每次搜索信息时，只需要搜集那些发生变化的或者新出现的站点信息即可，因此可以明显提高系统的搜索效率和降低资源消耗。

相比而言，增量搜集具有更大的适应能力和信息更新能力。

目前，Web信息采集的方式主要有手工和机器人、网络蜘蛛等自动抓取两种方式。

人工方式准确率高但技术含量低、工作量大，本节不予探讨，而注重讨论的是自动采集的方式。

由于自动采集器软件有多种名称和称呼，比如机器人（Robert）程序、Spider（网络蜘蛛）程序或Crawler（网络爬虫）程序，本节统称为采集器或网络爬虫。

4.3.1 采集器的工作流程 采集器的实质是一种用于从互联网上抓取Web信息的软件程序，它以Web页之间的相互链接为路径，依照一定的遍历策略试图爬行网络空间的每个角落，采集所需信息，并将其转存到本地存储器上，以便索引器进行标引处理和创建索引库。

采集器具有如下的特点：程序具备超强的执行力；在分析Web页和利用链接进行爬行等方面具有自动性和智能性；能自动将Web信息实施简单的存储操作等。

此外，采集器还具有可伸缩性、分布式、可定制性等特点。

<<Web信息检索系统的设计及应用>>

编辑推荐

《web信息检索系统的设计及应用》教学目标明确，注重理论与实践的结合，教学方法灵活，培养学生自主学习的能力，教学内容先进，满足专业人才的社会需求，教学模式完善，提供配套的教学资源解决方案。

版权说明

本站所提供下载的PDF图书仅提供预览和简介，请支持正版图书。

更多资源请访问:<http://www.tushu007.com>