**ElasticSearch IK分词器扩展说明**

简介：

本文主要描述IK分词器的扩展，当前扩展从缓存或DB加载需要指定过滤的词语。并简要说明设计思路以及配合全棉时代官网商品标签关联。

1. 源码扩展
2. Github下载elasticsearch-analysis-ik-6.1.1

<https://github.com/medcl/elasticsearch-analysis-ik/tree/v6.1.1，选择6.1.1>版本

1. 导入eclipse，maven工程，添加jedis和mysql的依赖包，pom文件添加：

<!-- customize redis and jdbc -->

<dependency>

<groupId>redis.clients</groupId>

<artifactId>jedis</artifactId>

<version>${jedis.version}</version>

<exclusions>

<exclusion>

<groupId>org.apache.commons</groupId>

<artifactId>commons-pool2</artifactId>

</exclusion>

</exclusions>

</dependency>

<dependency>

<groupId>org.apache.commons</groupId>

<artifactId>commons-pool2</artifactId>

<version>2.4.2</version>

</dependency>

<dependency>

<groupId>mysql</groupId>

<artifactId>mysql-connector-java</artifactId>

<version>5.1.14</version>

</dependency>

<dependency>

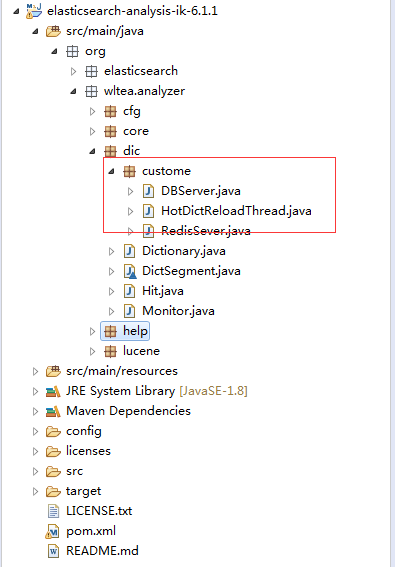
<groupId>com.alibaba</groupId>

<artifactId>druid</artifactId>

<version>1.0.29</version>

</dependency>

1. 添加org.wltea.analyzer.dic.custome包，并添加DBServer.java、RedisSever.java、HotDictReloadThread.java



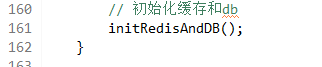
DBServer.java主要是数据库操作

RedisServer.java主要是对缓存操作

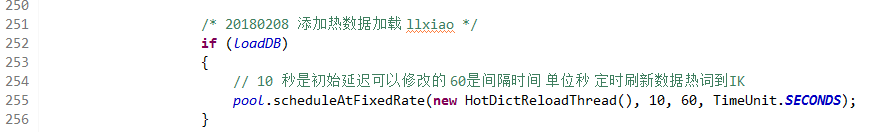
HotDictReloadThread.java定时刷新热词到IK分词器

1. 扩展org.wltea.analyzer.dic.Dictionary，修改点如下：

161行初始化缓存和DB连接，调用org.wltea.analyzer.dic.Dictionary.initRedisAndDB()方法

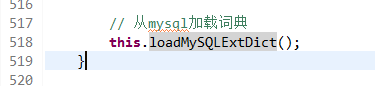


251行添加热加载，定时刷新缓存或DB的热词到IK分词器中

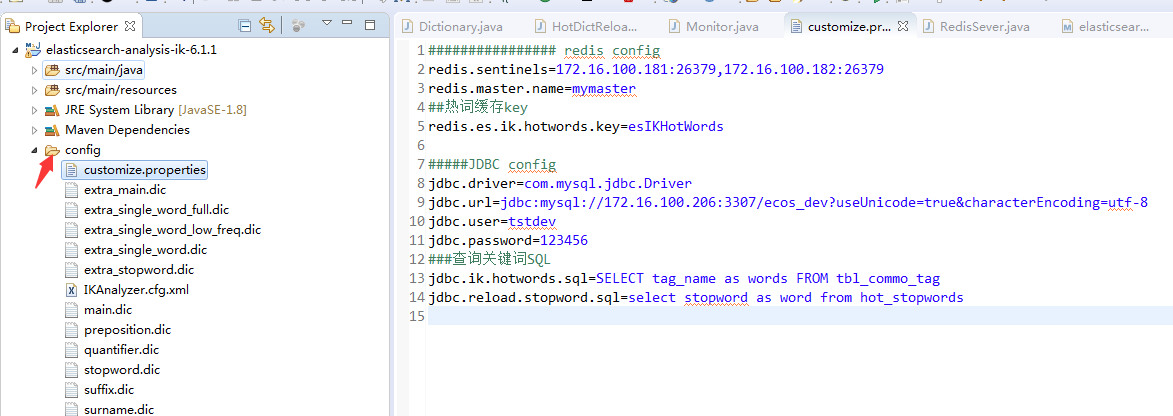


571行添加从缓存或mysql加载热词，

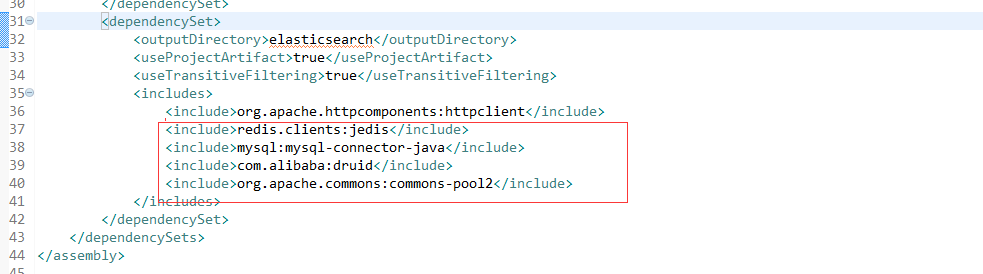
调用org.wltea.analyzer.dic.Dictionary.loadMySQLExtDict()方法。



1. 在config目录下添加customize.properties文件，增加缓存和DB的配置内容如下：

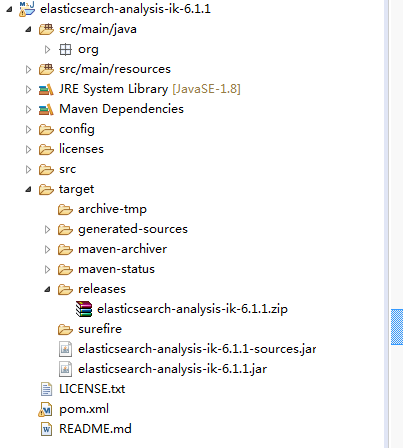


1. 修改打包文件添加依赖包 src/main/assemblies/plugin.xml文件

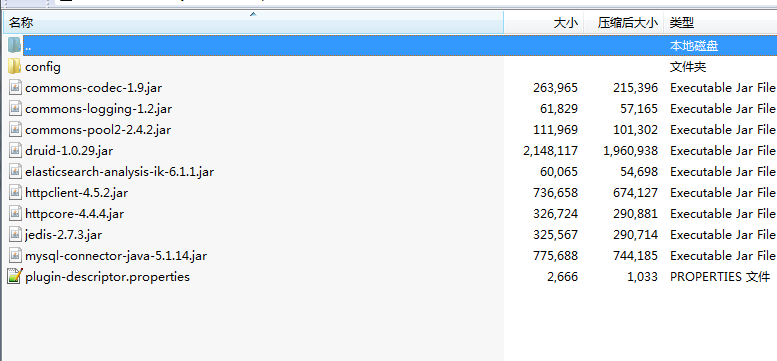


tips:打包成 zip的时候自动添加依赖包

1. 打包，mvn clean package，会在target/releases目录下看到 ik的zip包



Zip包包含内容：

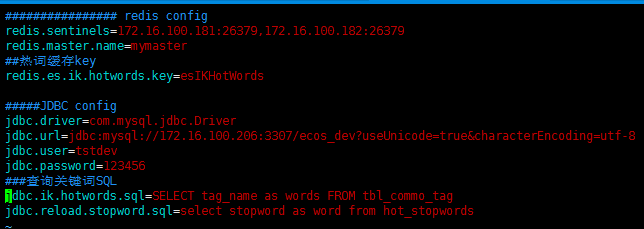


1. 新增/替换elasticsearch ik 插件为扩张后的插件

在elasticsearch\_home/plugins目录下，如有ik插件则删除，将扩展后的插件解压到plugins目录下，名称保持为ik

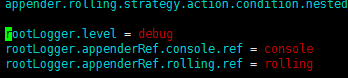


1. 修改ik/config/customize.properties文件的配置，db和redis

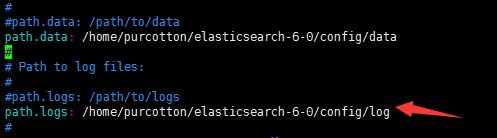


1. 日志观测

修改elasticsearch\_home/conf/log4j2.properties文件，将其设置为debug级别观测ik加载热分词日志



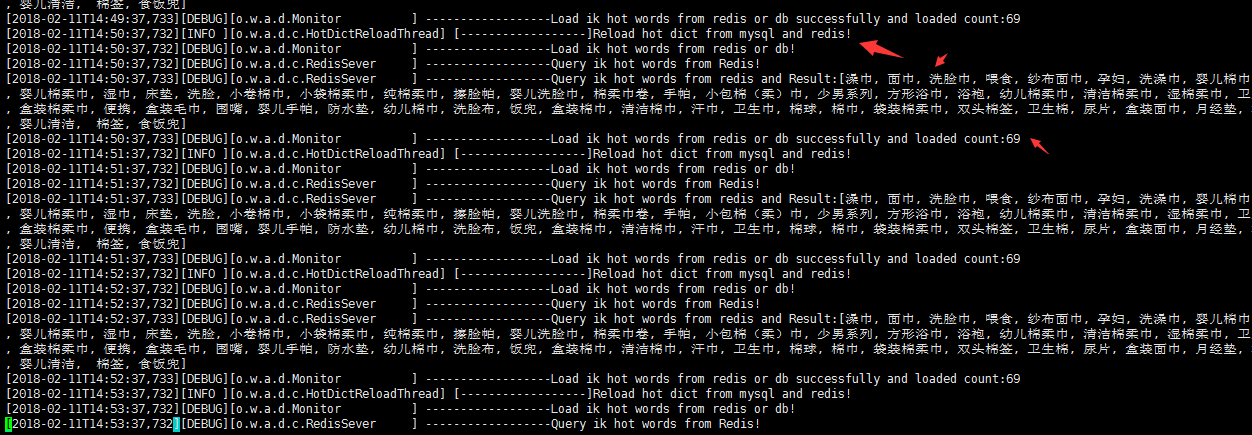
查看 elasticsearch\_home/conf/elasticsearch.yml中配置的日志路径文件夹



在该文件夹下回生成一个以集群名称为名字的日志文件



查看文件重新加载IK热词的日志信息

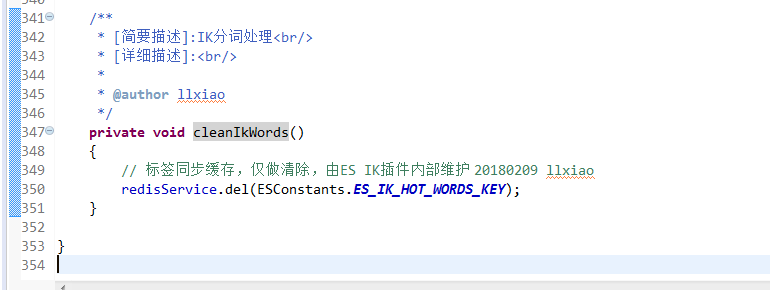


1. 官网Mgt配合更改

对商品标签进行增删改的时候，回清除缓存数据，缓存同步交给IK分词器去做同步维护，官网Mgt主要对标签的维护

增删改后调用清除缓存方法：

org.lazicats.ecos.view.commodity.viewcomponent.impl.CommonityTagViewCompinentImpl.cleanIkWords()



1. 修改JDK权限(代码内部连接Redis和DB，需要给一个权限)

[purcotton@pur181 ~]$ vi jdk1.8.0\_161/jre/lib/security/java.policy

加上如下内容：

permission java.net.SocketPermission "\*:\*","accept,connect,resolve";

permission java.security.AllPermission;

