

教你使用solr搭建你的全文检索 PDF转换可能丢失图片或格式
，建议阅读原文

https://www.100test.com/kao_ti2020/290/2021_2022__E6_95_99_E4_BD_A0_E4_BD_BF_E7_c104_290410.htm Solr 是一个可供企业使用的、基于 Lucene 的开箱即用的搜索服务器。对Lucene不熟？那么建议先看看下面两篇文档：实战Lucene，第1部分：初识 Lucene

： <http://www.ibm.com/developerworks/cn/java/j-lo-lucene1/>
用Lucene加速Web搜索应用程序的开发

： <http://www.ibm.com/developerworks/cn/web/wa-lucene2/> 一、solr介绍 solr是基于Lucene Java搜索库的企业级全文搜索引擎，目前是apache的一个项目。它的官方网址

在<http://lucene.apache.org/solr/> .solr需要运行在一个servlet 容器里，例如tomcat5.5.solr在lucene的上层提供了一个基

于HTTP/XML的Web Services，我们的应用需要通过这个服务与solr进行交互。二、 solr安装和配置 关于solr的安装和配置

，这里也有两篇非常好的文档，作者同时也是 Lucene Java 项目的提交人和发言人：使用Apache Solr实现更加灵巧的搜索

： <http://www.ibm.com/developerworks/cn/java/j-solr1/index.html>

<http://www.ibm.com/developerworks/cn/java/j-solr2/index.html> 下面主要说说需要注意的地方。 Solr的安装非常简单，下载solr

的zip包后解压缩将dist目录下的war文件改名为solr.war直接复制到tomcat5.5的webapps目录即可。注意一定要设置solr的主位置。有三种方法。我采用的是在tomcat里配置java

： `comp/env/solr/home`的一个JNDI指向solr的主目录（example目录下），建立`/tomcat55/conf/Catalina/localhost/solr.xml`文件

。 value="D:/solr/solr" override="true" /> 观察这个指定的solr主位置，里面存在两个文件夹：conf和data.其中conf里存放了对solr而言最为重要的两个配置文件schema.xml和solrconfig.xml.data则用于存放索引文件。 schema.xml主要包括types、fields和其他的一些缺省设置。 solrconfig.xml用来配置Solr的一些系统属性，例如与索引和查询处理有关的一些常见的配置选项，以及缓存、扩展等等。上面的文档对这两个文件有比较详细的说明，非常容易上手。注意到schema.xml里有一个url的配置，这里将url字段作为索引文档的唯一标识符，非常重要。三、加入中文分词对全文检索而言，中文分词非常的重要，这里采用了qieqie庖丁分词（非常不错：））。集成非常的容易，我下载的是2.0.4-alpha2版本，其中它支持最多切分和按最大切分。创建自己的一个中

```
文TokenizerFactory继承自solr的BaseTokenizerFactory./** *
Created by IntelliJ IDEA. * User: ronghao * Date: 2007-11-3 * Time:
14:40:59 * 中文切词对庖丁切词的封装 */public class
ChineseTokenizerFactory extends BaseTokenizerFactory { /** * 最
多切分 默认模式 */ public static final String
MOST_WORDS_MODE = "most-words". /** * 按最大切分 */
public static final String MAX_WORD_LENGTH_MODE =
"max-word-length". private String mode = null. public void
setMode(String mode) { if
(mode==null||MOST_WORDS_MODE.equalsIgnoreCase(mode)
|| "default".equalsIgnoreCase(mode)) {
this.mode=MOST_WORDS_MODE. } else if
(MAX_WORD_LENGTH_MODE.equalsIgnoreCase(mode)) {
```

```
this.mode=MAX_WORD_LENGTH_MODE. } else { throw new  
IllegalArgumentException("不合法的析器Mode 参数设置:"  
mode). } } @Override public void init(Map args) { super.init(args).  
setMode(args.get("mode")). } public TokenStream create(Reader  
input) { return new PaodingTokenizer(input,  
PaodingMaker.make(), createTokenCollector()). } private  
TokenCollector createTokenCollector() { if(  
MOST_WORDS_MODE.equals(mode)) return new  
MostWordsTokenCollector(). if(  
MAX_WORD_LENGTH_MODE.equals(mode)) return new  
MaxWordLengthTokenCollector(). throw new Error("never  
happened"). } } 100Test 下载频道开通，各类考试题目直接下载  
。详细请访问 www.100test.com
```