

<<搜索引擎>>

图书基本信息

书名：<<搜索引擎>>

13位ISBN编号：9787030146335

10位ISBN编号：7030146336

出版时间：2005-4

出版时间：科学出版社发行部

作者：李晓明

页数：247

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>

内容概要

本书系统地介绍了互联网搜索引擎的工作原理、实现技术及其系统构建方案。

全书分三篇共13章内容，从基本工作原理概述，到一个小型简单搜索引擎具体细节的实现，进而详细讨论了大规模分布式搜索引擎系统的设计要点及其关键技术；最后介绍了面向主题和个性化的web信息服务，阐述了中文网页自动分类等技术及其应用。

本书层次分明，由浅入深；既有深入的理论分析，也有大量的实验数据，具有学习和实用双重意义。

本书可作为高等院校计算机科学与技术、信息管理与信息系统、电子商务等专业的研究生或高年级本科生的教学参考书和技术资料，对广大从事网络技术、Web站点的管理、数字图书馆、Web挖掘等研究和应用开发的科技人员也有很高的参考价值。

<<搜索引擎>>

作者简介

李晓明：天网搜索引擎领域负责人
闫宏飞 王继民：天网搜索引擎项目负责人

书籍目录

前言第一章 引论 第一节 搜索引擎的概念 第二节 搜索引擎的发展历史 第三节 一些著名的搜索引擎上篇 Web搜索引擎基本原理和技术 第二章 Web搜索引擎工作原理和体系结构 第一节 基本要求 第二节 网页搜集 第三节 预处理 第四节 查询服务 第五节 体系结构 第三章 Web信息的搜集 第一节 引言 一、超文本传输协议 二、一个小型搜索引擎系统 第二节 网页搜集 一、定义URL类和Page类 二、与服务器建立连接 三、发送请求和接收数据 四、网页信息存储的天网格式 第三节 多道搜集程序并行工作 一、多线程并发工作 二、控制对一个站点并发搜集线程的数目 第四节 如何避免网页的重复搜集 一、记录未访问、已访问URL和网页内容摘要信息 二、域名与IP的对应问题 第五节 如何首先搜集重要的网页 第六节 搜集信息的类型 第七节 本章小结 第四章 对搜集信息的预处理 第一节 信息预处理的系统结构- 第二节 索引网页库 第三节 中文自动分词 第四节 分析网页和建立倒排文件 第五节 本章小结 第五章 信息查询服务 第一节 查询服务的系统结构 第二节 检索的定义 第三节 查询服务的实现 一、结果集合的形成 二、查询结果显示 第四节 本章小结中篇 对质量和性能的追求 第六章 可扩展搜集子系统 第一节 天网系统概述和集中式搜集系统结构 一、天网系统结构 二、集中式搜集系统 第二节 利用并行处理技术高效搜集网页的一种方案 一、节点间URL的划分策略 二、关于性能的讨论 三、性能测试和评价 四、系统的动态可配置性设计 第三节 本章小结 第七章 网页净化与消重 第一节 网页净化与元数据提取 一、引言 二、Docview模型 三、网页的表示 四、提取Docview模型要素的方法 五、模型应用及实验研究 第二节 网页消重算法 一、消重算法..... 第八章 高性能检索子系统 第九章 用户行为的特征及缓存的应用 第十章 相关排序与系统质量评估下篇 面向主题和个性化的Web信息服务 第十一章 中文网页自动分类技术 第十二章 搜索引擎个性化查询服务 第十三章 面向主题的信息搜集与应用参考文献附录 术语后记

<<搜索引擎>>

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>