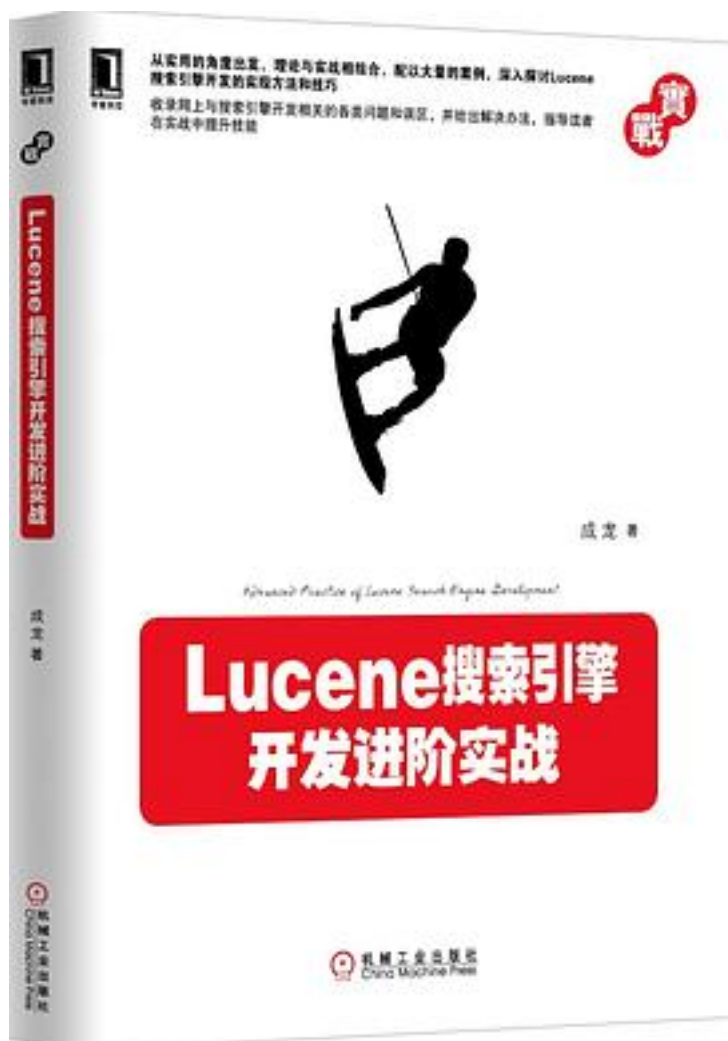


Lucene搜索引擎开发进阶实战



[Lucene搜索引擎开发进阶实战 下载链接1](#)

著者:成龙

出版者:机械工业出版社

出版时间:2015-1

装帧:平装

isbn:9787111488422

【编辑推荐】

从实用的角度出发，理论与实战相结合，配以大量的案例，深入探讨Lucene搜索引擎开发的实现方法和技巧

收录网上与搜索引擎开发相关的各类问题和误区，并给出解决办法，指导读者在实战中提升技能

【内容简介】

《Lucene搜索引擎开发进阶实战》结合笔者的实际开发经验，总结了一些新的开发技巧和开发思路，并对网上流传的一些错误技术点和代码进行验证，同时给出正确的思路，旨在给开发者提供一本清晰、完整、易懂的指导手册。本书既可为零起点的Lucene初学者提供系统全面的学习指导，也可帮助有相关经验的开发者解决在开发过程中遇到的一些难题和疑惑。

《Lucene搜索引擎开发进阶实战》共12章，分为四部分，第一部分首先介绍网络爬虫和Web搜索，然后讲述Lucene的概念和架构；第二部分通过一个小实例，并结合为文本创建索引来讲解其中的一些方法和类；第三部分主要基于数据库搜索创建搜索引擎实例，阐述技术疑难问题，讨论如何建立工程类的索引，如何使用分词技术等，并对创建索引过程中的jar包进行解析，然后介绍搜索引擎开发中关键的查询方法和高亮显示技术，以及查询结果排序和词频统计的相关知识，最后概述性能优化（包括代码优化、索引优化以及备份和恢复）等相关知识；第四部分总结目前的一些技术，并对未来的一些技术发展进行展望。

作者介绍:

成 龙

软件开发工程师，从事Lucene相关搜索引擎开发多年，曾在医药、教育等行业参与开发多个搜索引擎类项目，目前在进行一个搜索引擎项目的优化和二次开发工作，具有丰富的搜索引擎方向项目开发经验。工作之余也喜欢钻研疑难问题，并在实践验证后形成文档或经验与读者分享。

目录: 前 言

第1章 网络爬虫策略 1

1.1 信息时代的飞跃 1

1.1.1 搜索引擎的出现 1

1.1.2 搜索之网络爬虫的由来 4

1.2 网络爬虫 6

1.2.1 网络爬虫的基础框架 6

1.2.2 网络爬虫的策略分析 8

1.2.3 暗网获取 11

1.2.4 分布式爬虫 12

1.3 实现Web搜索 13

1.3.1 Web搜索的概念 13

1.3.2 经典小实例展示 13

1.4 疑难解析 18

1.4.1 本节技术概念 18

1.4.2 实例疑难解析 19

1.5 实践心得 19

1.5.1 如何快速实现Web搜索 19

1.5.2 如何解决和发现代码错误 20

1.6 本章小结	21
第2章 搜索引擎技术中的Lucene	22
2.1 Lucene到底是什么	22
2.1.1 Lucene的由来	22
2.1.2 Lucene的概念	23
2.1.3 Lucene的适用范围	23
2.2 Lucene的架构	24
2.2.1 Lucene=“完整的搜索程序”吗	26
2.2.2 搜索和索引组件介绍	27
2.2.3 其他模块	28
2.2.4 Lucene与应用的完美结合	29
2.3 Lucene小程序	29
2.3.1 创建索引	29
2.3.2 搜索索引	34
2.4 实践心得	35
2.4.1 实现创建和搜索的技术解析	36
2.4.2 实例创建过程中的个人心得	36
2.5 本章小结	36
第3章 创建索引实例	37
3.1 开发前的软硬件准备	37
3.1.1 开发语言和专业基础	37
3.1.2 开发环境基础	38
3.2 对文本文档进行索引	41
3.2.1 新建文本文档	41
3.2.2 基础的索引代码	43
3.2.3 索引结果	43
3.3 实例中用到的类和关键词	44
3.3.1 实例中涉及的类	44
3.3.2 实例中涉及的关键词	46
3.4 本章技术要点和关键点	51
3.4.1 本章技术要点	51
3.4.2 本章关键点	52
3.4.3 技术突破点	52
3.5 开发中的个人心得分享	53
3.6 本章小结	53
第4章 初建索引	54
4.1 建立数据库索引的前提和意义	54
4.1.1 建立数据库索引的前提	54
4.1.2 建立数据库索引的基本条件	55
4.1.3 建立数据库索引的意义	57
4.2 建立数据库索引实例	58
4.2.1 新建索引类	58
4.2.2 实例核心代码示例	66
4.3 实例中重要的工具：Luke	70
4.3.1 什么是Luke	70
4.3.2 Luke的作用	70
4.4 解决疑难问题的个人心得	71
4.4.1 多样化实例的参考	72
4.4.2 案例分析	74
4.5 SQL Server 2005、SQL Server 2008以及Oracle数据库的区别	75
4.6 本章小结	76
第5章 分词技术详解	77
5.1 分词的定义和意义	77
5.2 英文分词介绍	78

5.3 中文分词解析	80
5.3.1 中文分词的目的	80
5.3.2 中文分词的意义	80
5.3.3 中文分词的方法	82
5.4 实例的分词方法剖析	86
5.4.1 IKAnalyzer的诞生	86
5.4.2 IKAnalyzer的配置方法	86
5.4.3 IKAnalyzer的使用	88
5.5 分词技术的广泛应用	89
5.5.1 搜索引擎网站的分词技术应用	90
5.5.2 分词技术深入各个领域	91
5.6 实践见解	92
5.6.1 关于分词的问题	92
5.6.2 关于搜索引擎分词和查找的个人心得	93
5.7 本章小结	94
第6章 jar包应用详解	95
6.1 jar包的定义	95
6.2 基本的jar包介绍	96
6.2.1 连接SQL Server的jar包	97
6.2.2 Lucene常用的jar包	98
6.3 实例中的分词jar包IKAnalyzer3.2.8	98
6.3.1 概念	98
6.3.2 适用范围和基本配置要求	99
6.3.3 使用案例分析	100
6.4 实例中的其他jar包应用	102
6.4.1 实例包含的其他jar包	102
6.4.2 适用范围和案例分析	103
6.5 jar包的调用方法	103
6.6 扩展：如何生成jar包和运行jar包	104
6.7 实践分享	106
6.7.1 如何在自己的项目中灵活运用jar包	106
6.7.2 jar的路径问题	107
6.7.3 jar包中class文件的反编译	108
6.8 本章小结	109
第7章 构建应用程序的实现	110
7.1 建立实例的项目	110
7.1.1 src目录	110
7.1.2 webroot目录	116
7.2 详解扩展词典和停止词	121
7.2.1 为什么需要扩展词典和停止词	121
7.2.2 扩展词典的应用	122
7.2.3 停止词应用	122
7.3 应用程序测试	124
7.3.1 应用程序在MyEclipse下的调试	124
7.3.2 搜索结果的分页技术	125
7.3.3 实现界面搜索	129
7.4 开发过程中的疑难问题分享	129
7.4.1 停止词的准确应用	130
7.4.2 扩展词典问题	130
7.4.3 界面分页显示的实例分析	131
7.4.4 localhost与127.0.0.1的异同	131
7.4.5 src目录和webroot目录	132
7.5 本章小结	132
第8章 查询方法的实现	133

8.1 查询的概念和意义	133
8.1.1 搜索引擎实现查询的概念	133
8.1.2 搜索引擎查询的意义	136
8.2 查询的分类	136
8.2.1 完全匹配查询	136
8.2.2 模糊查询	137
8.2.3 多域查询	138
8.2.4 通配符查询	140
8.2.5 其他查询	140
8.2.6 组合查询	142
8.3 实例分析	142
8.3.1 完全匹配查询实例解析	142
8.3.2 模糊查询实例分析	143
8.3.3 多域查询实例分析	144
8.3.4 通配符查询实例分析	145
8.3.5 其他查询实例及分析	146
8.3.6 组合查询实例及分析	146
8.4 实践过程中的重难点解析	148
8.4.1 使用什么查询方法	148
8.4.2 模糊查询案例剖析	149
8.4.3 Occur.MUST与Occur.SHOULD	150
8.5 本章小结	150
第9章 高亮显示技术	151
9.1 高亮显示的概念和目的	151
9.1.1 高亮显示的概念	151
9.1.2 高亮显示的目的	152
9.2 高亮显示的模块介绍	152
9.2.1 高亮显示的步骤	152
9.2.2 高亮显示的重要模块	154
9.2.3 其他相关内容	155
9.3 高亮实现搜索结果	156
9.3.1 高亮显示的实例	156
9.3.2 高亮显示的搜索结果	159
9.3.3 高亮显示界面效果	161
9.4 高亮显示的应用	162
9.5 高亮显示的技术疑难分析	163
9.5.1 如何解决高亮查询结果显示不完全的问题	163
9.5.2 高亮显示的对应问题解析	164
9.6 本章小结	165
第10章 结果排序和词频统计	166
10.1 排序和词频统计概念	166
10.1.1 什么是搜索结果排序	166
10.1.2 搜索结果排序的意义	167
10.1.3 什么是词频统计	168
10.1.4 词频统计的意义何在	168
10.2 排序分类	169
10.2.1 根据域值排序	169
10.2.2 索引顺序排序	170
10.2.3 根据相关性排序	170
10.2.4 根据词频率排序	171
10.2.5 其他	172
10.3 实例分析	172
10.3.1 根据域值排序的实例解析	172
10.3.2 根据索引顺序排序的实例分析	173

- 10.3.3 根据相关性排序的实例分析 173
- 10.3.4 根据词频排序实例分析 175
- 10.3.5 其他排序方法实例分析 177
- 10.4 实践心得 178
 - 10.4.1 查询结果排序的问题 178
 - 10.4.2 关于词性处理的问题解析 181
- 10.5 本章小结 183
- 第11章 开发中的性能优化概述 184
 - 11.1 代码的优化 184
 - 11.1.1 什么是代码优化 184
 - 11.1.2 代码优化的意义 185
 - 11.1.3 如何实现代码的优化 185
 - 11.2 索引优化 189
 - 11.2.1 索引优化的目的 189
 - 11.2.2 索引优化的方法和途径 189
 - 11.2.3 索引优化的效果 192
 - 11.3 索引的备份和错误修复 192
 - 11.3.1 如何实现索引备份 193
 - 11.3.2 恢复索引的实现 193
 - 11.3.3 修复索引 194
 - 11.4 本章小结 195
- 第12章 对瓶颈技术的未来设想 196
 - 12.1 海量数据库资源搜索 196
 - 12.1.1 什么是海量数据库 196
 - 12.1.2 海量数据库资源搜索的现状和瓶颈 197
 - 12.1.3 海量数据库搜索的未来设想 197
 - 12.2 高亮显示查询结果的未来设想 199
 - 12.2.1 高亮显示出现的意义 199
 - 12.2.2 高亮显示目前现状 199
 - 12.2.3 黄褪技术概述 200
 - 12.2.4 对高亮显示查询结果的未来展望 200
 - 12.3 搜索引擎开发的规范性约束 201
 - 12.3.1 搜索引擎开发的现状 201
 - 12.3.2 版本的控制 202
 - 12.3.3 未来国际规范性约束的展望 202
 - 12.4 本章小结 202
- 附录A Java安装详解 203
- 附录B SQL Server 2005安装详解 207
- 附录C MyEclipse 安装详解 213
- 附录D Lukeall在本书中的应用 215
- 附录E MyEclipse快捷键详解 218
- • • • • ([收起](#))

[Lucene搜索引擎开发进阶实战_下载链接1](#)

标签

lucene

搜索引擎

程序设计

总体还不错

P

评论

很差的一本书，有10%有用的东西，其他都是网上随手可以找到的东西，质量还比网上的精品内容差很多。

看起来像是初学者写的书，不过简单入门还可以

没看完，太烂了。
1.代码重复太多，里面的代码基本就相当于helloworld级别，前几章的代码直接从lucene官网拿的。 2.内容非常浅，与进阶不符合 3.15年出的书用的居然是lucene3。
4.不提供源代码下载，书中要操作的数据库不给脚本。

v3.3

妈蛋

在城际上看了这本，对于我这个搞过一阵es的来说，也就有10页有用。把进阶实战改成入门吧。。现在书都这么好写了么？

[Lucene搜索引擎开发进阶实战_下载链接1](#)

书评

有这么写书的嘛？你书名可是进阶，你这什么内容？明明就是入门，网上都能找到的东西。你能写点有深度的嘛？源码你都没解析过，好意思说进阶？都是些最基础的demo，现在出版社都这么不负责任，不知道你这书名是不是属于误导消费者，犯不犯法。从来没看过这么气人的书，希望看到...

[Lucene搜索引擎开发进阶实战_下载链接1](#)