

图书基本信息

书名：<<网络信息检索技术及搜索引擎系统开发>>

13位ISBN编号：9787030261434

10位ISBN编号：7030261437

出版时间：2010-2

出版时间：科学出版社

作者：高凯，郭立炜，许云峰 编著

页数：243

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

本书较系统地论述了网络信息检索技术的基本原理，并进一步阐述了其在搜索引擎系统开发及其智能化实现中的应用。

本书分为三部分。

第一部分是基础知识和相关背景介绍部分，包括从第1章到第3章的内容。

其中，第1章概要介绍了信息检索的起源和发展、信息检索模型及方法、网络信息检索的过程、网络信息检索性能评价、网络信息智能化处理、网络信息检索技术的未来发展等问题；第2章简要介绍了Web信息下载、页面分析与信息抽取方法、基于链接分析的网页相关性算法、检索结果排序、自然语言处理等问题；第3章就搜索引擎的发展、分类、功能、资源等进行了介绍。

第二部分为利用开源资源实现搜索引擎系统的部分，包括从第4章到第8

章的内容，主要介绍如何利用Lucene等开源资源来构建自己的搜索引擎。

作为开源项目中的一朵奇葩，Lucene提供了强大的全文索引和检索功能，并在搜索引擎、桌面检索系统、网站站内搜索、企业级内部文档管理与检索、情报分析系统、知识管理系统、数字图书馆检索系统中发挥了很好的作用。

本书第4章介绍了Lucene的索引与检索机制及其应用、开发平台的搭建与配置等；第5章介绍Lucene中的中英文分词处理及其效果；第6章介绍检索结果排序及处理技术；第7章介绍如何利用开源资源来获取网络信息；第8章介绍如何对常见格式的网络资源进行解析与预处理。

第三部分为搜索引擎智能化的研究与实现部分，包括从第9章到第14章的内容。

其中，第9

章对信息采集中的网页去重与相关网页聚类进行了研究；第10章讨论了信息的动态采集与更新策略，以期搜索引擎能根据网站及其更新速度的不同，动态调整其信息采集与更新的频度；第11章则是面向自然语言提问的理解与处理，提供面向大众的支持自然语言提问的智能检索接口不仅能使人机交互更加人性化，还能促进搜索引擎的应用普及；第12章则给出一种参照多数用户在检索类似问题时的经验，为用户提供一些关联性和扩展性的相关概念反馈的方法；第13章给出一种相近检索与检索结果排序方法；第

14章阐述了一种基于用户浏览兴趣的网页预取策略。

全书理论联系实际，涉及面广，体系完整，内容新颖，条理清晰，组织合理，图例丰富，说明详细，既可作为高等院校计算机应用技术专业和图书馆等相关专业的教材，也可作为工程技术人员的参考资料。

书籍目录

- 第1章 绪论
- 第2章 网络信息处理
- 第3章 搜索引擎
- 第4章 Lucene的索引与检索机制及其应用
- 第5章 分词处理
- 第6章 检索结果排序及处理
- 第7章 网络信息获取
- 第8章 网络信息预处理
- 第9章 信息采集中的网页去重与相似网页聚类
- 第10章 信息的动态采集与更新
- 第11章 面向自然语言提问的理解与处理
- 第12章 相关概念反馈
- 第13章 相近检索与检索结果排序
- 第14章 基于用户浏览兴趣的网页预取



编辑推荐

《网络信息检索技术及搜索引擎系统开发》是由科学出版社出版的。

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>