

Elasticsearch大数据搜索引擎



[Elasticsearch大数据搜索引擎_下载链接1](#)

著者:罗刚

出版者:电子工业出版社

出版时间:2018-1

装帧:

isbn:9787121332333

作者介绍:

罗刚，猎兔搜索技术创始人 曾经担任国防大学科研处 技术顾问 工信部 舆情开发顾问
东南大学 社会导师 首都师范大学 金融课程讲师 北京石油化工学院 社会导师
北大光华管理学院 技术顾问 蓝汛公司搜索集群技术咨询顾问 新东方 创新研究院
研究员

目录: 第1章 使用Elasticsearch 1

- 1.1 基本概念 1
 - 1.2 安装 2
 - 1.3 搜索集群 5
 - 1.4 创建索引 6
 - 1.5 使用Java客户端接口 9
 - 1.5.1 创建索引 11
 - 1.5.2 增加、删除与修改数据 14
 - 1.5.3 分析器 16
 - 1.5.4 数据导入 17
 - 1.5.5 通过摄取快速导入数据 17
 - 1.5.6 索引库结构 17
 - 1.5.7 查询 18
 - 1.5.8 区间查询 22
 - 1.5.9 排序 23
 - 1.5.10 分布式搜索 23
 - 1.5.11 过滤器 24
 - 1.5.12 高亮显示 24
 - 1.5.13 分页 25
 - 1.5.14 通过聚合实现分组查询 26
 - 1.5.15 文本列的聚合 27
 - 1.5.16 遍历数据 28
 - 1.5.17 索引文档 29
 - 1.5.18 Percolate 29
 - 1.6 RESTClient 30
 - 1.6.1 使用摄取 31
 - 1.6.2 代码实现摄取 33
 - 1.7 使用Jest 33
 - 1.8 Python客户端 37
 - 1.9 Scala客户端 40
 - 1.10 PHP客户端 43
 - 1.11 SQL支持 44
 - 1.12 本章小结 48
- ## 第2章 开发插件 49
- 2.1 搜索中文 49
 - 2.1.1 中文分词原理 49
 - 2.1.2 中文分词插件原理 51
 - 2.1.3 开发中文分词插件 53
 - 2.1.4 中文AnalyzerProvider 55
 - 2.1.5 字词混合索引 57
 - 2.2 搜索英文 60
 - 2.2.1 句子切分 60
 - 2.2.2 标注词性 62
 - 2.3 使用测试套件 64
 - 2.4 本章小结 68
- ## 第3章 管理搜索集群 69

3.1 节点类型	69
3.2 管理集群	69
3.3 写入权限控制	70
3.4 使用X-Pack	71
3.5 快照	72
3.6 Zen发现机制	73
3.7 联合搜索	74
3.8 缓存	74
3.9 本章小结	75
第4章 源码分析	76
4.1 Lucene源码分析	76
4.1.1 Ivy管理依赖项	76
4.1.2 源码结构介绍	76
4.2 Gradle	77
4.3 Guice	77
4.4 Joda-Time	79
4.5 Transport	80
4.6 线程池	80
4.7 模块	80
4.8 Netty	81
4.9 分布式	81
4.10 本章小结	82
第5章 搜索相关性	83
5.1 BM25检索模型	83
5.1.1 使用BM25检索模型	86
5.1.2 参数调优	86
5.2 学习评分	86
5.2.1 基本原理	87
5.2.2 准备数据	87
5.2.3 Elasticsearch学习排名	89
5.3 本章小结	91
第6章 搜索引擎用户界面	92
6.1 JSP实现搜索界面	92
6.1.1 用于显示搜索结果的自定义标签	93
6.1.2 使用Listlib	98
6.1.3 实现翻页	100
6.2 使用Spring实现的搜索界面	102
6.2.1 实现REST搜索界面	102
6.2.2 REST API中的HTTP PUT	104
6.2.3 Spring-data-elasticsearch	106
6.2.4 Spring HATEOAS	112
6.3 实现搜索接口	113
6.3.1 编码识别	113
6.3.2 布尔搜索	116
6.3.3 搜索结果排序	116
6.4 实现相似文档搜索	117
6.5 实现AJAX搜索联想词	119
6.5.1 估计查询词的文档频率	119
6.5.2 搜索联想词总体结构	119
6.5.3 服务器端处理	120
6.5.4 浏览器端处理	125
6.5.5 拼音提示	127
6.5.6 部署总结	127
6.5.7 Suggester	128

- 6.6 推荐搜索词 129
 - 6.6.1 挖掘相关搜索词 130
 - 6.6.2 使用多线程计算相关搜索词 132
- 6.7 查询意图理解 133
 - 6.7.1 拼音搜索 133
 - 6.7.2 无结果处理 133
- 6.8 集成其他功能 134
 - 6.8.1 拼写检查 134
 - 6.8.2 分类统计 135
 - 6.8.3 相关搜索 141
 - 6.8.4 再次查找 144
 - 6.8.5 搜索日志 144
- 6.9 查询分析 146
 - 6.9.1 历史搜索词记录 146
 - 6.9.2 日志信息过滤 147
 - 6.9.3 信息统计 148
 - 6.9.4 挖掘日志信息 150
 - 6.9.5 查询词意图分析 150
- 6.10 部署网站 150
 - 6.10.1 部署到Web服务器 151
 - 6.10.2 防止攻击 152
- 6.11 本章小结 156
- 第7章 OCR文字识别 157
 - 7.1 Tesseract 157
 - 7.2 使用TensorFlow识别文字 161
 - 7.3 OpenCV 164
 - 7.3.1 预处理 166
 - 7.3.2 文字区域提取 169
 - 7.3.3 纠正偏斜 171
 - 7.3.4 Linux环境支持 172
 - 7.4 JavaCV 172
 - 7.5 本章小结 174
- 第8章 问答式搜索 176
 - 8.1 生成表示语义的代码 176
 - 8.2 信息整合 181
 - 8.2.1 实体对齐 181
 - 8.2.2 编辑距离 181
 - 8.2.3 Jaro-Winkler距离 187
 - 8.2.4 比较器 189
 - 8.2.5 Cleaner 189
 - 8.2.6 运行过程 190
 - 8.2.7 遗传算法调整参数 192
 - 8.3 自动问答 193
 - 8.3.1 问句处理器 193
 - 8.3.2 自动发现答案 198
 - 8.4 本章小结 199
- 第9章 Elastic系统监控 201
 - 9.1 Logstash 201
 - 9.1.1 使用Logstash 201
 - 9.1.2 插件 203
 - 9.1.3 数据库输入插件 206
 - 9.2 Filebeat 207
 - 9.3 消息过期 208
 - 9.4 Kibana 208

9.5 Flume 209
9.6 Kafka 210
9.7 Graylog 211
9.8 物联网数据 215
9.9 本章小结 216
• • • • • ([收起](#))

[Elasticsearch大数据搜索引擎_下载链接1](#)

标签

搜索

大数据

elasticsearch

ELK

评论

书的目录设计得不错，但书中大量内容都是在敷衍了事，甚至在书中发现完全重复的页面（p131 和 p143），价值不高。

[Elasticsearch大数据搜索引擎_下载链接1](#)

书评

[Elasticsearch大数据搜索引擎_下载链接1](#)