

```
支付宝打赏:
                                 lter indexWriter = new IndexWriter(directory, analyzer, true,
                                 IndexWriter.MaxFieldLength.UNLIMITED);
                                 有需要建立索引的文件
                                 iles = new File(dataDirectory).listFiles();
                             4篇
                                 i = 0; i < files.length; <math>i++) {
                             6篇
                                 文件是第几个
                            10篇
                                 !em.out.println("这是第" + i + "个文件-----");
                                 文件的完整路径
                             4篇
                                 iem.out.println("完整路径: " + files[i].toString());
                            12篇

英取文件名称
                                 ing fileName = files[i].getName();
                                 失取文件后缀名,将其作为文件类型
微信公众号
                                 Ing fileType = fileName.substring(fileName.lastIndexOf(".") + 1,
                                      fileName.length()).toLowerCase();
                                 文件名称
                                 iem.out.println("文件名称: " + fileName);
                                 文件类型
                                 :em.out.println("文件类型: " + fileType);
                                 iment doc = new Document();
                                 itring fileCode = FileType.getFileType(files[i].toString());
                                 至看各个文件的文件头标记的类型
                                 system.out.println("fileCode=" + fileCode);
PL/SQL developer连接oracle出现"ORA-
                                 itStream in = new FileInputStream(files[i]);
1215年NS:could not resolve the connect
                                 itStreamReader reader = null;
阅读量:41914
                                 fileType != null && !fileType.equals("")) {
关于Python中的闭包
最新评论
TensorFlow基础学习——"Only call
                                 if (fileType.equals("doc")) {
`sparse_softmax_cross_entropy_with_logits
                                      // 获取doc的word文档
                                     WordExtractor wordExtractor = new WordExtractor(in);
qq_40285082:客户端和服务端都使用的话,就出
                                      // 创建Field对象,并放入doc对象中
施可數了Walkanyespylietualwempager创建
                                     // Field的各个字段含义如下:
Python原拟 预i 版的安装、...
                                     // 第1个参数是设置field的name,
dengdoms822基础学习siked更用frentin激谢您的
                                     // 第2个参数是value, value值可以是文本(String类型, Reader类型或者是预分享的TokenStream)
莱楝均餐变海魚
                                     // 二进制 (byet[]), 或者是数字 (一个 Number类型)
windows下壁in/學字差、...
                                     // 第3个参数是Field.Store,选择是否存储,如果存储的话在检索的时候可以返回值
sikedaima: 為特一介 tr.control_dependencies()函数用法
                                     // 第4个参数是Field.Index,用来设置索引方式
利用iQuery UI为CMS网站..
                                     doc.add(new Field("contents", wordExtractor.getText(),
      . 我是在夕急需源码,
                                             Field.Store.YES, Field.Index.ANALYZED));
                                     // 关闭文档
                                     wordExtractor.close();
                                     System.out.println("注意: 已为文件"" + fileName + ""创建了索引");
       223 : [reply]liuxiang66666[/reply] 谢谢
很高短能够帮到您事况 评论
阅读第: 3363 294<sup>2</sup>篇 203
                                 } else if (fileType.equals("docx")) {
                                     // 获取docx的word文档
        亚马逊AWS基础学习
                                     XWPFWordExtractor xwpfWordExtractor = new XWPFWordExtractor(
          ■ 访问:104万+
|读量:53972 7 章
                                             new XWPFDocument(in));
                                      // 创建Field对象,并放入doc对象中
                                     doc.add(new Field("contents", xwpfWordExtractor.getText(),
                                             Field.Store.YES, Field.Index.ANALYZED));
                                     // 关闭文档
                                      xwpfWordExtractor.close();
                  下载CSDN APP
                                      System.out.println("注意: 已为文件"" + fileName + ""创建了索引");
              28454 8篇
kefu@csdn.net
                                 } else if (fileType.equals("pdf")) {
                                      // 获取pdf文档
                                     PDFParser parser = new PDFParser(in);
                                     parser.parse();
                                     PDDocument pdDocument = parser.getPDDocument();
                                     PDFTextStripper stripper = new PDFTextStripper();
                                      // 创建Field对象,并放入doc对象中
                                     doc.add(new Field("contents", stripper.getText(pdDocument),
                                             Field.Store.NO, Field.Index.ANALYZED));
                                      // 关闭文档
                                      pdDocument.close();
                                      System.out.println("注意: 已为文件"" + fileName + ""创建了索引");
```

```
支付宝打赏:
                                   } else if (fileType.equals("txt")) {
                                       // 建立一个输入流对象reader
                                       reader = new InputStreamReader(in);
                                       // 建立一个对象,它把文件内容转成计算机能读懂的语言
                              4篇
                                       BufferedReader br = new BufferedReader(reader);
                                       String txtFile = "";
                              6篇
                                       String line = null;
                              10篇
                                       while ((line = br.readLine()) != null) {
                              4篇
                                           // 一次读入一行数据
                              12篇
                                           txtFile += line;
                                       }
                                       // 创建Field对象,并放入doc对象中
                                       doc.add(new Field("contents", txtFile, Field.Store.NO,
                                               Field.Index.ANALYZED));
                                       System.out.println("注意: 已为文件"" + fileName + ""创建了索引");
                                   } else {
                                       System.out.println();
                                       continue;
                                   训建文件名的域,并放入doc对象中
PL/SQL developer连接oracle出现"ORA-
                                   add(new Field("filename", files[i].getName(), Field.Store.YES,
12154:TNS:could not resolve the connect
                                       Field.Index.NOT_ANALYZED));
阅读量:41914
                                   训建时间的域,并放入doc对象中
                                   add(new Field("indexDate", DateTools.dateToString(new Date(),
关于Python中的闭包
                                       DateTools.Resolution.DAY), Field.Store.YES,
最新评论
TensorFlow基础学习——"Only call
                                       Field.Index.NOT_ANALYZED));
`sparse_softmax_cross_entropy_with_logits
                                   員入IndexWriter
                                   exWriter.addDocument(doc);
qq_40285082:客户端和服务端都使用的话,就出
                                   负行
施可數了Walkanyespacetual wrangager创建
                                   :em.out.println();
Python原拟 预i 版的安装、...
dengdoffg822基础学习siked更同trentin激谢您的
                                   ndexWriter里面有多少个索引
奏楝/释摩座德息
                                   but.println("numDocs=" + indexWriter.numDocs());
windows下壁in學字差、...
sikedaima: 為特工介
tt.control_dependencies()函数用法
                                   .ter.close();
利用iQuery UI为CMS网站...
      . 我是在夕急需源码,
       223 : [reply]liuxiang66666[/reply] 谢谢
很高知能够帮到您事况 评论
阅注明: 3363 294<sup>2</sup>篇 203
                                   test.poi;
         亚马逊AWS基础学习
           ile:
                                   OException;
                                   ne.lucene.analysis.standard.StandardAnalyzer;
                                   ie.lucene.document.Document;
                                   ie.lucene.queryParser.ParseException;
                   下载CSDN APF
                                   ie.lucene.queryParser.QueryParser;
                                   ie.lucene.search.IndexSearcher;
                3454
■ kefu@csdn.net
                                   ne.lucene.search.Query;
                                   ie.lucene.search.ScoreDoc;
                                   ie.lucene.search.TopDocs;
                                   ie.lucene.store.Directory;
                                   ie.lucene.store.SimpleFSDirectory;
                                   ne.lucene.util.Version;
                                   :eneSearch {
                                   c void main(String[] args) throws IOException, ParseException {
                                   ndexDirectory = "D:\\Studying\\poi\\test\\indexDirectory";
                                   irectory对象 , 也就是分词器对象
```

```
使用Lucene对doc、docx、pdf、txt文档进行全文检索功能的实现 - 干勾鱼的CSDN博客 - CSDN博客
支付宝打赏:
                                y directory = new SimpleFSDirectory(new File(indexDirectory));
                                IndexSearcher对象,相比IndexWriter对象,这个参数就要提供一个索引的目录就行了
                                ircher indexSearch = new IndexSearcher(directory);
                                ueryParser对象,
                                参数表示Lucene的版本,
                            4篇
                                表示搜索Field的字段,
                                表示搜索使用分词器
                            6篇
                                ser queryParser = new QueryParser(Version.LUCENE_30,
                            10篇
                                "contents", new StandardAnalyzer(Version.LUCENE_30));
                                uery对象
                            4篇
                                lery = queryParser.parse("百度");
                           12篇
                                课 TopDocs里面有scoreDocs[]数组,里面保存着索引值
                                 hits = indexSearch.search(query, 10);
                                 totalHits表示一共搜到多少个
微信公众号
                                out.println("找到了" + hits.totalHits + "个");
                                its.scoreDocs数据,并使用indexSearch.doc方法把Document还原,再拿出对应的字段的值
                                i = 0; i < hits.scoreDocs.length; i++) {</pre>
                                reDoc sdoc = hits.scoreDocs[i];
                                ment doc = indexSearch.doc(sdoc.doc);
                                :em.out.println(doc.get("filename"));
                                rch.close();
                                3有了,就不做过多解释了。需要的jar包如下:
PL/SQL developer连接oracle出现"ORA-
12154:TNS:could not resolve the connect
阅读量:41914
                                -4.1.jar
关于Python中的闭包
最新评论
TensorFlow基础学习——"Only call
qq_40285082:客户端和服务端都使用的话,就出
施可數了Walkanyespacetual wrangager创建
PXth8V束状颓槁的安装、...
dengdoffg822基础学习siked更同trentin激谢您的
奏楝/≱變率德息
windows下壁in/學字差、...
sikedaima: 激赞 tt.control_dependencies()函数用法
                                16.jar
利用iQuery UI为CMS网站...
      . 我是在夕急需源码,
                                 ,读取pdf的类到Apache PDFBox官网下载,这里用的1.8.13版本,2.0版本的调用方式与1.0版本已经
       223 : [reply]liuxiang666666[/reply] 谢谢
很高學能够帮到您事次 评论
阅读9: 3363 294<sup>2</sup>篇 203
        亚马逊AWS基础学习
          下载CSDN APP
             28454 8篇
kefu@csdn.net
```







微信公众号

PL/SQL developer连接oracle出现"ORA-

1215年NS:could not resolve the connect 阅读量:41914

关于Python中的闭包

最新评论 TensorFlow基础学习——"Only call

`sparse_softmax_cross_entropy_with_logits qq_40285082:客户端和服务端都使用的话,就出 施可數了Walkanyespacetual wrangager创建

Python原拟 预i 版的安装、...

dengdoms822基础学习siked更用frentin激谢您的 **李操/移變**來海魚

Windows下壁訊學家装、…

sikedaima: 為特一介 tr.control dependencies()函数用法

利用jQuery UI为CMS网站...

chenkuntian:楼主,急需源码, **博夫**5.0043@qq.com,谢谢



*装Oracle11g. PhoneGab学习 223: [reply]liuxiang66666[/reply] 谢谢 很高兴能够帮到您! 阅读量:3363 2 篇



工作时间 8:00-22:00

关于我们 招聘 广告服务 图址论图 ©1999-2018 江苏乐知网络技术有限公司 汪菊知達为计算机有限公司 北京创新乐知篇 **信息技术**有限公司版权所有

网络449报警服务 经营性网站备案信息7篇 北京互联网违法和不良信息举报中心 中国互联网举报中心

5篇 Hadoop

indows平台下的文件搜索工具,可以按照指定类型(word, excel, txt, java..php等文档文本类型)内...

 \blacksquare

ir包:

9开源项目,是著名的Apache Jakarta大家庭的一员,并且基于Apache软件许可。同样,Lucene是当前引

叟索的流程 - 算法魔功

因此,你不需要深入理解Lucene的信息索引和检索... 因为它的简单。 来自: 算法质

orld - 青鸟飞鱼

elloworld; import org.apache.lucene.analysis.standard.StandardAnalyzer...

_36470908的博客 commons-io-2.6.jar,

lucene-analyzers-common-7.2... 来自: qq_36470908的博客

(doc,docx,ppt,pptx,xls,xlsx,pdf,txt等) - 明天很美--

tx,xls,xlsx,pdf,txt等各类文件内容,包含代码和jar包

型文件内容 - soj_1998的博客

3容 由于文件可能会很大,一次性读取可能会造成内存溢出,所以分块... 来自: soj_1998的博客

-Lucene的介绍 - 李阿昀的博客 @ 2035

支术——Lucene,现在这个技术用到的比较多,我觉得大家还是应该掌握... 来自: 李阿昀的博客

服务器的基本使用(一)——SolrJ的使用 - yyf的博客 @ 659

门直接用Lucene虽然可以实现很多自己想要的自定义功能,但是对于一... 来自: yyf的博客

应各种平台的在线 Office 文档预览 - tongxiaomingicesword的专栏

7309 前言 在许多Web应用中都需要预览文档的功能。而用户可能用不... 来自: tongxiaomingiceswor...

全文索引 - wei15801110272

@ 322

1761

@ 15⁻

⊚ 22

© 650

来自:青鸟7

也许词不达意,也许有其他问题,作为一个新手,第一次使用lucene,贴...

DF文档 - 断弯刀

henyong2013

文搜索 (支持双层pdf)

us全文检索pdf文件

索pdf文件,网上看了很多开源框架,最后用了lucene+pdfbox+lius完成,支持建立包含子目录索引,全文检索多级目录的功能。 附源码

PDF文本进行内容的解析 - 杨鑫newlife的专栏

FBox.zip的包*利用lucene对PDF文本进行内容的解析*读取pdf文件的... 来自: 杨鑫newlife的专栏

lerZS的博客 2857

入门 3.Lucene进阶 全文检索一, 生活中的搜索: 1.Windows系统中的有搜... 来自: CoderZS的博客

并搜索关键字,为其盖章-sdizoea的博客 3977

字定位pdf的盖章位置的相关需求,其中关键字可配置多个(包含pdf文档... 来自: sdizoea的博客

文件存储和读取技术详解 -401

组研发技术人员,负责搜索引擎架构的设计和研发,搜索集群健康状况... 来自: 达观数据

页览办公文件(doc、docx、xls等)-天蓝色blue

、docx、xls、xlsx、pdf等),转化效果个人觉得比较理想

x, Pdf的内容 - GhostHouse

⊚ 3599

卖取doc, docx,pdf文件内容。在网上搜罗许久,还是发现有些好东西可... 来自: GhostHouse

〔统 java源码 (信息检索技术) lucene又鬥稅索系統...lava源码...可执行程序,支持pdf,doc,xls,ppt,html,txt。 归档 **索详解及开发示例 - 纸上得来终觉浅, 绝知此事要躬行。** @ 28 2018年11月 4篇 s.com/xing901022/p/3933675.html 讲解之前, 先来分享一些资料 来自:纸上得来终觉浅,红 2018年10月 6篇 2018年9月 10篇 pdf、doc、rtf、txt以及doc转docx - 磁悬浮青蛙呱呱呱 @ 26 ··· 2018年8月 4篇 弓存为其它格式,比如pdf、doc、rtf、txt,或者将doc升级为docx,用VBA... 来自:磁悬浮青蛙呱呱 2018年7月 12篇 Excel2003、word2007、Excel2007、PowerPoint、Text 可用于Luce... @ 20 展开 ava.io.BufferedReader; import java.io.File; import java.io.FileInputStream... 来自: JohnBanana的€ 执门文章 i) - PDF(带目录) - winseii 划第2版)》基于Apache的Lucene 3.0,从Lucene核心、Lucene应用、案例分析3个方面详细系统地介绍了Lucene,包括认识 Oracle中TO_NUMBER()函数的用法 阅读量:84737 s+txt+多层文件 - 伏龙的专栏 1742 使用MySQL Workbench进行数据库设计 --MySQL Workbench使用方法总结 多层文件2009-09-21 20:33/* * To change this template, choose Tools | T... 来自:伏龙的专栏 阅读量:61348 .5和IKAnalyzer2012全文分词检索技术文档 如何进入、退出docker的container 阅读量:59888 实现基础的全文检索实现 修改nginx运行端口 阅读量:46151 4,实现文件查找 - 都市桃源 1863 PL/SQL developer连接oracle出现"ORAhe软件基金会4 jakarta项目组的一个子项目,是一个开放源代码的全文... 来自:都市桃源 12154:TNS:could not resolve the connect 阅读量:41914 endeguang的专栏 ⊚ 2093 :个包当然是关于lucene索引的数据结构单元——document。这个包里面... 来自: shendeguang的专栏 最新评论 与实现(资深Redis技术专家撰写,深入了解Redis技术内幕的必读之作。从源... 在iava中使用spring整合c... qq_40285082:客户端和服务端都使用的话,就出 整地讲解了Redis的内部机制与实现方式,对Redis的大多数单机功能以及所有...

qq_40285082:客户端和服务端都使用的话,就出现问题了No namespace on "html&quo...

windows下nginx的安装、.

dongdong9223: [reply]sikedaima[/reply] 谢谢您的

支持!多多交流!

windows下nginx的安装、...

sikedaima: ▲赞一个

利用jQuery UI为CMS网站...

chenkuntian:楼主,急需源码, 381540043@qq.com,谢谢

linux下安装Oracle11g...

dongdong9223: [reply]liuxiang66666[/reply] 谢谢 您的认可,很高兴能够帮到您!

联系我们





微信客服

下载CSDN APP

■ QQ客服

kefu@csdn.net

● 客服论坛

2 400-660-0108

工作时间 8:00-22:00

关于我们 招聘 广告服务 网站地图

營 百度提供站內搜索 京ICP证09002463号 ◎1999-2018 江苏乐知网络技术有限公司 江苏知之为计算机有限公司 北京创新乐知 信息技术有限公司版权所有

网络110报警服务 经营性网站备案信息 北京互联网违法和不良信息举报中心 中国互联网举报中心