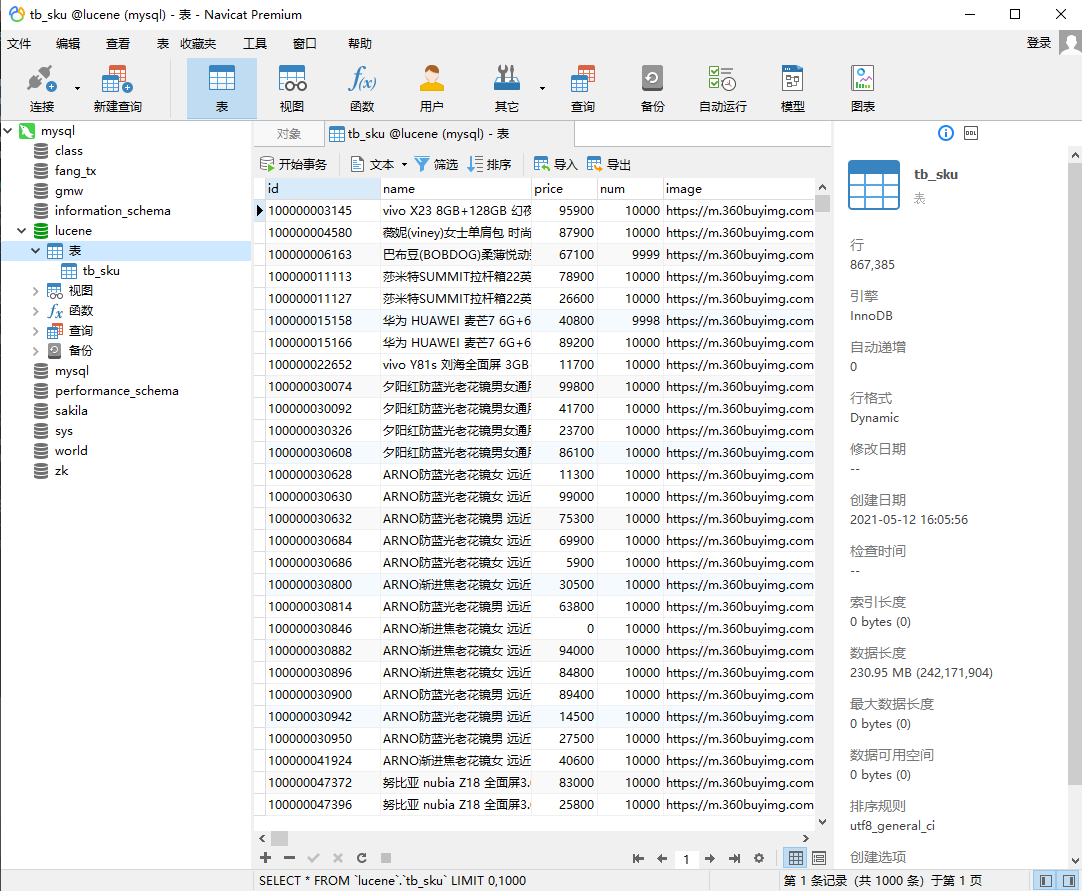
**Lucene全文检索技术项目开发记录**

1. **数据准备**

在这里，我们是准备的mysql语句，进行直接的插入，在使用之前，可以线创建一个数据库，在mysql建库语句里面插入use ‘database\_name’，才能在mysql workbench上面运行（在Navicat Premium上直接运行不需要加入，当然加入也没有影响。）

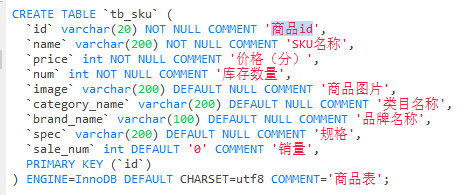
数据准备完成：



一共是938864条数据，数据的价格有问题。

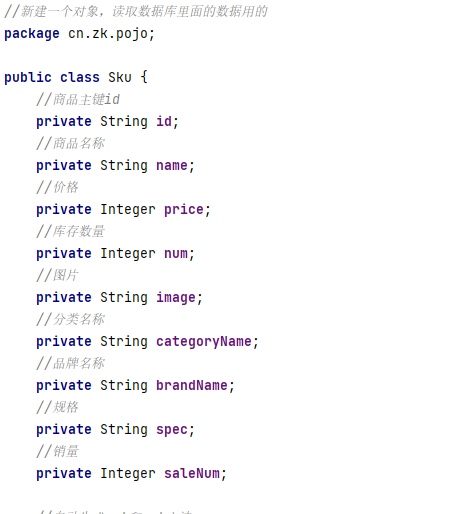
在使用mysql workbench运行时，由于mysql的建库语句太多，数据太大，导致导入的时候卡死，不建议使用。

表信息：



1. **导出数据**

（1）首先是建立一个对象，用于存放mysql导出的每一条数据。Sku类



再创建属于这个类的get和set方法，就是取和赋值（略）

（2）、在就是连接数据库，以及查询数据，获得查询的数据，将每一行数据都赋值类一个Sku类，在将每个Sku类都添加到list中。



1. 、创建一个接口函数，启动querySkuList()这个函数，将返回的list导出。

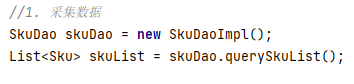
此时的list里面应该包含了938864个Sku类。

1. **创建索引库**

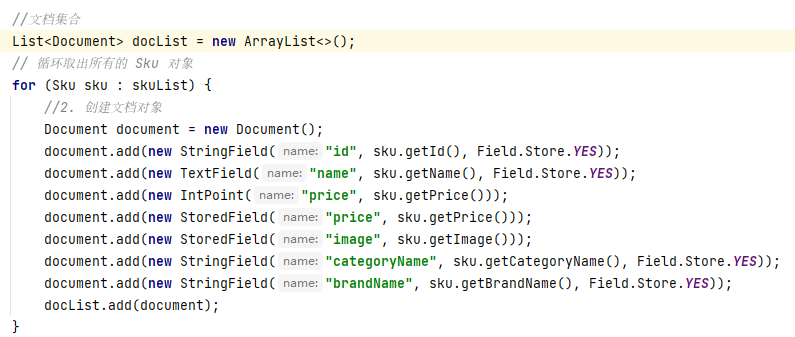
其实创建索引库就是将文本或者标题进行分词后，形成的一系列的一个文件。

在更目录下创建一个dir，将创建的索引库导入到dir中。

1. 、对mysql导出的数据进行采集



1. 使用for循环将list的Sku类都拿出来，放到Document中



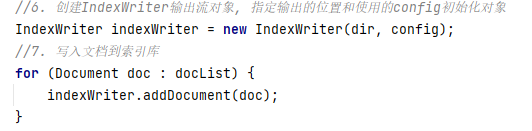
1. 、创建一个分词器



1. 、创建Directory目录对象，目录对象表示索引库的位置



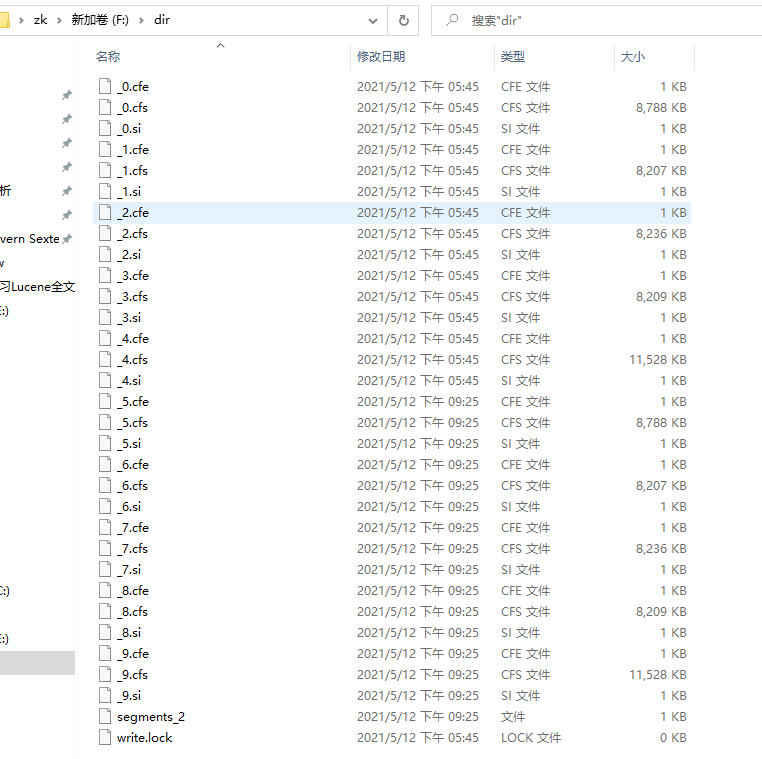
1. 、分配指定的分词器，并将结果保存到F://dir中



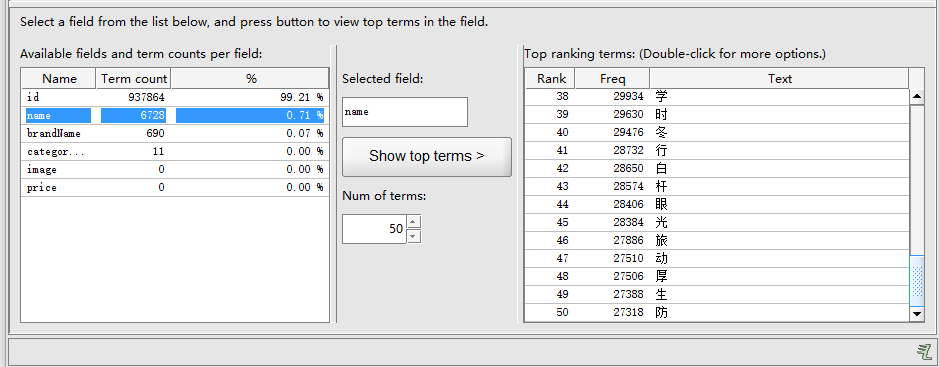
1. 、释放资源，防止内存溢出



1. 、查看索引库文件（F://dir）



可以使用luke.bat进行查看。



至此，索引库创建完成！

1. **配置文内容（pom.xml）**

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

<project xmlns="http://maven.apache.org/POM/4.0.0"

xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"

xsi:schemaLocation="http://maven.apache.org/POM/4.0.0 http://maven.apache.org/xsd/maven-4.0.0.xsd">

<modelVersion>4.0.0</modelVersion>

<!--配置文件-->

<groupId>cn.zk</groupId>

<artifactId>zk</artifactId>

<version>1.0-SNAPSHOT</version>

<properties>

<maven.compiler.source>1.8</maven.compiler.source>

<maven.compiler.target>1.8</maven.compiler.target>

<project.build.sourceEncoding>UTF-8</project.build.sourceEncoding>

<project.reporting.outputEncoding>UTF-8</project.reporting.outputEncoding>

<skipTests>true</skipTests>

</properties>

<parent>

<groupId>org.springframework.boot</groupId>

<artifactId>spring-boot-starter-parent</artifactId>

<version>2.1.4.RELEASE</version>

</parent>

<dependencies>

<dependency>

<groupId>commons-io</groupId>

<artifactId>commons-io</artifactId>

<version>2.6</version>

</dependency>

<dependency>

<groupId>org.apache.lucene</groupId>

<artifactId>lucene-core</artifactId>

<version>7.7.2</version>

</dependency>

<dependency>

<groupId>org.apache.lucene</groupId>

<artifactId>lucene-analyzers-common</artifactId>

<version>7.7.2</version>

</dependency>

<dependency>

<groupId>org.apache.lucene</groupId>

<artifactId>lucene-queryparser</artifactId>

<version>7.7.2</version>

</dependency>

<!-- 测试 -->

<dependency>

<groupId>junit</groupId>

<artifactId>junit</artifactId>

<version>4.12</version>

<scope>test</scope>

</dependency>

<!-- mysql数据库驱动 -->

<dependency>

<groupId>mysql</groupId>

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

<version>5.1.48</version>

</dependency>

<!--web起步依赖-->

<dependency>

<groupId>org.springframework.boot</groupId>

<artifactId>spring-boot-starter-web</artifactId>

</dependency>

<!-- 引入thymeleaf -->

<dependency>

<groupId>org.springframework.boot</groupId>

<artifactId>spring-boot-starter-thymeleaf</artifactId>

</dependency>

<!-- Json转换工具 -->

<dependency>

<groupId>com.alibaba</groupId>

<artifactId>fastjson</artifactId>

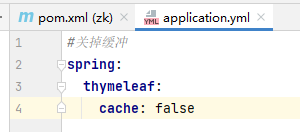
<version>1.2.51</version>

</dependency>

</dependencies>

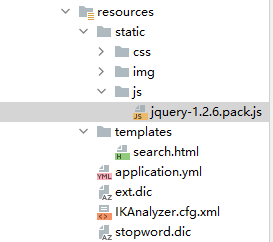
</project>

1. **配置resources文件夹**
2. 、先配置一个applicatio.yml文件



关闭缓冲，否则在几次的搜索后，回导致内存的溢出。

1. 、导入前段需要的css代码、img以及js等。目录如下：



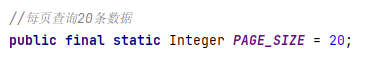
在static里面保存的是css文件，，以及要用的图片和js

Search.html是前端的页面html代码，这里不对前端多做描述。

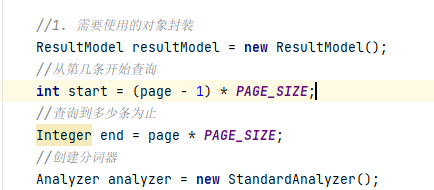
1. **实现在前端的搜索功能**
2. 、先定义一个项目的启动函数（Application）

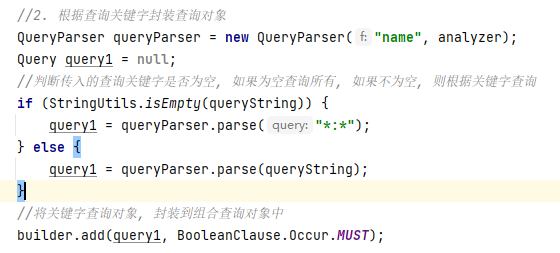


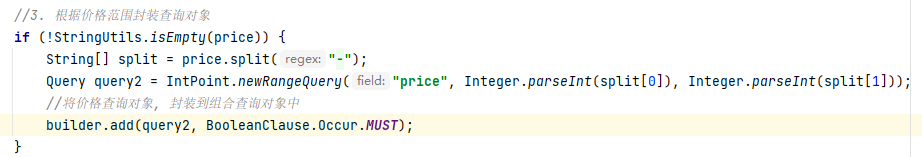
1. 、定义查询函数searchServiceImpl
2. 、定义每页的显示数量（需要与网页的一页的数量进行匹配）



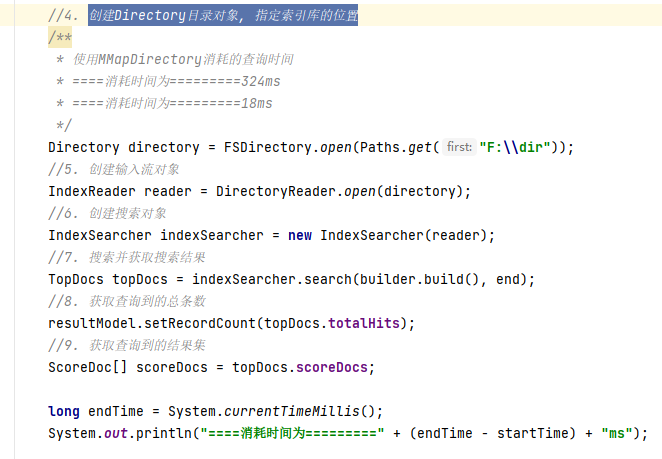
1. 、在定义一个函数query（查询函数）先要知道网页的是第几页的数据，防止重复显示，先定义开始的显示数据是第几条，在设置结束的数据是第几条



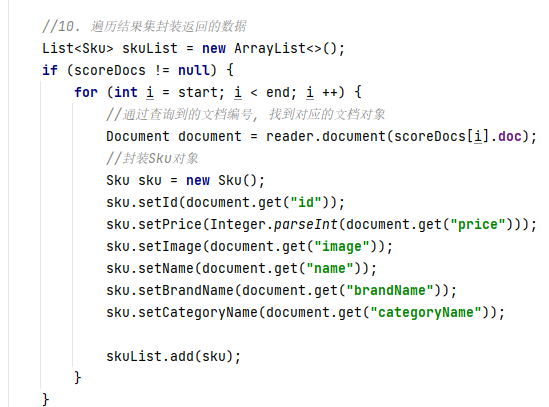
1. 、创建查询对象，导入前端输入的关键字（这里需要与前端的变量名进行同步，防止获取不到），这里需要判断用户输入的关键字是否为空，是空，这返回所有的数据库的文件，不是这等于用户输入的关键字。  
   
2. 、看用户是否用价格方面的要求，对搜索到的内容进行价格的删选



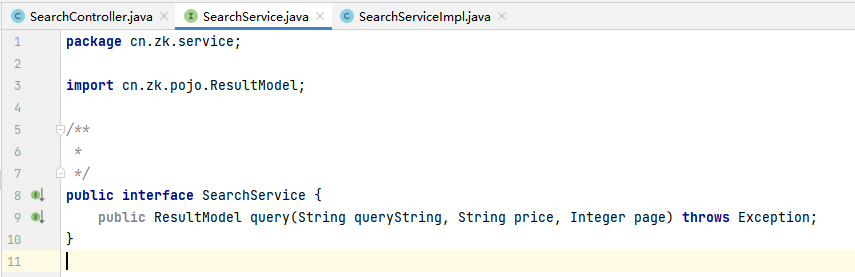
1. 、指定索引库的位置，创建相关对象，将搜索的条数返回，促成最大的页数，将查询的结果集返回，并且在后台返回消耗的时间（可以删去，对页面没有影响）



6）、遍历返回的每一条数据，将其封装到每一个Sku对象中，方便后面的页面显示

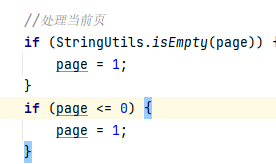


1. 、创建searchServiceImpl函数的接口



1. 、定义与页面的交互函数SearchController

1）、处理页面的当前页



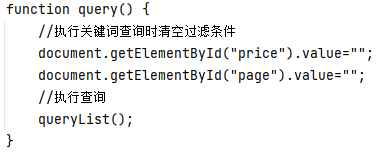
1. 、调用service查询，使用model对象去承接返回值



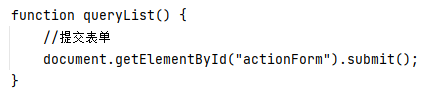
3）、将查询到的页面，返回到前端页面中，这里要注意前端页面定义的相关值的函数名，一定要一致，否则无法赋值成功。



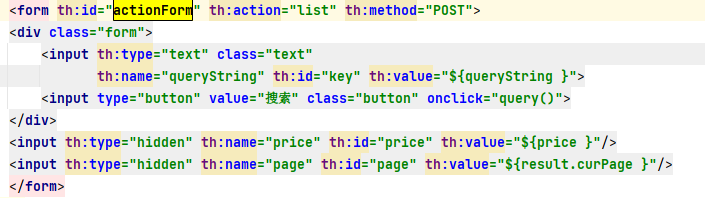
1. **简单介绍前端页面的js部分**
2. 、在查询的时候一定要清空之前的查询条件，如价格、当前的页面码等，防止查询的时候查询数据与查询的内容不符合或者查询出来的数据少的尴尬情况。



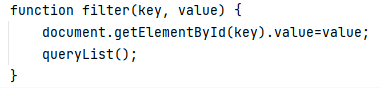
1. 、提交表单，也就是前后端交互，将用户输入的关键字导入到java代码中（这里一定要注意函数名，防止不一致，无法传递）



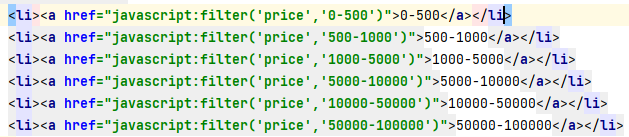
我们在页面定位搜索框，发现是以post方式发送的



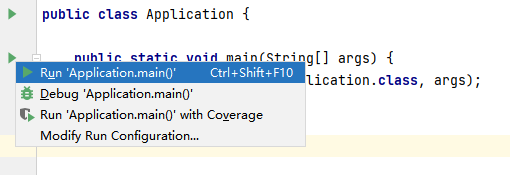
1. 、页面的价格删选，当用户点击价格删选后，将值的范围传递到java代码中，这里要注意的还是变量名，以及他传入的方式，是传入的字符串还是整数。

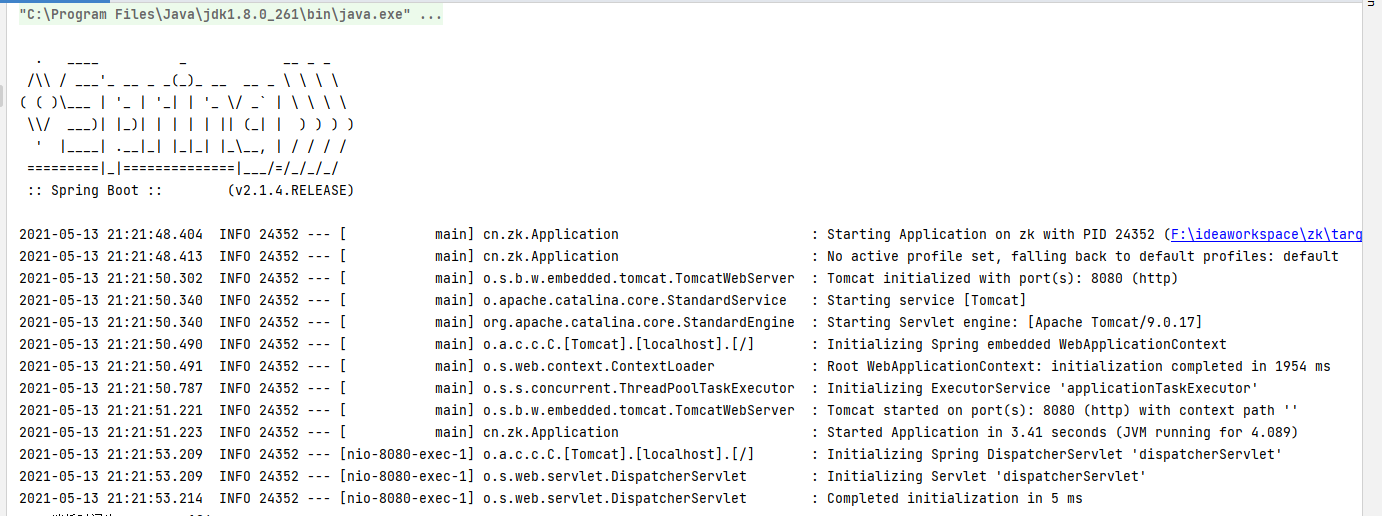


在页面找到对于的选择框所在地，查看发现是字符串，在处理的时候使用字符串处理的方法。



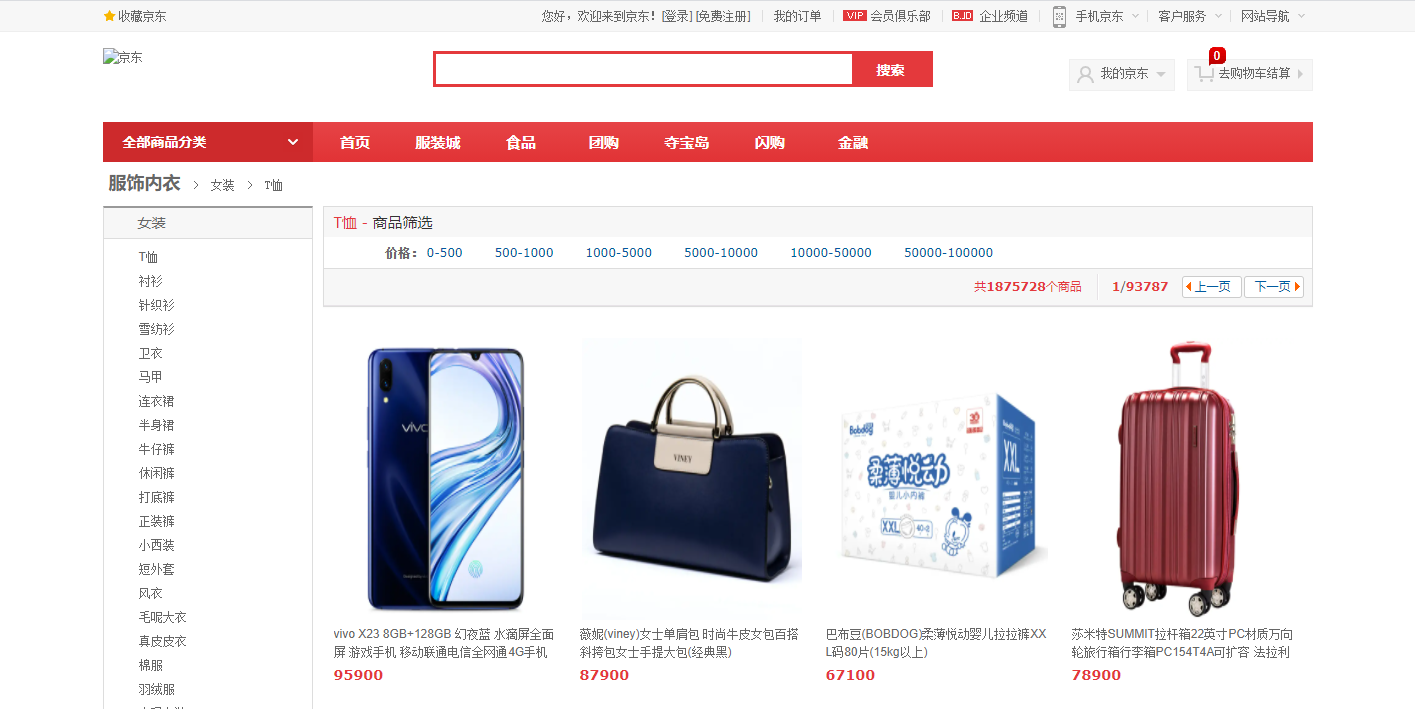
1. **项目启动截图**
2. 、在Application中启动项目



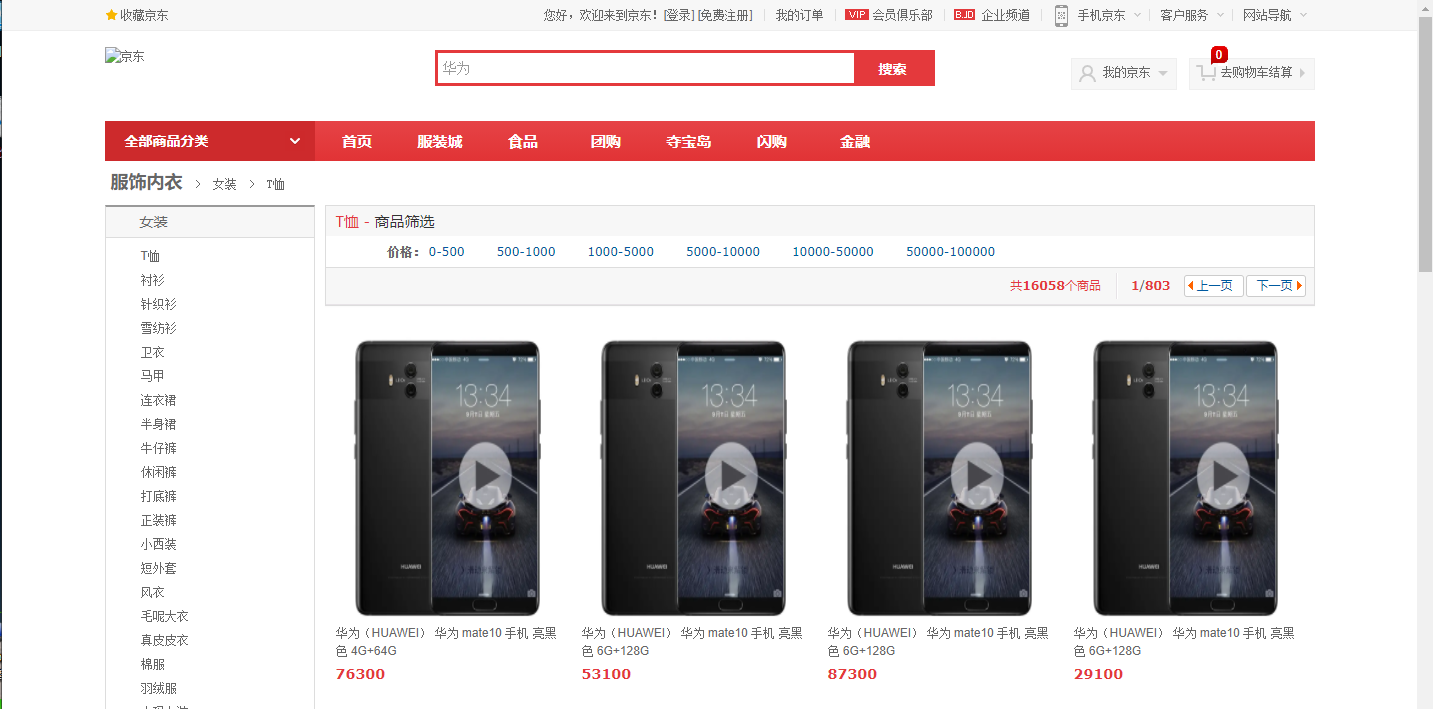


项目启动完成

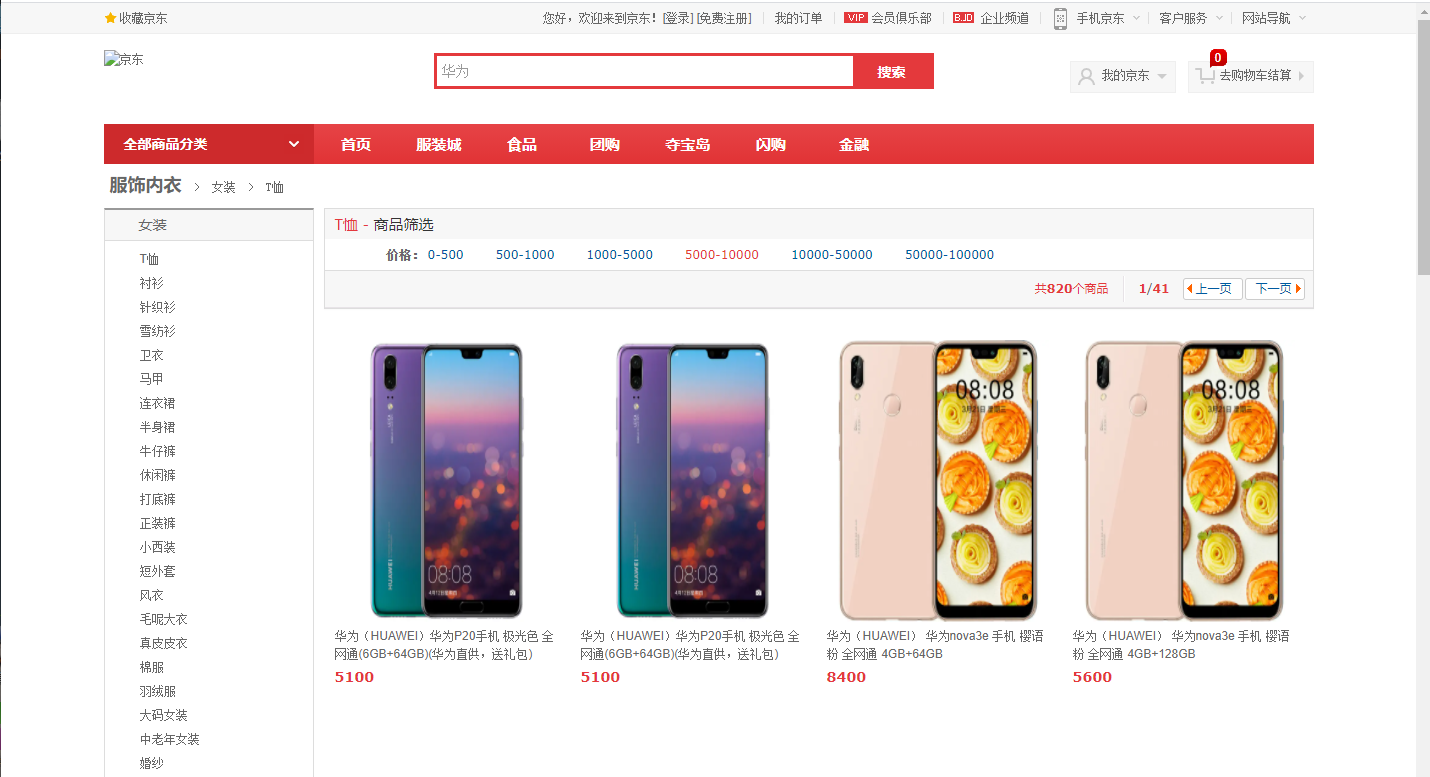
1. 、前端显示页面



1. 、搜索华为页面截图



1. 、进行价格的删选（5000-10000）



1. **项目结语**

通过这次的项目，我了解很多的Lucene全文检索技术的知识，也学习到了一些前后端加护的相关知识，在这次项目中，要注意一下几点：

1. 、在实验过程中，由于第三方的分词库无法使用，所以直接使用的standardAnalyzer()分词器
2. 、需要注意直接的dir文件的保存路径，由于索引库只创建一次，后面都是直接进行使用的，所以导入的时候，要注意路径的问题。
3. 、推荐使用本地的java包，不然在网上下载会影响项目启动的速度。

到此，学习该项目的学习记录到此结束，后附加zk.iml的配置内容，以及修改后的代码的下载地址：

<https://github.com/xx20000722/Search-Engines.git>

附zk.iml的配置

