**Luke:Lucene索引查看工具**

转载 2014年04月22日 18:50:35

**Luke介绍**

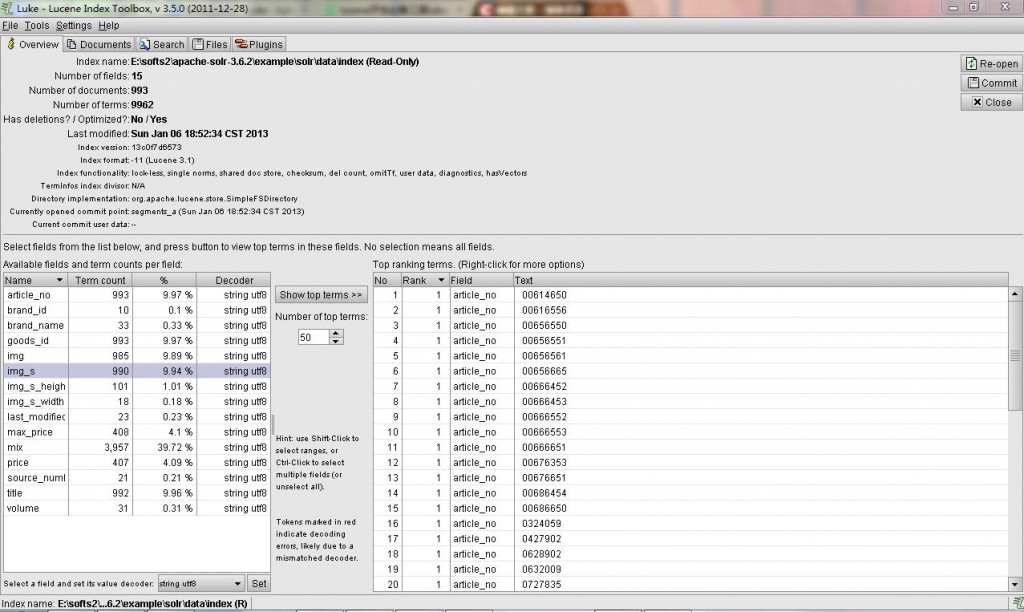
[Luke](http://code.google.com/p/luke/)是一个方便的索引查看和诊断工具，可以访问Lucene构建的索引文件，显示和修改某些索引内容。能提供：

* 通过document编号或term浏览索引
* 查看document内容，可复制到剪贴板
* 对频率最高的term的索引字段提供排名后的浏览
* 执行搜索语句并浏览搜索结果
* 分析搜索结果
* 从索引中选择性删除文件
* 重建原始文档字段，对其进行编辑，然后重新插入的索引
* 优化索引
* 可以打开hadoop文件系统内的索引文件

**Luke使用**

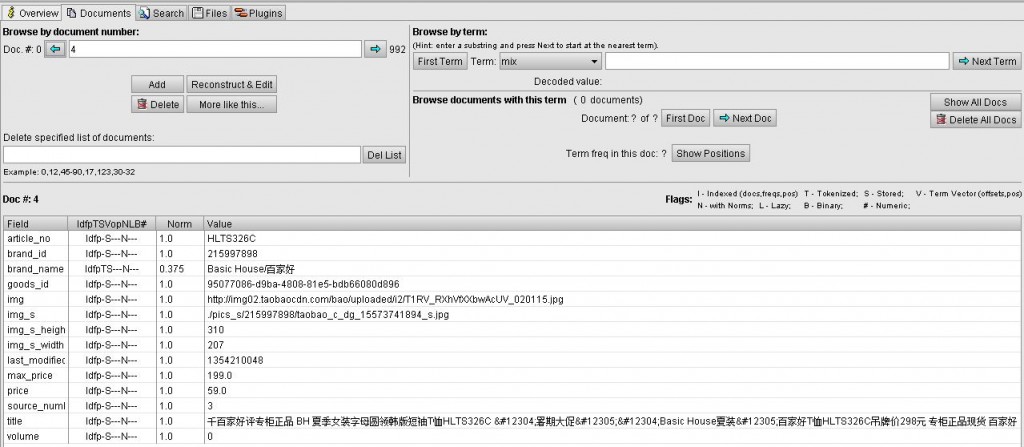
从[google code](http://code.google.com/p/luke/downloads/list)里下载lukeall的jar包，直接**java -jar lukeall-version.jar** 即可启动。

启动后选择你的索引文件路径，选择read-only打开：

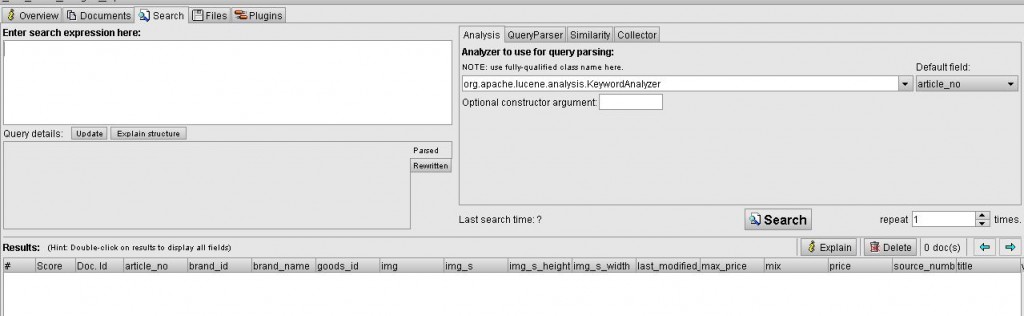


overview界面是用来进行索引的一般性查看和操作的，比如**索引目录，域信息，版本，term信息，Rank排名**等信息。**注意，索引文件里Analyze却不Store的字段信息还是不可见的，也就是只能看STORE了的内容。**

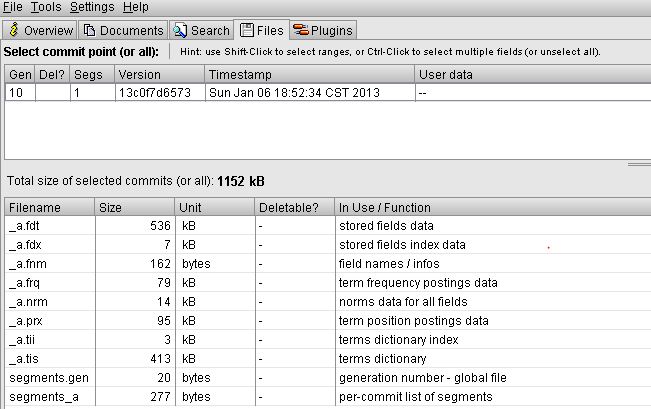
documents界面是用来进行文档的操作和查看的，能根据文档编号和词进行查找，其实这个就是**搜索功能**。



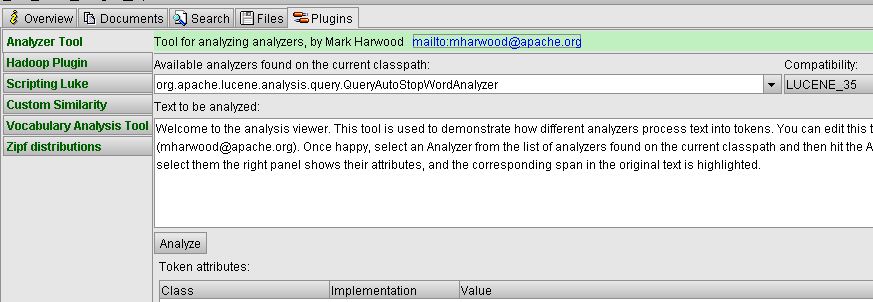
search界面是可以进行索引的**搜索测试**，可以编写lucene搜索语句，看到语句解析后的query树，还可以选择进行搜索的分词器、默认字段和重复搜索次数，然后下面的listview中就会列出一个搜索的的文档的所有保存的（store）字段的值，可以看到查询花费的时间



file界面，故名思义，这个就是用来查看每个索引相关文件的一些属性的界面，具体的话，可以通过这个界面分析下索引文件的多少，是否需要优化或者合并等等



最后一个plugins界面，就是可以看到luke提供的各种插件。比较有用的还是分词工具，提供一个分词的类，然后下面文本框输入一段文本，然后就可以让这个工具帮你分词，你可以看到详细的分词信息，对自定义分词器的调试或者测试。还有一个hadoop插件，支持从hadoop节点中获取节点中文件的相关信息，对分布式搜索引擎搭建有用，算是**支持多平台的lucene索引文件块的查看**。



**个人理解**

其实Lucene构建的索引，无论从结构上说还是功能上说，和一个DBMS数据库很相似，你可以认为Luke做的事情就是包装了Lucene的IndexReader和IndexSearcher之后，变成一个界面化的索引展示和管理工具。你完全可以自己写程序在cmd里查看，但是没有Luke提供的展示那么直观和多样。

Luke只提供带桌面界面的操作系统，就是说你的linux服务器，如果是命令行界面的话，是使用不了Luke来查看的，这是不是一个可以想办法改进的地方？

luke源码结构简单，通过使用luke并阅读源码，对这个工具一定可以有更全面的认识。

