<<开发自己的搜索引擎--Lucen>>

图书基本信息

书名：<<开发自己的搜索引擎--Lucene 2.0+Heriterx>>

13位ISBN编号：9787115160003

10位ISBN编号：7115160007

出版时间：2007-6

出版时间：人民邮电出版社

作者：邱哲,符滔滔

页数：521

字数：662000

<<开发自己的搜索引擎--Lucen>>

内容概要

本书是一本针对搜索引擎开发的书籍。
通过学习本书，读者可以独立构建出一个企业级的搜索引擎网站。
本书详细讲解了搜索引擎与信息检索基础，Lucene入门实例，Lucene索引的建立，使用Lucene进行搜索，排序，过滤和分页，Lucene的分析器，对Word、Excel和PDF格式文档的处理，Compass搜索引擎框架，Lucene分布式和Google Search API，爬虫Heritrix，HTMLParser，DWR等内容。
最后综合使用所讲述的技术，构建了一个典型的垂直搜索系统，该系统具有很强的商业实用价值。

本书是一本介绍如何使用Lucene 2.0和Heritrix来构建搜索引擎的书。
通过对相关API和源代码的分析，力求使读者在掌握应用的基础上能够深入其核心，自行扩展和开发相应组件，开发出更有创意的搜索引擎产品。

本书适合从事计算机软件开发的人员阅读，同时也可以作为搜索引擎爱好者的入门书籍。
阅读本书需要具备Java语言基础。

作者简介

邱哲，北京理工大学硕士，现为某公司技术经理，主要从事欧美软件外包开发。
在J2EE方面有4年的开发经验，在搜索引擎与“爬虫”方面有3年的开发经验，著有《征服Ajax+Lucene构建搜索引擎》一书。

<<开发自己的搜索引擎--Lucen>>

书籍目录

版权说明