

原

Lucene对office的word, Excel操作

2015年04月05日 16:41:54 杨鑫newlife 阅读数：1373 更多

版权声明：学习交流为主，未经博主同意禁止转载，禁止用于商用。 <https://blog.csdn.net/u012965373/article/details/44888875>

首先说明Lucene如果想对office进行操作的话，需要额外的包

这里有几种方法

一是使用POI

二是使用jacob

这里的代码使用的是jacob对于word进行处理

代码：

```
package jacob;

import com.jacob.activeX.ActiveXComponent;
import com.jacob.com.Dispatch;
import com.jacob.com.Variant;

public class WordReader {

    public WordReader() {
        // TODO Auto-generated constructor stub
    }

    public static void extractDoc(String inputFile, String outputFile){
        boolean flag = false;

        //打开word应用程序
        ActiveXComponent app = new ActiveXComponent("Word.Application");

        try{
            //设置word不可见
            app.setProperty("Visible", new Variant(false));

            //打开word文件
            Dispatch doc1 = app.getProperty("Document").toDispatch();
            Dispatch doc2 = Dispatch.invoke(doc1, "Open", Dispatch.Method, new Object[]{inputFile, new Variant(false),new Variant(true)}, new int[1]).toDispatch();
            //作为txt格式保存到临时文件
            Dispatch.invoke(doc2, "SaveAs", Dispatch.Method, new Object[]{outputFile, new Variant(7)}, new int[1]);
            //关闭Word
            Variant f = new Variant(false);
            Dispatch.call(doc2, "Close", f);
        } catch (Exception e) {
            e.printStackTrace();
        }
    }
}
```

```
e.printStackTrace();
}finally{
app.invoke("Quit", new Variant[]{});
}
if(flag == true){
System.out.println("Transformed Successfully");
}else{
System.out.println("Transform Failed");
}
}

}

public static void main(String[] args) {
// TODO Auto-generated method stub
WordReader.extractDoc("E:\\Hadoop.doc", "E:\\jacob.txt");
}

}
```

远离假传奇！这游戏爆率+999倍，有充值有VIP算我输！

贪玩游戏 · 顶新



想对作者说点什么

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 3622

转载请注明出处：http://blog.csdn.net/dongdong9223/article/details/76273859 本文出自【我是干勾鱼... 来自：干勾鱼的CSDN博客

lucene读取word,excel,pdf - aidayei的专栏 4229

前面在写lucene入门的时候，例子只能对txt文档建立索引，不能对word,excel,pdf建立索引，要读取这... 来自：aidayei的专栏

解析pdf、word2003、Excel2003、word2007、Excel2007、PowerPoint、Text 可用于Lu... 2044

所用到的jar 包 可以到 import java.io.BufferedReader; import java.io.File; import java.io.FileInputStrea... 来自：JohnBanana的专栏

Python全栈学完需要多少钱？

零基础学爬虫，你要掌握学习那些技能？需要学多久？

Lucene检索WORD等文件 - 努力一点 704

Lucene是什么？lucene是apache开源的全文检索的框架，不像百度那样的搜索引擎拿来就能用！Luce... 来自：努力一点

Lucene对pdf、word、html等文件的处理 - aitcax的专栏 2678

Lucene在创建索引的过程中，yuanc PDFBox是一个在java环境中对pdf文件进行解析处理的开源软件... 来自：aitcax的专栏

lucene全文检索word内容 - fanglm的专栏 6369

创建索引 Lucene 3.0+ package test; import java.io.File; import java.io.FileInputStream; import java.io... 来自：fanglm的专栏

VC操作word和excel文件，查询与读写[依赖office环境] - 总有一人，触不可及 1513

2016-07-26 10:15AM 田海涛>>记录>>合肥工业大学 由于孟师兄的碗扣式支架建模系统需求，须在程... 来自：总有一人，触不可及

lucene实现pdf，doc，xls，ppt，htm，html等格式文件的检索 - 脑瓜子的程序人生 1089

代码如下，代码没有优化，仅实现功能 该代码复制到程序中不能直接使用，需要下载文章最后的例子... 来自：脑瓜子的程序人生

远离假传奇！这游戏爆率+999倍，有充值有VIP算我输！

贪玩游戏 · 顶新

poi word excel ppt pdf 读取纯文件流 适用于lucene全文检索 - jc橙子的博客 514

package com.fzky.diams.web.luncene; import java.io.BufferedReader; import java.io.File; impo... 来自：jc橙子的博客

相关热词

LuceneLUCENElucene查看luceneLucene去重

一步一步跟我学习lucene（15）---java读取word excel pdf及lucene搜索之正则表达式查询...

3216

今天快下班的时候收到了一个群友的问题，大意是读取文本文件中的内容，找出文件中的手机号和邮...

来自：wuyinggui的专栏



bit_kaki

关注

115篇文章



天外野草

关注

236篇文章



数据架构师

关注

659篇文章

lucene全文检索 - yangmy的博客

219

lucene全文检索可识别txt,doc,docx,xls,xlsx,pdfLucene 原理与代码分析完整版.pdfpackage com.abc;i...

来自：yangmy的博客

Lucene实现各种常见文档的全文检索 - ::: 移动互联网时代开发者 :::

4140

Lucene是apache软件基金会jakarta项目组的一个子项目，是一个开放源代码的全文检索引擎工具...

来自：::: 移动互联网时代开...

手把手教你如何玩转Solr（包含项目实战）- Cs_hnu_scw的博客

1.5万

备注：学习Solr最好先了解一下Lucene的基本内容，不需要很熟，但是知道个基础即可。一：Solr简...

来自：Cs_hnu_scw的博客



免费云主机试用一年

百度广告

Solr的使用 - 没有伞的孩子必须努力奔跑！一周家林

1998

1、 schema.xml schema.xml文件在SolrCore的conf目录下，在此配置文件中定义了域以及域的类型等...

来自：没有伞的孩子必须努...

solr教程，值得刚接触搜索开发人员一看 - awj3584

23.4万

Solr调研总结 开发类型 全文检索相关开发 Solr版本 4.2 文件内容 本文介绍solr的功能使用及相关注意...

solr优缺点 - qq_39791917的博客

4255

Solr是一个可伸缩性的、能够即时部署的、致力于信息检索的、可以处理海量信息的、以处理自然语...

来自：qq_39791917的博客

solr在java中的使用 - 实践是最好的方法

2.1万

学习solr的基础知识：http://blog.csdn.net/u012385190/article/details/51682380SolrJ是操作Solr的JAV...

来自：实践是最好的方法

利用lucene对文件内容进行关键字检索 - zh_winer的博客

1750

一、概述 关于lucene的具体介绍，请自行百度。 二、实例讲解 在具体实现之前，请根据自...

来自：zh_winer的博客



大数据时代下的数据可视化研究

百度广告

lucene4.1 小例子 全文检索 - 既认准这条路，又何必在意要走多久

2640

自己需要到一个全文检索，就搜了一下整理了一个简单可用的代码 实现很简单 首先需要 lucen 的包...

来自：既认准这条路，又何...

全文检索之lucene的优化篇--查询篇 - suchy

780

全文检索的查询是很重要的，里面的很多的查询方式，就像是Google和Baidu中的高级查找。 首先...

来自：suchy

Lucene——解析PDF，word，excel - sunghosts的专栏

948

PDF 用到的java PDF类库是：PDFBox，下载后其中lib和externa文件夹下的jar文件需要引入项目 解...

来自：sunghosts的专栏

python应用系列教程——python操作Office办公软件（excel） - 全栈工程师开发手册（原创）

5218

全栈工程师开发手册（作者：栾鹏）python教程全解 python操作Office办公软件（excel）。前提条件...

来自：全栈工程师开发手册...

下载

C#代码操作Excel,不需要安装office - 第七支队

01-14

com.Ole.excel 源码,不需要在服务器上安装Excel,完全使用C#代码实现,更多excel支持的高级功能可能没有,但实际开发中基本够用了.附加了一个示例和一个Table导出到excel的示例

[推动全社会公益氛围形成，使公益与空气和阳光一样触手可及。](#)
公益缺你不可，众多公益项目等你PICK——百度公益 让公益像「空气和阳光」一样触手可及！
gongyi.baidu.com

https://blog.csdn.net/u012965373/article/details/44888875?utm_source=blogxgwz2

3/7

- C++**操作Word**学习笔记（四） - aleyuan的专栏

1028

【当前博文转自http://blog.sina.com.cn/s/blog_4aac71b01000brk.html】 【本文有打印相关操作】 1... 来自： aleyuan的专栏
- 总结下python读写**word、excel**文件时碰到的坑 - feng_vs_sunzhilong的博客

367

巨坑一：新建word或者excel时 编写程序时，遇到的奇葩现象：当在命令行界面，新建word或者excel... 来自： feng_vs_sunzhilong...
- java**操作word/excel/pdf**等文件技术方案 - 山月记

1225

最近项目中遇到很多对word/excel/pdf等文件的操作，解决方案有好多，开源免费有:利用openoffice组... 来自： 山月记
- xls和xlsx的打开方式 - orichisonic的专栏

2132

AccessDatabaseEngine.exe自然是要装的，但是关键是Provider=Microsoft.Jet.Oledb.4.0一定要替换... 来自： orichisonic的专栏
- office文档全文检索以及在线预览 - xiezunjin的博客

5873

office文档(word pdf excel txt) 的lucene 全文检索以及在线预览 来自： xiezunjin的博客



如何购买真正的云主机

百度广告

- 在网页中在线查看文档（ doc、docx、xls、xlsx、pdf、swf ） - lin062854的专栏

4.3万

需求：用户上传附件后，点击查看，可以在页面直接查看到附件内容，样式排版需要和附件文档里一... 来自： lin062854的专栏
- wifi交互过程 - tinyso

2661

三、实例分析 仅列举几个特例，其他的报文大同小异。一个正常的报文一般是有请求，则会有回应。 ...
- lucene3.6全文检索**word2007** - 程序员！我当定了！

1472

最近做了一个项目,有个需要是要全文检索上传的word文档的内容,经过经理推荐用lucene,我下载了最... 来自： 程序员！我当定了！
- Solr原理 - 汪翰翔的Blog

1.5万

Solr原理 Solr对外提供标准的http接口来实现对数据的索引的增加、删除、修改、查询。在 Solr 中， ... 来自： 汪翰翔的Blog
- Solr(二)——Java使用solr全文搜索 - 东天里的冬天

1187

使用SolrJ可以实现索引库的增删改查操作。 一.步骤 1添加 第一步：把solrJ的jar包添加到工程中。 第... 来自： 东天里的冬天
- [推动全社会公益氛围形成，使公益与空气和阳光一样触手可及。](#)
公益缺你不可，众多公益项目等你PICK——百度公益 让公益像「空气和阳光」一样触手可及！
gongyi.baidu.com

- 【Solr】——Solr7安装教程 - 不负时光

8182

前提 solr已经升级7.1，但是我们公司的solr还是使用的4.4，你们说low不low！！重要的是，人家... 来自： 不负时光
- 什么是Solr - mischen520的博客

4379

Solr 是Apache下的一个顶级开源项目，采用Java开发，它是基于Lucene的全文搜索服务器。Solr提... 来自： mischen520的博客
- solr 高级进阶，解决以下五个问题，就可以投产了 - 阿胆Amazing的博客

315

一、多表、多数据源 数据导入的问题 之前介绍过通过data-config.xml 导入数据的问题，实际开发过程... 来自： 阿胆Amazing的博客
- 关于C#**操作WPS**和**office**兼容性的问题 - 风之翼的博客

4823

最近一直在做的开发是关于导出word的功能，一开始的做法是在VS中直接添加引用office PIA，Micros... 来自： 风之翼的博客
- 用 Ruby on Rails 实现适应各种平台的在线 **Office** 文档预览 - tongxiaomingices**word**的专栏

481

http://ruby-china.org/topics/17309 前言 在许多Web应用中都需要预览文档的功能。而用户可能用不同... 来自： tongxiaomingiceswor...

陈小青坦言：这游戏不充钱都能当全服大哥，找到充值入口算我输！

贪玩游戏 · 顶新

下载

office竞赛试题（ word、excel、ppt ）

office竞赛试题（ word、excel、ppt ），有一定难度，不过还是很不错的！希望对你有帮助！

- 怎么样在WEB上实现office在线编辑的功能 - u011258203的博客

3367

可以使用POI将 word转变为html进行保存。然后使用H5的特性进行编辑。详情见附件。 ...

来自： u011258203的博客
- 在VS2010中利用MFC读取word、ppt、excel格式文件 - hnu_wan的博客

6060

本文源码的应用环境说明：Windows 7，Microsoft Visual Studio 2010，Microsoft Office Excel 2013 ...

来自： hnu_wan的博客
- lucene对文件做简单的索引 - 成就源于知识，知识在于积累。

797

package com.mylucene; import java.io.File; import java.io.FileReader; import java.io.IOException; im...

来自： 成就源于知识，知识...
- 使用Lucene对预处理后的文档进行创建索引（可运行） - 杨鑫newlife的专栏

999

搜索引擎

来自： 杨鑫newlife的专栏

远离假传奇！这游戏爆率+999倍，有充值有VIP算我输！

贪玩游戏 · 顶新

- solr简介、学习详细过程！（超详细~） - 邵小小的博客

1.1万

solr是什么呢？一、Solr它是一种开放源码的、基于 Lucene Java 的搜索服务器，易于加入到 Web 应...

来自： 邵小小的博客
- 项目中如何使用solr - 程高伟

4.2万

总结了solr在项目如何使用

来自： 程高伟
- Solr的基本使用 - happy_meng

1.4万

Solr的基本使用 1.schema.xml 1.1.Field和fieldType（重点） schema.xml，在SolrCore的conf目录下...

来自： happy_meng
- C++中auto型的详细解释 - 小白_努力

2456

auto：用来声明自动变量。它是存储类型标识符，表明变量（自动）具有本地范围，块范围的变量声...

来自： 小白_努力
- solr教程 - 李木一的博客

2042

Solr调研总结：开发类型全文检索相关开发Solr版本4.2文件内容本文介绍solr的功能使用及...

来自： 李木一的博客

鲲已经过时了！这游戏开局带虎，无限进阶，0元满回馈！

贪玩游戏 · 顶新

- solr7.4.0集群搭建 - 龚道松

964

三、solr集群搭建 注意，在搭建solr集群前，建议最好有一个solr服务是已经搭建好的，可以简化大量...

来自： 龚道松



代码整洁之道(Robert C. Martin力作，韩磊献译)

软件质量，不但依赖于架构及项目管理，而且与代码质量紧密相关。这一点，无论是敏捷...

杨鑫newlife

关注

原创

1529

粉丝

737

喜欢

184

评论

287

等级：

博客 日

访问：

225万+

积分：

3万+

排名：

162

勋章：



数据可视化工具



博客介绍

wechat:812716131

技术交流群请联系上面wechat

Mail: yangxin812716131@163.com

[LinkedIn](#)

博客专注大数据 && AI && 算法

混迹于北京

家乡佳木斯

最新文章

[Python实现决策树应用之判断隐形眼镜的类型](#)

[Python实现决策树应用之判断鱼类与非鱼类](#)

[Python实现DescionTree决策树 --- 选择切分数据集的最佳特征](#)

[Python实现DescionTree决策树 --- 划分数数据集](#)

[Java实现时间工具类](#)

博主专栏



[搜索引擎](#)
阅读量：138024 123 篇



[C指针编程之道学习笔记](#)
阅读量：16386 13 篇



[数据库学习笔记](#)
阅读量：320218 188 篇



[Linux/Unix学习笔记](#)
阅读量：363454 120 篇



[大数据学习笔记](#)

[展开](#)

个人分类

[C and C++](#)

399篇

[Java](#)

347篇

[Python](#)

191篇

Scala	65篇
Haskell	16篇
展开	
归档	
2018年11月	10篇
2018年10月	21篇
2018年9月	27篇
2018年8月	18篇
2018年7月	16篇
展开	

最新评论

IT行业技术，内推交流群

qq_40612528：大佬，我想进群学习

Cassandra学习笔记 ----

HeiShuaiGuo：可以的。就是为什么呢？好多软件都是这样需要写真实ip。

数据结构之---C语言实现括号匹配...


qq_41253659：清空栈函数貌似没用到

数据结构之---C语言实现括号匹配...


qq_24470943：请问能解释一下
Parenthesis_match函数里面算法的意思吗 不太懂

数据结构之---C语言实现括号匹配...

qq_24470943：请问主函数定义的int flag=0 0代表true 1代表false吗



大数据可视化



联系我们



微信客服



区块链大本营

 QQ客服

 kefu@csdn.net

 客服论坛

 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

 百度提供站内搜索 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