# 一、相关工具

tomcat 8.5.6

solr 6.2.1

solr-dataimport-scheduler 1.2.jar

IKAnalyzer-lucene5x-1.0.jar

mysql-connector-java-5.1.22-bin.jar

pinyin4j-2.5.0.jar

pinyin-analyzer-lucene5x-1.0.ja

# 二、构建方法

## 1、tomcat集成solr

第一步：

solr-6.2.1\server\solr-webapp\webapp文件夹拷贝到apache-tomcat-8.5.6\webapps文件夹中改名为solr。

第二步：

将solr-6.2.1\server\lib\ext下的所有jar包拷贝到

apache-tomcat-8.5.6\webapps\solr\WEB-INF\lib

第三步：

将solr-6.2.1\server\resources下的log4j.properties配置文件拷贝到

apache-tomcat-8.5.6\webapps\solr\WEB-INF\classes，如果WEB-INF下没有classes文件那么 就创建一个classes文件夹。

第四步：

将solr-6.2.1\server\solr拷贝到apache-tomcat-8.5.6\webapps\solr中，改名为solrhome

修改apache-tomcat-8.5.6\webapps\solr\WEB-INF\web.xml

<env-entry>

<env-entry-name>solr/home</env-entry-name>

<!--solrhome的地址-->

<env-entry-value>

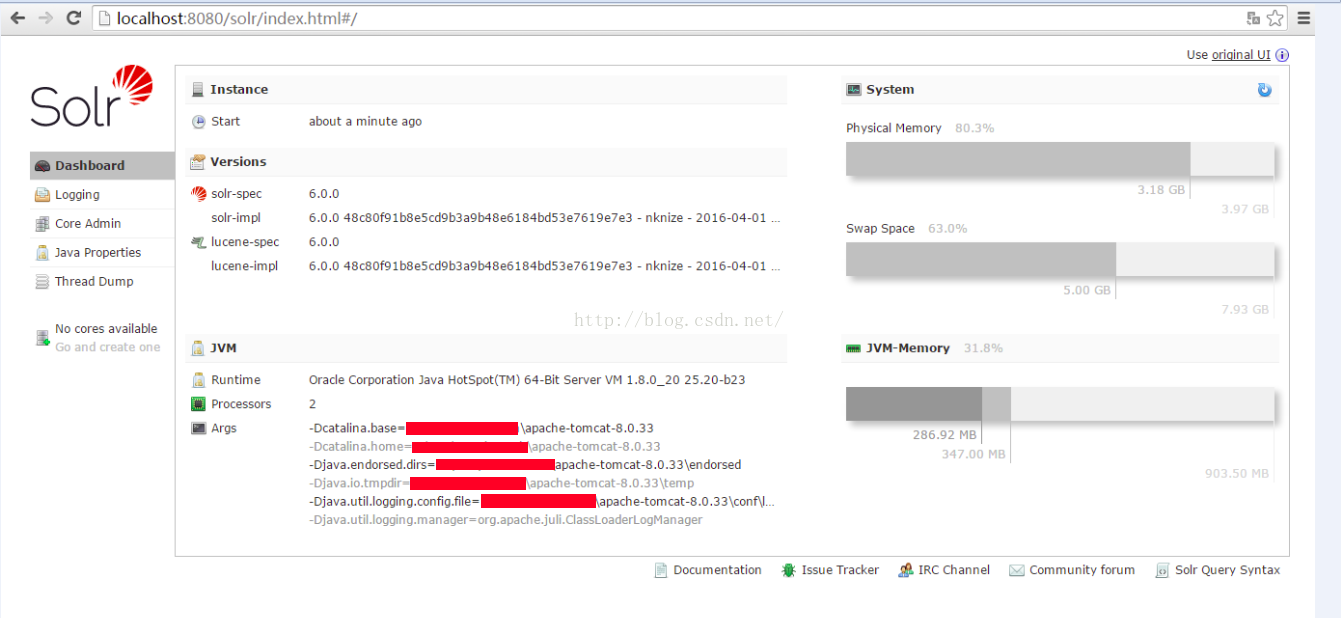
C:\work\zhongxin\sousuo\solr\apache-tomcat-8.5.6\webapps\solr\solrhome

</env-entry-value>

<env-entry-type>java.lang.String</env-entry-type>

</env-entry>

运行tomcat，运行成功后访问：http://localhost:8080/solr/index.html 即可得到如下界面：



## 2、创建core

修改solrhome下的solr.xml文件，注解掉zookeeper搭建集群配置。

在solrhome下创建goods文件夹

在goods文件夹中添加core.properties配置，内容如下

name=goods

这个name的值实际上就core的名称

将solr-6.2.1\example\example-DIH\solr\solr下的conf文件夹拷贝到goods文件夹下。

solr-5.0 以上默认对schema的管理是使用managed-schema，不能手动修改，需要使用Schema Restful的API操作。如果要想手动修改配置，把【conf】文件夹中managed-schema拷贝一份修改为schema.xml，在solrconfig.xml中加入如下：

<!-- 解除managed-schema管理模式 start -->

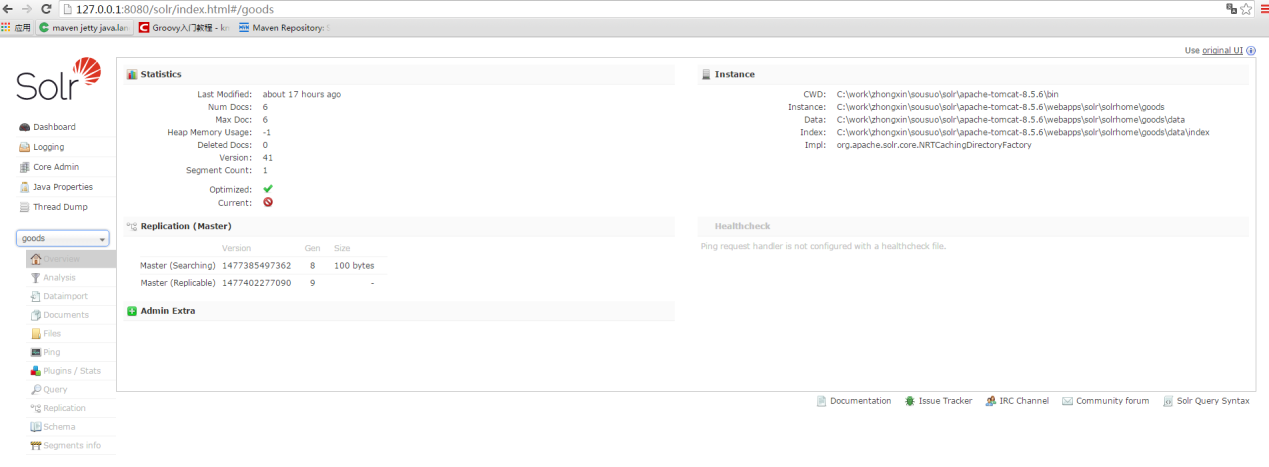
<schemaFactory class="ClassicIndexSchemaFactory"/>

<!-- 解除managed-schema管理模式 end -->

将solr-6.2.1\dist下的solr-dataimporthandler-6.2.1.jar和solr-dataimporthandler-extras-6.2.1.jar

拷贝到apache-tomcat-8.5.6\webapps\solr\WEB-INF\lib下

启动tomcat



## 3、加入索引（需要根据实际业务修改）

1. 链接数据库

修改apache-tomcat-8.5.6\webapps\solr\solrhome\goods\conf\solr-data-config.xml

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

<dataConfig>

<dataSource type="JdbcDataSource" driver="com.mysql.jdbc.Driver" url="jdbc:mysql://172.16.75.48:3306/solrtest" user="root" password="123456" batchSize="-1" />

　　<document name="testDoc">

<entity name="user" pk="id"

query="select \* from user">

　　　 <field column="id" name="id"/>

　　　 <field column="userName" name="userName"/>

<field column="userAge" name="userAge"/>

<field column="userAddress" name="userAddress"/>

　　　 </entity>

　　</document>

</dataConfig>

1. 修改schema.xml,加入相应field

<field name="id" type="int" indexed="true" stored="true" required="true" multiValued="false" />

<!--<field name="id" type="int" indexed="true" stored="true" required="true" multiValued="false"/> -->

<field name="userName" type="text\_general" indexed="true" stored="true" />

<field name="userAge" type="int" indexed="true" stored="true" />

<field name="userAddress" type="text\_general" indexed="true" stored="true" />

1. 引入mysql库

拷贝mysql-connector-java-5.1.22-bin.jar到apache-tomcat-8.5.6\webapps\solr\WEB-INF\lib



## 4、定时更新索引（需要根据实际业务修改）

1. 更新数据库

加入updateTime用来判定记录是否需要重新索引。

新增一个字段updateTime，类型为timestamp，默认值为CURRENT\_TIMESTAMP。

加入isDeleted用来判断是否删除

1. 设置更新策略

deltaQuery：增量索引查询主键ID 注意这个只能返回ID字段

deltaImportQuery：增量索引查询导入的数据

deletedPkQuery：增量索引删除主键ID查询 注意这个只能返回ID字段

修改apache-tomcat-8.5.6\webapps\solr\solrhome\goods\conf\solr-data-config.xml

<entity name="user" pk="id"

query="select \* from user where isDeleted=0"

deltaImportQuery="select \* from user where id='${dih.delta.id}'"

deltaQuery="select id from user where updateTime> '${dataimporter.last\_index\_time}' and isDeleted=0"

deletedPkQuery="select id from user where isDeleted=1">

1. 设置更新时间

apache-solr-dataimportscheduler.jar放到apache-tomcat-8.5.6\webapps\solr\WEB-INF\lib目录下。solr6.1以上必须使用1.2版本，否则启动失败。

修改solr的WEB-INF目录下面的web.xml文件：为<web-app>元素添加一个子元素<listener> <listener-class>org.apache.solr.handler.dataimport.scheduler.ApplicationListener</listener-class>

</listener>

在apache-tomcat-8.5.6\webapps\solr\solrhome\conf新建文件dataimport.properties



## 5、中文分词

IKAnalyzer-lucene5x-1.0.jar放到apache-tomcat-8.5.6\webapps\solr\WEB-INF\lib

在schema.xml中加入

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

<analyzer type="index">

<tokenizer class="org.wltea.analyzer.lucene.IKAnalyzerSolrFactory" isMaxWordLength="false"/>

</analyzer>

<analyzer type="query">

<tokenizer class="org.wltea.analyzer.lucene.IKAnalyzerSolrFactory" isMaxWordLength="true"/>

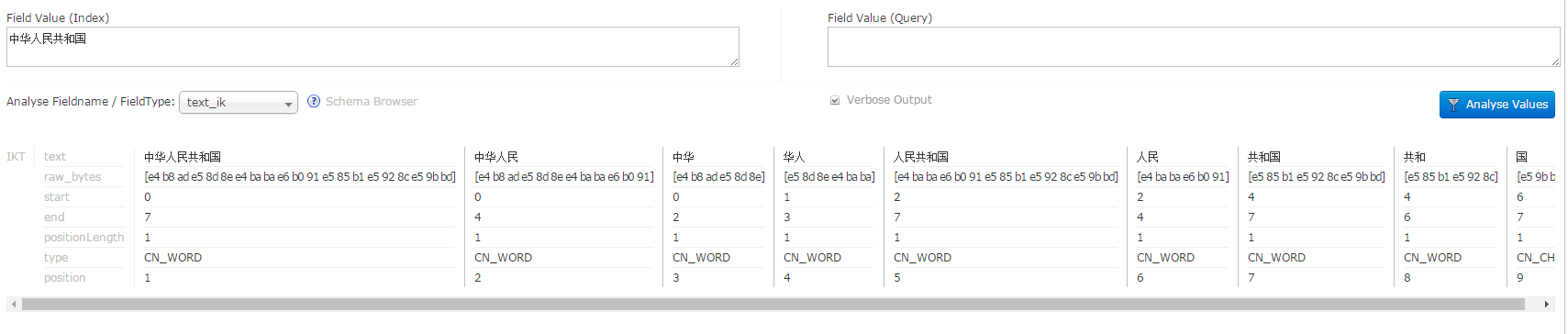
<filter class="solr.LowerCaseFilterFactory" />

</analyzer>

</fieldType>

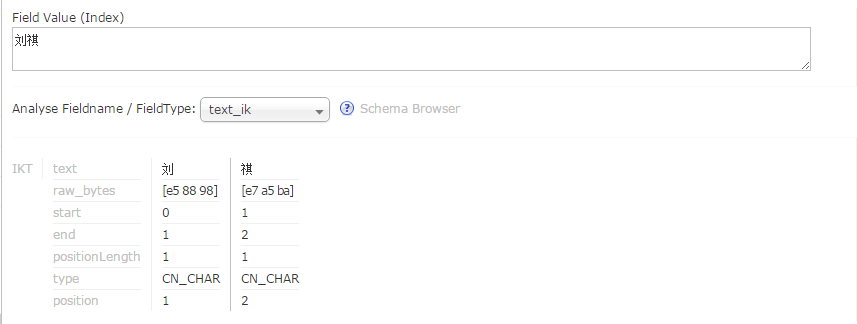
需要中文分词的字段设置成text\_ik,如

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



## 6、扩展中文分词词库

扩展前的效果

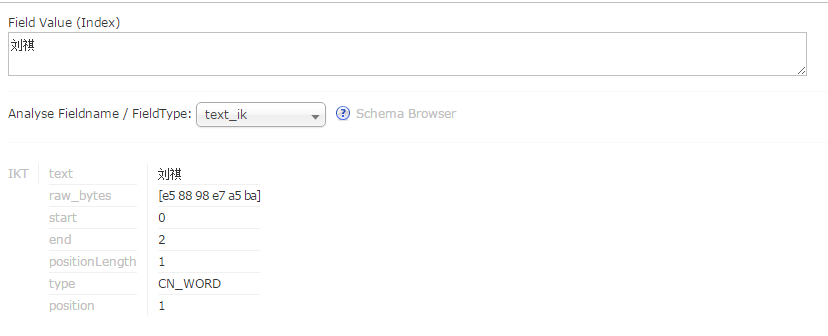


将IKAnalyzer.cfg.xml放在apache-tomcat-8.5.6\webapps\solr\WEB-INF\classes，设置扩展文件

<entry key="ext\_dict">product.dic;</entry>

添加product.dic在apache-tomcat-8.5.6\webapps\solr\WEB-INF\classes；编码方式设置成UTF-8 无BOM的编码方式

在product.dic加入“刘祺”，重启tomcat

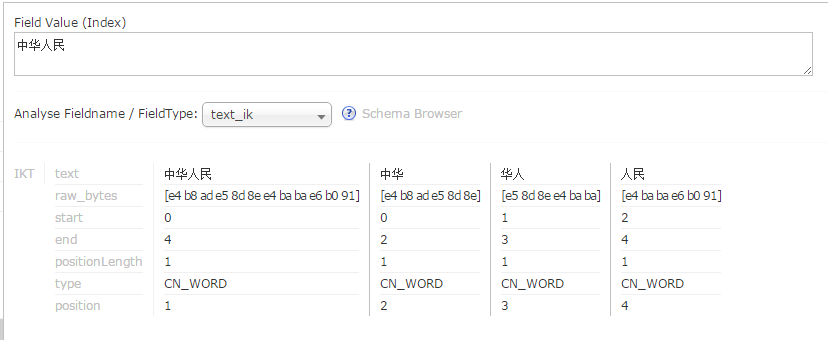


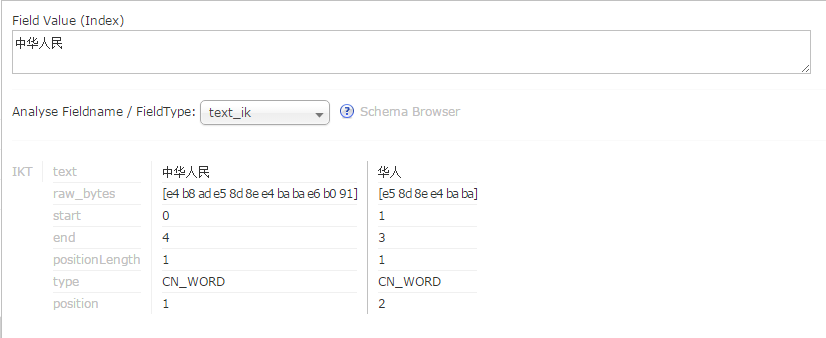
配置停止词典：对加入的词取消分词

IKAnalyzer.cfg.xml中加入<entry key="ext\_stopwords">stopword.dic;</entry>

添加stopword.dic在apache-tomcat-8.5.6\webapps\solr\WEB-INF\classes；编码方式设置成UTF-8 无BOM的编码方式

在stopword.dic中加入“中华”“人民”，重启tomcat





## 7、字母全部转换成小写

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

<analyzer type="index">

<tokenizer class="org.wltea.analyzer.lucene.IKAnalyzerSolrFactory" isMaxWordLength="false"/>

<filter class="solr.LowerCaseFilterFactory" />

</analyzer>

<analyzer type="query">

<tokenizer class="org.wltea.analyzer.lucene.IKAnalyzerSolrFactory" isMaxWordLength="true"/>

<filter class="solr.LowerCaseFilterFactory" />

</analyzer>

</fieldType>

## 8、汉字拼音索引

在中文分词的基础上，建立全拼和首字母索引。

pinyin4j-2.5.0.jar、pinyin-analyzer-lucene5x-1.0.jar加入到apache-tomcat-8.5.6\webapps\solr\WEB-INF\lib

在schema.xml配置

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

<analyzer type="index">

<tokenizer class="org.wltea.analyzer.lucene.IKAnalyzerSolrFactory" isMaxWordLength="false"/>

<filter class="org.apache.lucene.analysis.pinyin.solr5x.PinyinTokenFilterFactory" minTermLength="1"/>

<filter class="solr.LowerCaseFilterFactory" />

<filter class="solr.RemoveDuplicatesTokenFilterFactory"/>

</analyzer>

<analyzer type="query">

<tokenizer class="org.wltea.analyzer.lucene.IKAnalyzerSolrFactory" isMaxWordLength="true"/>

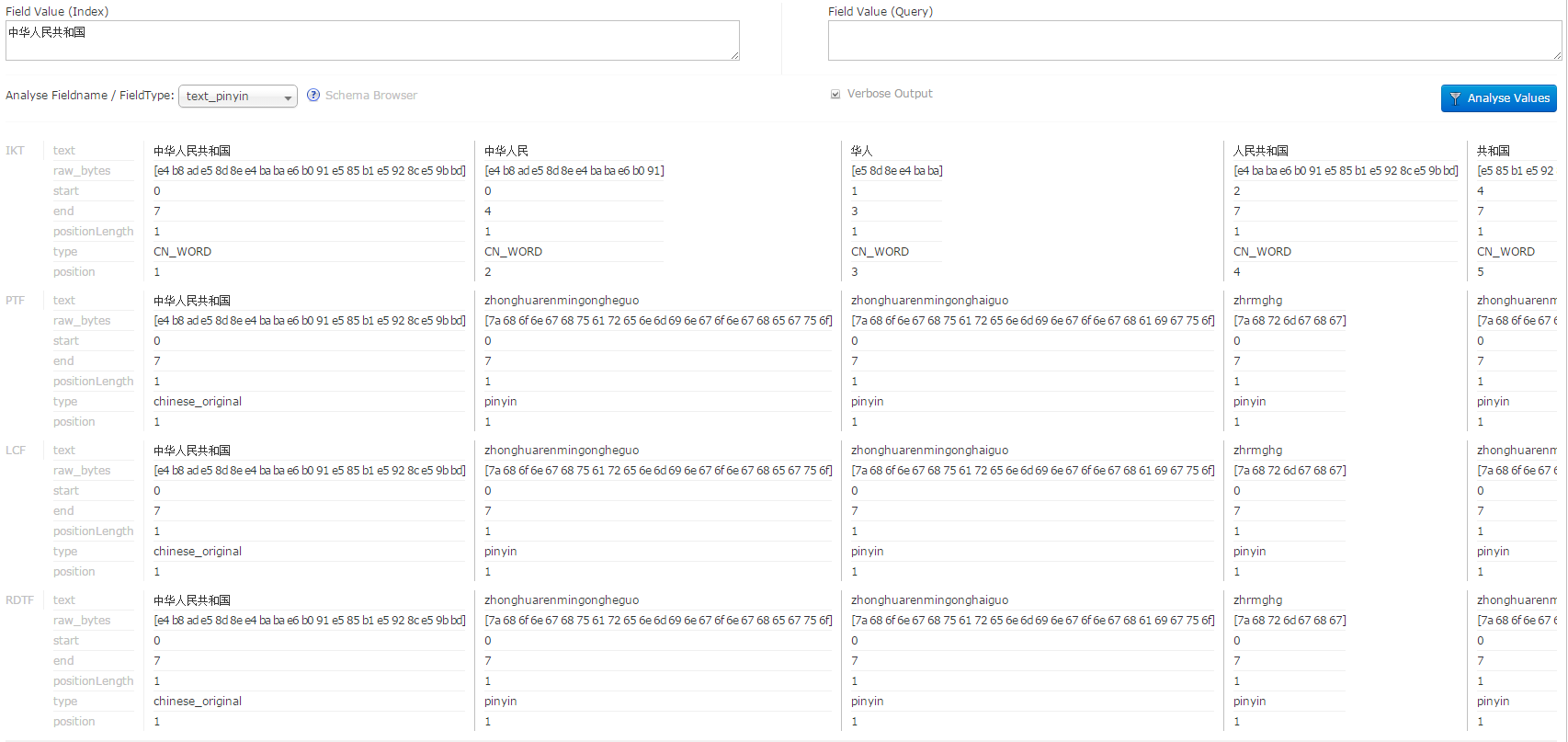
<filter class="solr.LowerCaseFilterFactory" />

</analyzer>

</fieldType>

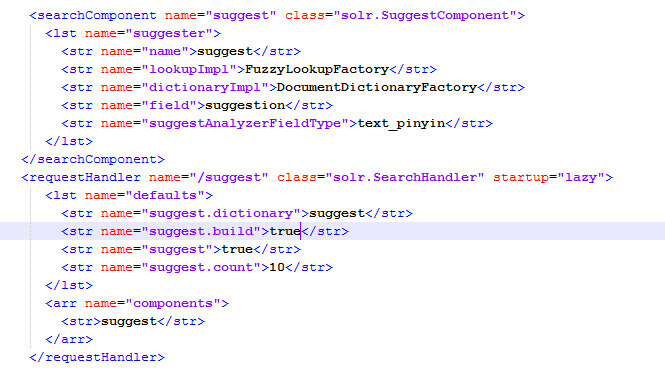
需要拼音搜索的字段设置成text\_pinyin,如

<field name="userName" type="text\_pinyin" indexed="true" stored="true" />



## 9、搜索提示词列表

在solrconfig.xml中加入



在schema.xml中加入

<field name="suggestion" type="text\_pinyin" indexed="true" stored="true" termVectors="true" multiValued="true" />

<copyField source="userName" dest="suggestion" />

把userName加入到suggestion列表中。

