

<<搜索引擎原理、实践与应用>>

图书基本信息

书名：<<搜索引擎原理、实践与应用>>

13位ISBN编号：9787121048104

10位ISBN编号：7121048108

出版时间：2007-9

出版时间：第1版 (2007年9月1日)

作者：卢亮

页数：289

字数：210000

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>

<<搜索引擎原理、实践与应用>>

内容概要

本书是搜索引擎业界资深的搜索引擎专家长久以来经验的积累与研究工作的心路历程。

本书从搜索引擎的历史和现状开始展开，为广大读者展现了搜索引擎简单界面背后的复杂技术、原理和应用，从数据挖掘到搜索引擎的数据结构，从搜索引擎爬虫到分布式搜索引擎的设计均是作者精心研究的结果和过程，对研究搜索引擎的人士颇有实用价值和参考意义。

本书适合刚搜索引擎行业乃至互联网行业的从业人员、如网站设计者、程序员、个人网站的站主，本书还适合大中学院相关专业及培训机构的相关培训作为参考书籍。

<<搜索引擎原理、实践与应用>>

书籍目录

第1章 搜索引擎的历史和现状 1.1 搜索引擎的历史 1.2 搜索引擎的分类 1.3 搜索引擎的现状 1.4 世界主要搜索引擎介绍 1.4.1 Google 1.4.2 百度 1.4.3 Technorati 1.4.4 A11theweb 1.4.5 Ask.com 1.4.6 MSN Search 小结 参考文献第2章 数据挖掘 2.1 数据挖掘概览 2.1.1 为什么要做数据挖掘 2.1.2 数据挖掘的任务 2.1.3 关联挖掘和分析 2.1.4 多子类 2.1.5 聚类分析 2.1.6 序列模式分析 2.1.7 偏差分析 2.2 数据挖掘的常用技术 2.2.1 人工神经网络 2.2.2 统计分析 小结 参考文献第3章 搜索引擎的数据结构 3.1 存储结构 3.1.1 四种基本存储方法 3.2 索引 3.2.1 倒排索引 3.3 结构化查询语言 3.4 海量数据系统 小结 参考文献第4章 搜索引擎的基本结构 4.1 网络爬虫 4.2 排序 4.2.1 网页的权重 4.2.2 向量空间模型VSM 4.2.3 扩展相关信息 4.3 索引系统 4.4 缓存机制 小结 参考文献第5章 搜索引擎爬虫 5.1 深度优先与广度优先 5.1.1 网页链接情况概览 5.1.2 取得有效的网页文字第6章 搜索引擎索引系统第7章 分布式搜索引擎设计第8章 Google搜索引擎的结构第9章 中文分词第10章 分类与聚类第11章 内容消重和SPAM消除第12章 图像搜索引擎附录

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>