# Solr使用

# Solr

## Solr简介

Solr是apache公司下的一个开源项目，是一个基于lucene的全文检索服务。它是一个服务，可以独立运行在tomcat容器中，他提供了索引和搜索两个服务。

全文检索服务

Solr具体的应用场景就是搜索引擎和站内搜索，作为web开发，使用较多的是站内搜索。

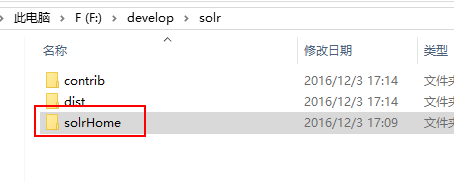
# Solr配置

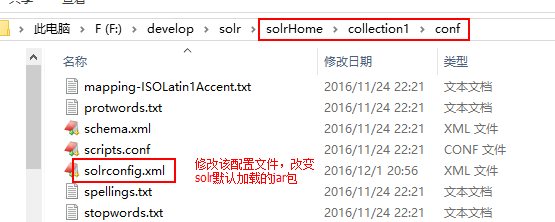
## 配置solrHome目录

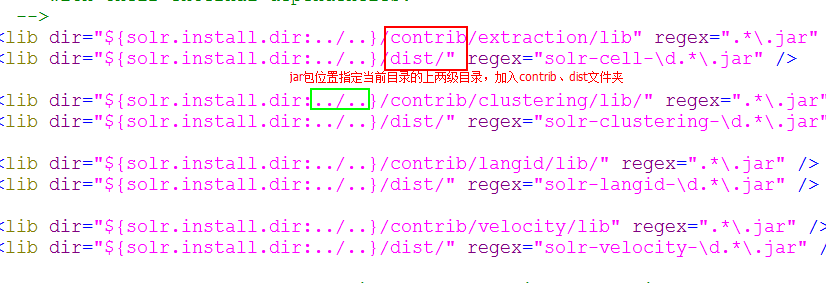
### 第一步、复制example/solr下的所有文件，放置在solrHome中



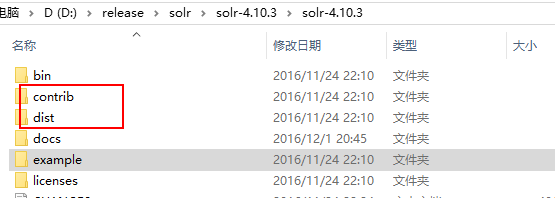
### 第二步、修改collection\conf\solrconfig.xml配置文件，改变默认加载的jar包





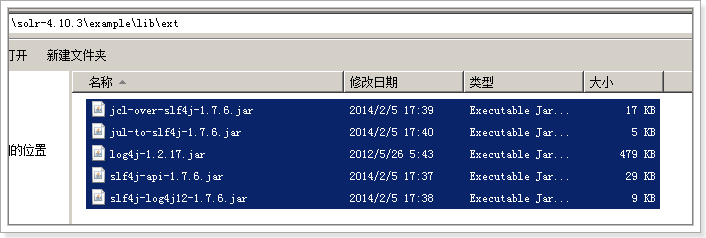


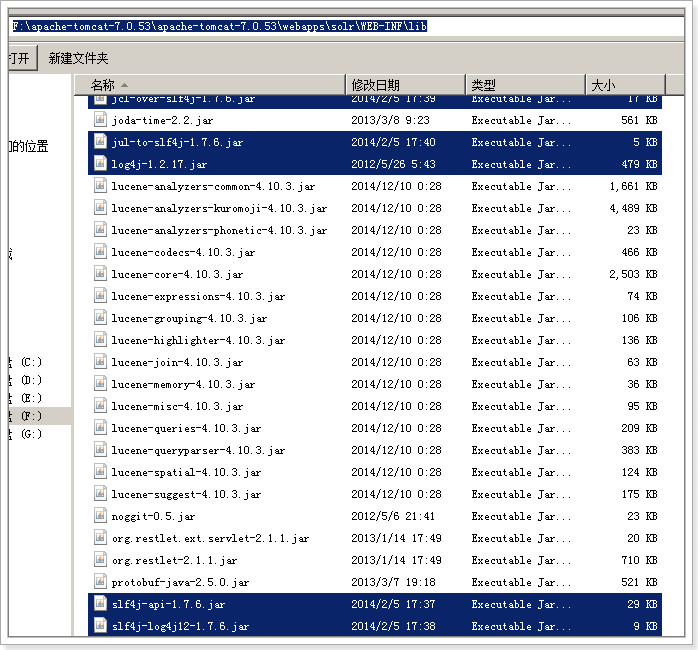
### 第三步：向/develop/solr目录中添加/solr-4.10.3/contrib、dist目录



## Solr测试工程部署

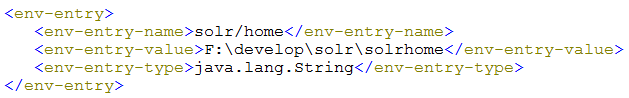
1. 将dist\solr-4.10.3.war拷贝到Tomcat的webapp目录下改名为solr.war
2. 启动tomcat后，solr.war自动解压，将原来的solr.war删除。
3. 拷贝example\lib\ext 目录下所有jar包拷贝到Tomcat的webapp\solr\WEB-INF\lib目录下





1. 修改Tomcat目录 下webapp\solr\WEB-INF\web.xml文件，如下所示：

设置Solr home



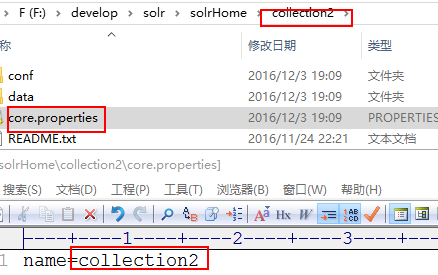
1. 拷贝log4j.properties文件

在 Tomcat下webapps\solr\WEB-INF目录中创建文件 classes文件夹，

复制Solr目录下example\resources\log4j.properties至Tomcat下webapps\solr\WEB-INF\classes目录

## 多个solrCore配置

复制collection1文件夹，改名为collection2，需要修改core.properties中的配置文件



注意：此操作会复制solrCore中的索引库

## Schema文件配置

注意：1）关于主键id，查询时，必须有该项，否则报错



2）关于copyField，配置目标域时，必须有multiValued="true"

## 安装中文分词器

* + 1. IKAnalyzer部署

只需要将fieldType中的<analyzer></analyzer>中的<takenizer>中的class改为自己想要的即可

1.将IKAnalyzer2012FF\_u1.jar拷贝到 solr\contrib\IKAnalyzer\lib下，在solrconfig.xml中添加：



2.在Tomcat的webapps/solr/WEB-INF/下创建classes目录

3.将IKAnalyzer.cfg.xml、ext\_stopword.dic mydict.dic copy到 Tomcat/webapps/solr/WEB-INF/classes下

注意：ext\_stopword.dic 和mydict.dic必须保存成无BOM的utf-8类型。

* + 1. 修改schema.xml文件

FieldType

首先需要在types结点内定义一个FieldType子结点，包括name,class,等参数，name就是这个FieldType的名称，class指向org.apache.solr.analysis包里面对应的class名称，用来定义这个类型的行为。在FieldType定义的时候最重要的就是定义这个类型的数据在建立索引和进行查询的时候要使用的分析器analyzer,包括分词和过滤

修改Solr的schema.xml文件，添加FieldType：

<!-- IKAnalyzer-->

<fieldType name="text\_ik" class="solr.TextField">

<analyzer type="index" isMaxWordLength="false" class="org.wltea.analyzer.lucene.IKAnalyzer"/>

<analyzer type="query" isMaxWordLength="true" class="org.wltea.analyzer.lucene.IKAnalyzer"/>

</fieldType>

在index或query分别配置isMaxWordLength，查询采用IK自己的最大分词法设置为true，索引则采用它的细粒度分词法设置为false。

Field：

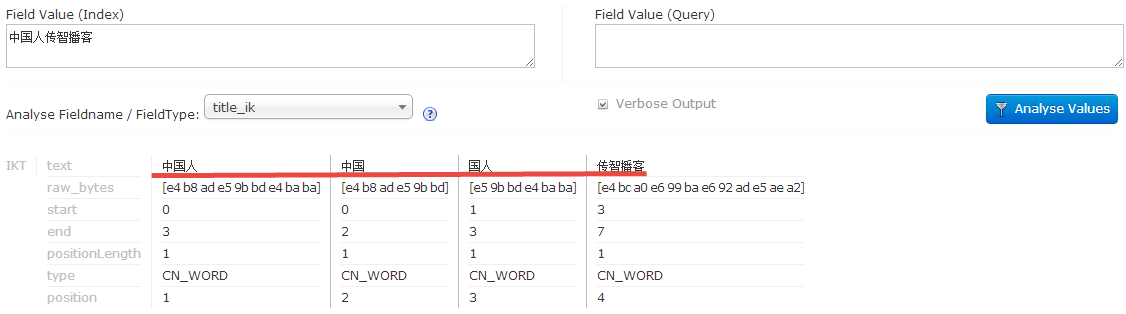
FieldType定义好后就可以在fields结点内定义具体的field，filed定义包括name,type（即FieldType）,indexed（是否被索引）,stored（是否被储存），multiValued（是否有多个值）等

<!--IKAnalyzer Field-->

<field name="title\_ik" type="text\_ik" indexed="true" stored="true" />

<field name="content\_ik" type="text\_ik" indexed="true" stored="false" multiValued="true"/>

* + 1. 测试



# 项目案例

## 配置数据导入处理器dataimport handler

目的：为关系型数据库批量建立索引到solr索引库中。

前提准备：我们需要定义自定义的field域。（参考中文分词器）

具体实现步骤：

安装dataimport-Handler从关系数据库将数据导入到索引库。

### 第一步：加载dataimport的jar包

将solr\dist\ solr-dataimporthandler-4.10.3.jar拷贝至solr\contrib\dataimporthandler下，

将mysql数据库驱动的jar拷贝至solr\contrib\databaseDriver下

在solrconfig.xml加载以上jar包：



### 第二步：编辑data-config.xml文件，存放在SolrCore的conf目录

<?xml version="1.0" encoding="UTF-8" ?>

<dataConfig>

<dataSource type="JdbcDataSource"

driver="com.mysql.jdbc.Driver"

url="jdbc:mysql://localhost:3306/lucene"

user="root"

password="mysql"/>

<document>

<entity name="product" query="SELECT pid,name, catalog ,catalog\_name,price,description,picture FROM products ">

<field column="pid" name="id"/>

<field column="name" name="product\_name"/>

<field column="catalog " name="product\_catalog "/>

<field column="catalog\_name" name="product\_catalog\_name"/>

<field column="price" name="product\_price"/>

<field column="description" name="product\_description"/>

<field column="picture" name="product\_picture"/>

</entity>

</document>

</dataConfig>

说明：

<field column="pid" name="id"/>必须有一个id域，这里使用Solr默认的id域，域值是从关系数据库查询的pid列值。

下边以“product\_”开头的Field都是在schema.xml中自定义的商品信息Field。

### 第三步：修改solrconfig.xml，添加requestHandler：

<requestHandler name="/dataimport"

class="org.apache.solr.handler.dataimport.DataImportHandler">

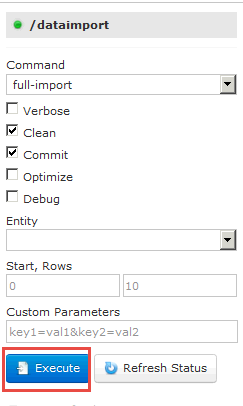
<lst name="defaults">

<str name="config">data-config.xml</str>

</lst>

</requestHandler>

### 第四步：重启Tomcat，进入管理界面--》SolrCore--》dataimport下执行导入



第五步：查看导入结果

进入管理界面--》SolrCore--》dataimport下

