<<Lucene分析与应用>>

图书基本信息

书名：<<Lucene分析与应用>>

13位ISBN编号：9787111249924

10位ISBN编号：7111249925

出版时间：2008-9

出版时间：机械工业出版社

作者：吴众欣，沈家立　编著

页数：279

前言

Google被人熟知，Baidu在中国成功推广，“搜索”吸引着IT界的眼球，也吸引了更多开发者的好奇心。
于是诞生了Lucene，一个开源的全文检索API（Application Program Interface，应用程序界面）。
并在LHcene的基础上，衍生出了一个全文检索引擎(Nutch)和分布式文件系统(Hadoop)。
大家一定很好奇，Google的搜索引擎是如何工作的？
采用什么样的文件系统？
提供什么样的服务？
……我们无法得知。
Lucene与其相关的项目Nutch和Hadoop弥补了这个不足，让我们有机会了解到搜索引擎、分布式文件系统的内部工作原理。
如果介绍一个软件或者一套框架如何使用是比较容易的，但是要从源代码剖析内核，却不容易。
老吴与家立在写作期间，辗转难眠，思索如何表述才能够准确地把Lucene的设计精髓展现给读者。
最终确定通过对Lucene源代码的解说、辅以图表，并通过一些具体实例把所有源代码进行组织与剖析，完整地展示Lucene从建立索引到查询的完整过程。
并通过介绍一些Lucene的应用，和读者分享Lucene在具体项目开发中的应用环境。
同时，插入一些Lucene开发实例，抛砖引玉，试图让读者也能亲自体会Lucene本身的强大功能。
最后，为了进一步说明Lucene的应用环境，本书简单地介绍了 Nutch和Hadoop。
老吴很早就开始研读Lucene的源代码，并阅读了Dong Cutting的相关论文，对Lucene的内核具有深刻的认识。
我们很想与大家分享自己的学习体会和研究成果，于是决定把它写出来，家立负责Lucene多处应用部分的写作。
Lucene是一个很活跃的开源项目，因为老吴研究得比较早，版本以1.4.3为主。
为了能够跟上Lucene的步伐，家立推荐采用了较新的1.9-2.1版本进行分析。
但是该版本的内核变化比较大，因此需要重新分析、调试、总结。
为了尽快完成，我们日日熬夜，真所谓痛并快乐着。
在此非常感谢家人的支持，朋友的鼓励。
在此，向我的爱妻张信健对我的一贯支持表示感谢！
谢谢你，我的爱人！
希望对搜索引擎内核与运行机制感兴趣的朋友阅读此书，由于时间仓促，难免有所疏漏，请读者批评指正。
吴众欣

<<Lucene分析与应用>>

内容概要

本书对Lucene搜索引擎的源代码进行分析讲解，并用一些具体实例把所有源代码进行组织与剖析，完整地展示Lucene从建立索引到查询的过程。
本书通过介绍Lucene的应用，分析Lucene具体项目开发的应用环境。
最后简单地介绍了Nutch和Hadoop。
本书适用于开发搜索引擎的技术人员、Lucene爱好者等读者。

作者简介

吴众欣，西安交通大学在读博士，主攻搜索引擎与服务组合。
喜欢研读，头脑虽慢，滴水石穿。
好奇心重，兴趣广泛。

<<Lucene分析与应用>>

书籍目录

章节摘录

1.1搜索引擎与Lucene简介从最初的图书检索到链接查询，对图片、多媒体的搜索，直至现在的人肉搜索，搜索引擎作为信息融合平台将万千世界带到你的周围，让你触手可得，悄悄改变着你的生活，同时也可能将你暴露于众目睽睽之下。
有心人可能会考虑它背后的机理，以体味搜索引擎给我们的生活带来的变化。
现今的商业搜索引擎还是Google一家独大，微软也提供了MSN搜索引擎，但技术与经验积累还不够。而百度则是中文搜索中的佼佼者。
在商业搜索引擎中，核心技术与外部世界之间隔着一扇沉重的大门，幸好开源社区常常会将这扇门撬开些许缝隙，让我们能窥得冰山一角。
Lucene与MG4J正是开源搜索引擎项目，虽然“代码之前，了无秘密”，但是冰山一角也非轻易窥得。
本章我们也是浮光掠影地谈一谈搜索引擎，力图能给大家一些新的信息，以便从多个角度来认识搜索引擎与Lucene。

<<Lucene分析与应用>>

版权说明