



《爬虫期末作业》

题 目： 网络爬虫的数据爬取

分 院： 理工分院

班 级： 16计算机三班

姓 名： 张驰

学 号： 16219111311

完成日期： 2019年6月18日

温州大学瓯江学院教务部

二○一二年十一月制

# 目录

[目录 2](#_Toc7196205)

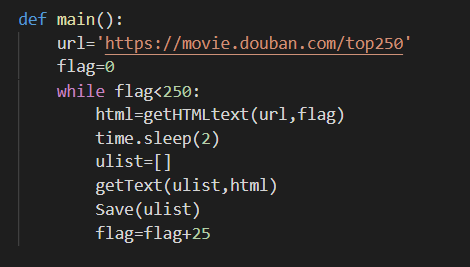
[一． 豆瓣TOP250电影网页的静态爬虫 3](#_Toc7196206)

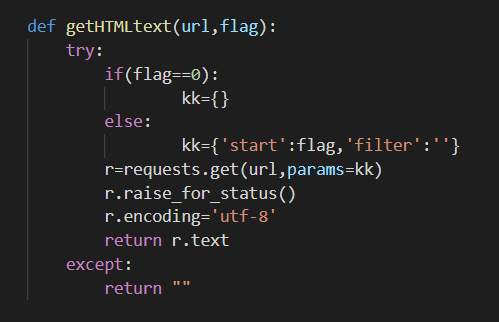
[二． 动态爬取京东搜索手机信息 8](#_Toc7196207)

三． 12306登录……………………………………………………………………………….12

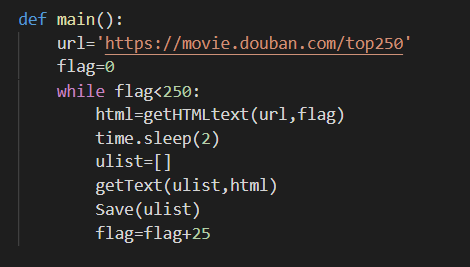
# 豆瓣TOP250电影网页的静态爬虫

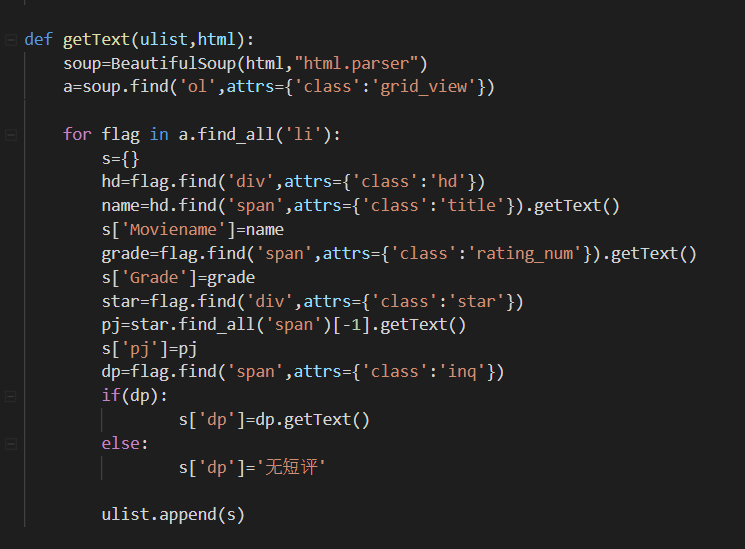
1. 用request库将网页代码爬取下来：



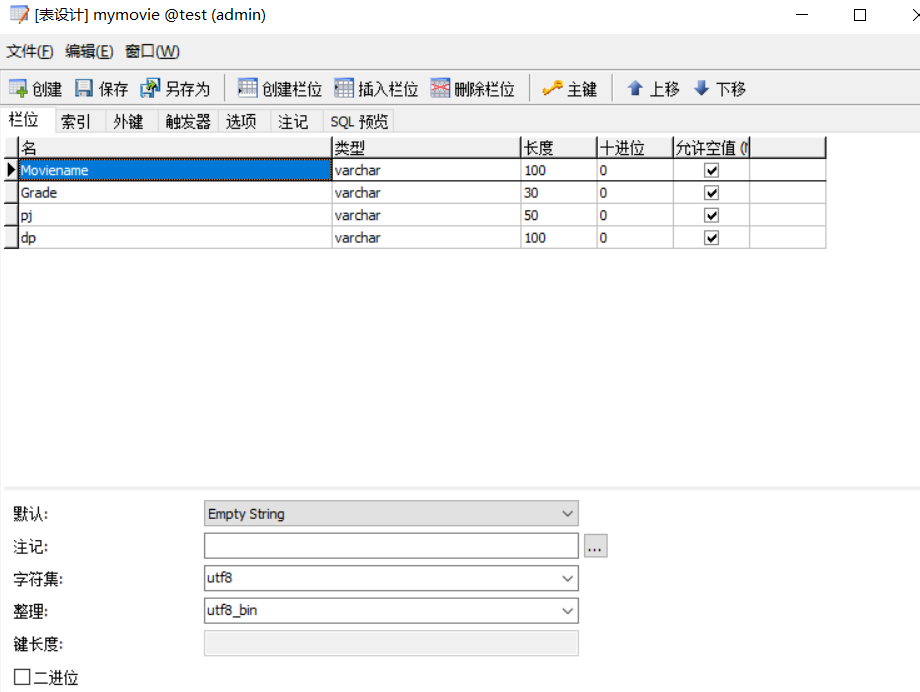


1. 用BeautifulSoup库将相应的内容爬去下来并保存到列表里：



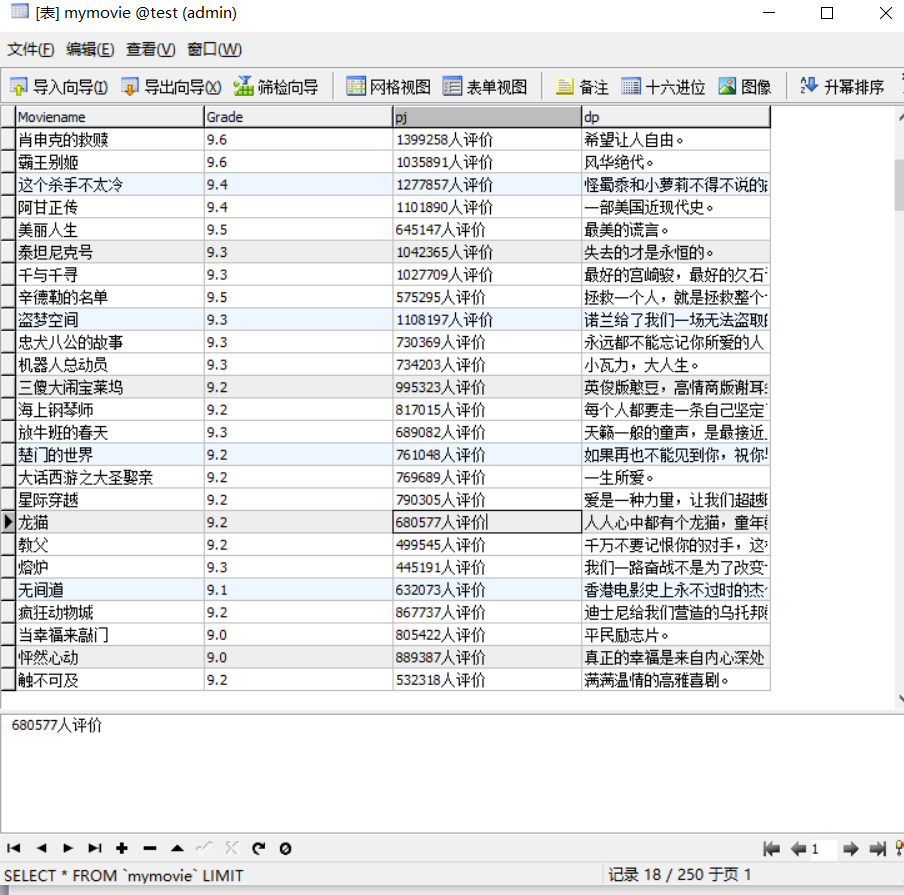


1. 将爬取的内容输入到数据库：
2. 首先在Mysql中创建数据表名为mymovie:

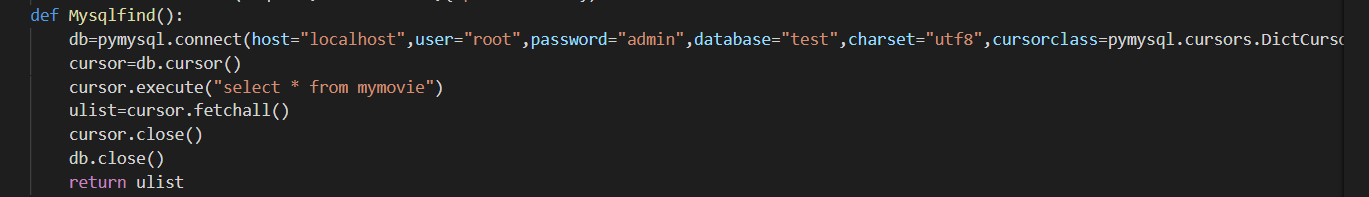


1. **用代码将列表内数据输入到数据库中：**





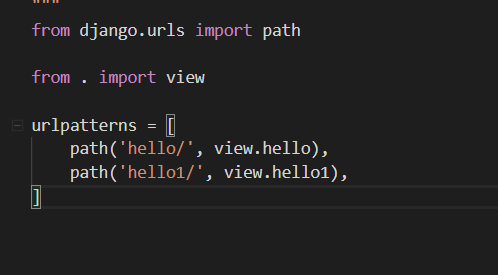
**（3）。将数据库内的内容输出并且展示在django框架中：**



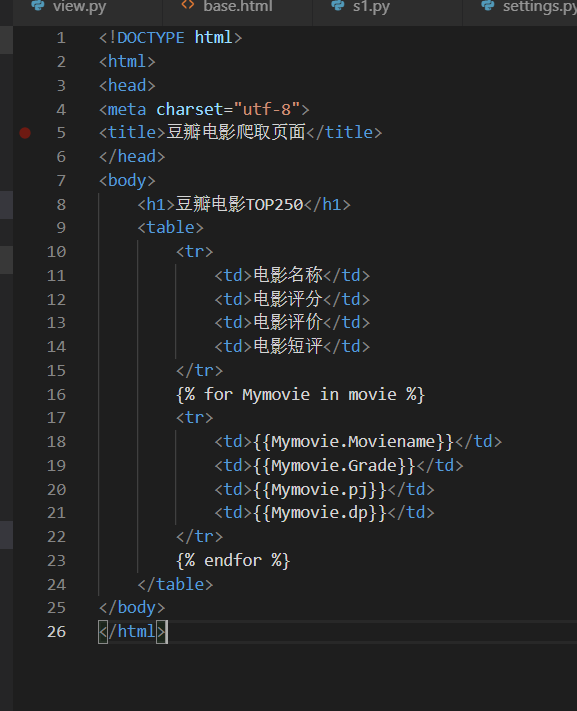
**因为爬取的数据已经输入到数据库，所以输入数据库代码可以省略：**



**url.py文件：**



**Hello.html文件：**

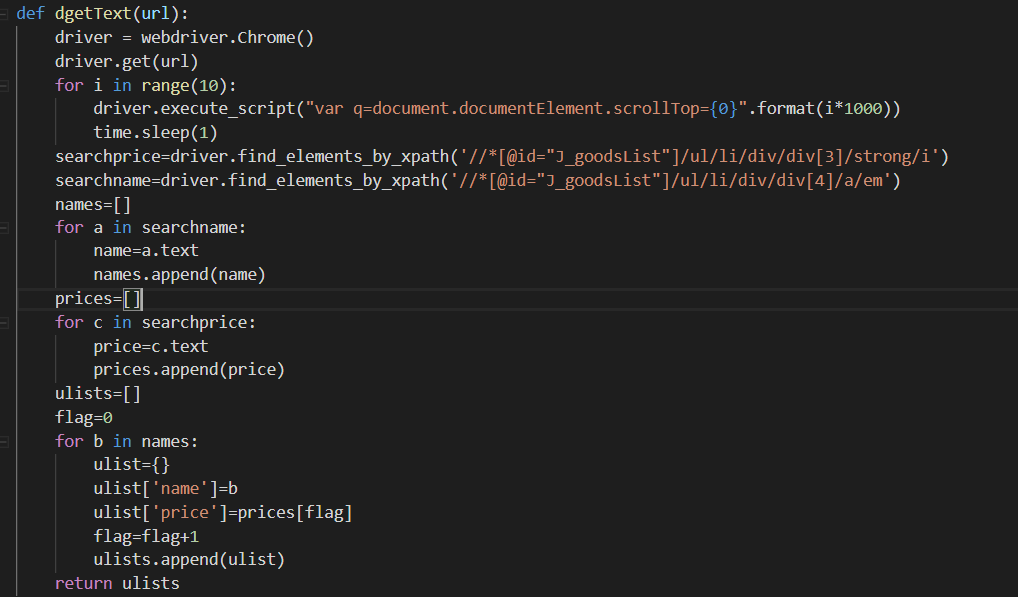


**运行后的效果：**

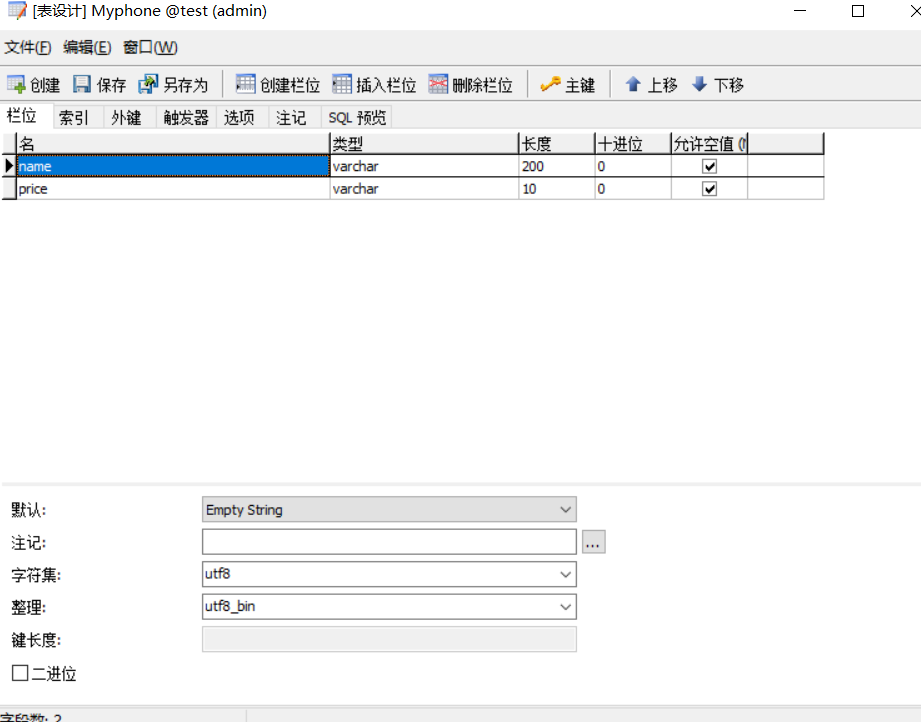


# 动态爬取京东搜索手机信息

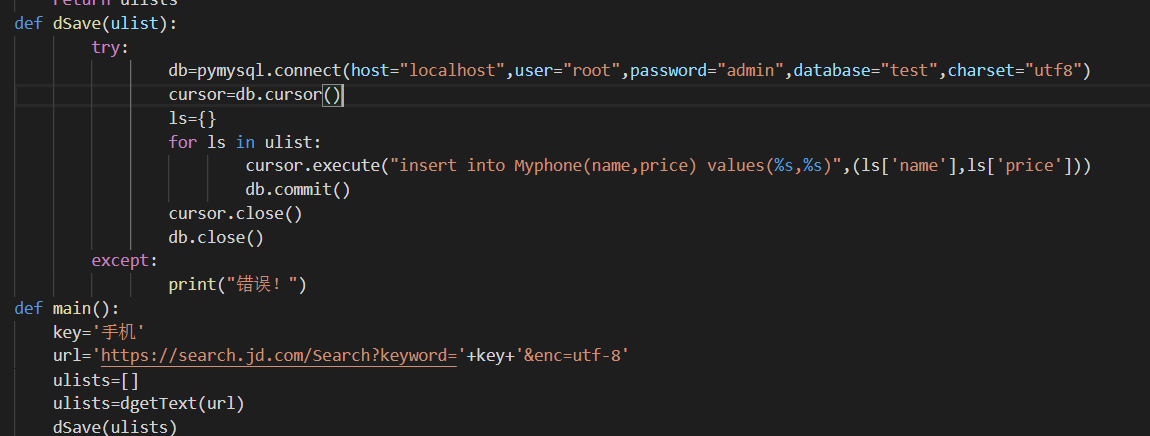
1. 用selenium直接获取所需信息并保存到列表中：

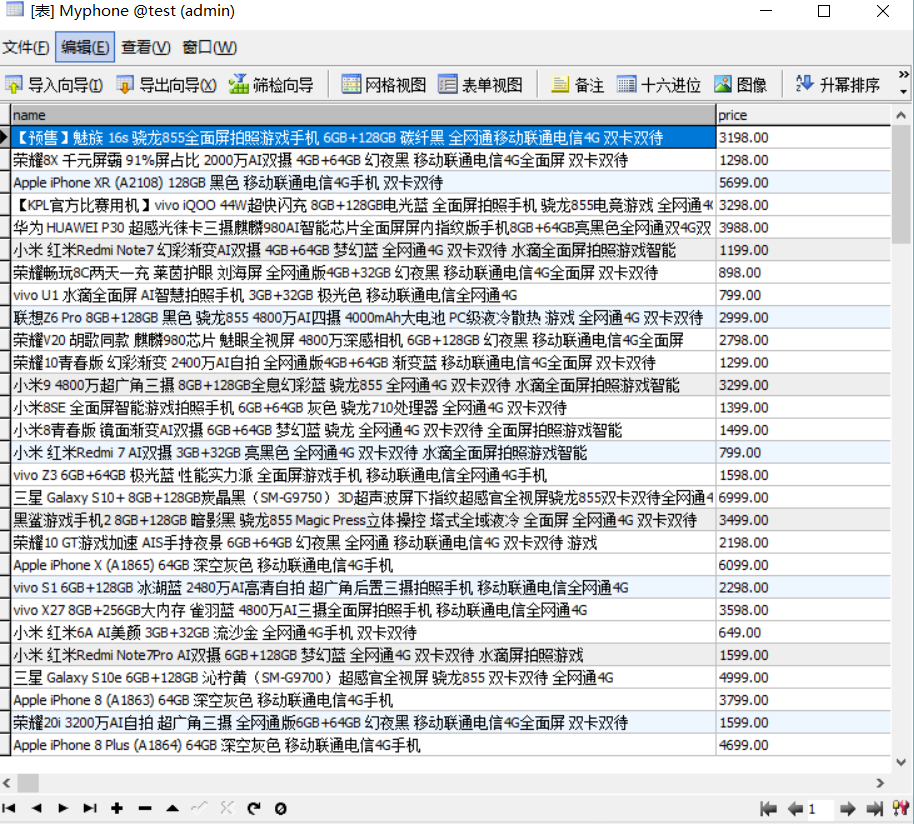


1. 创建名为Myphone的数据表：

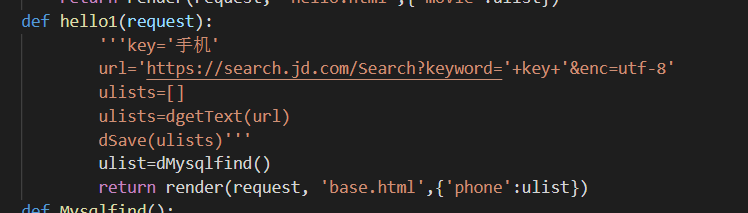


1. 将数据保存到Myphone数据表中：

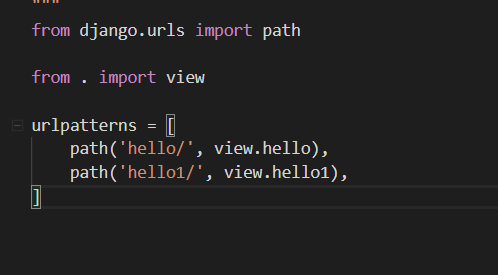




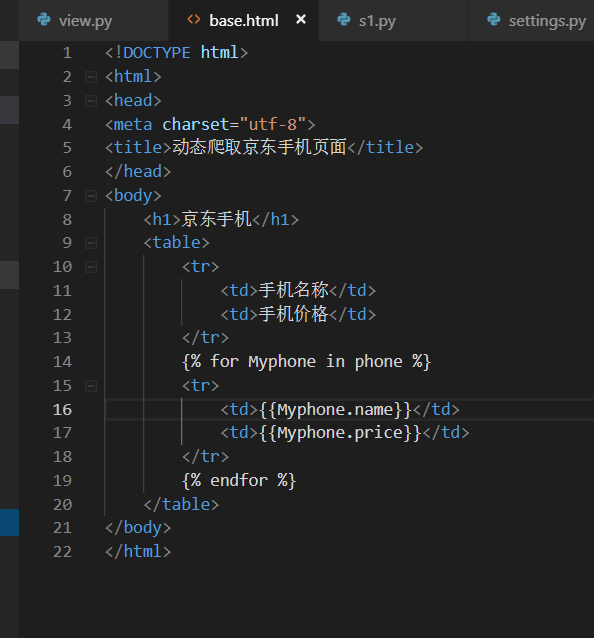
1. 将数据表内的内容输出到django框架：



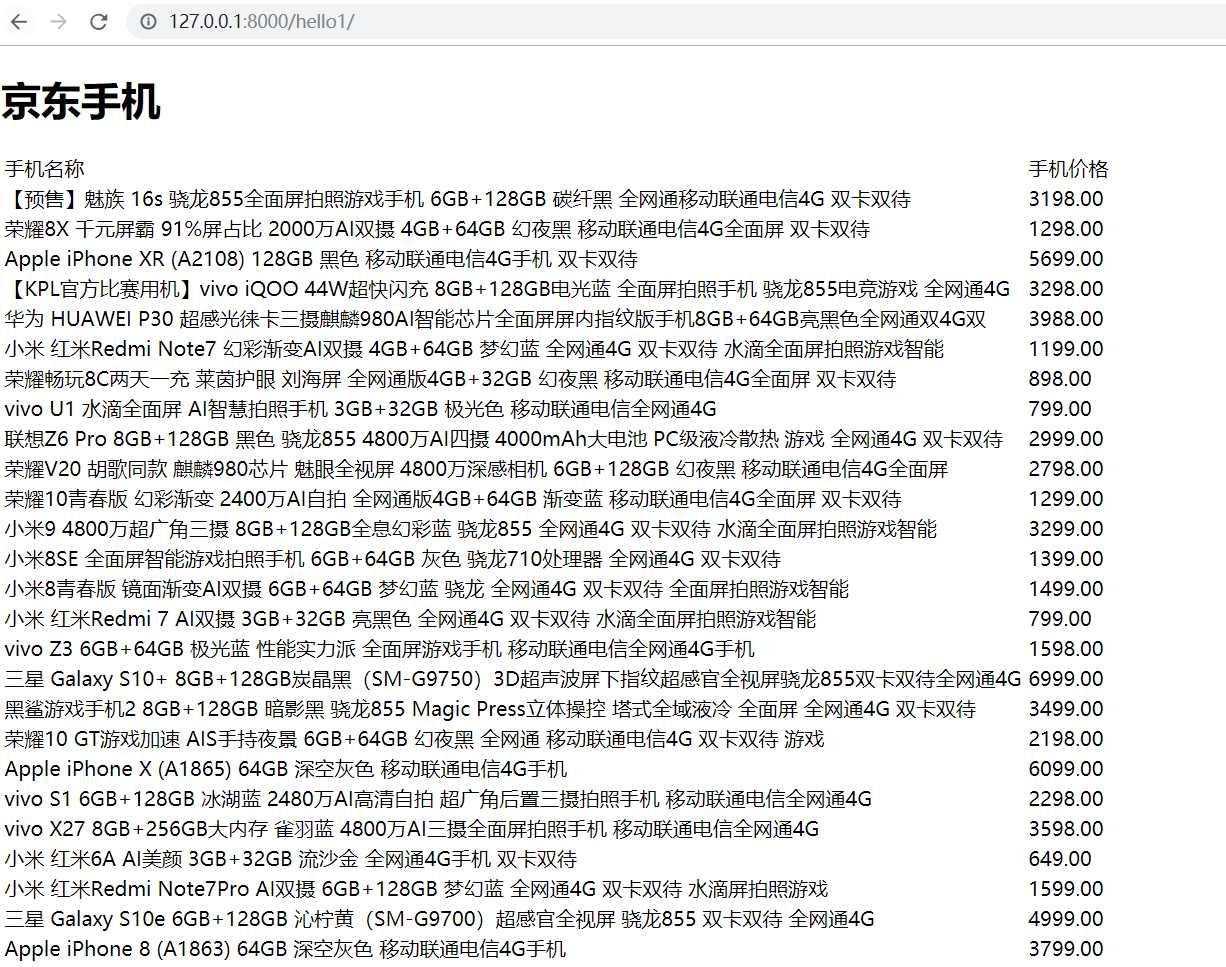
1. urls.py文件：



Base.html文件：

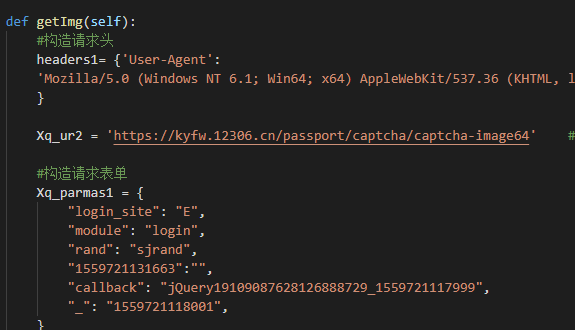


运行后输出的网页：



# 12306登录

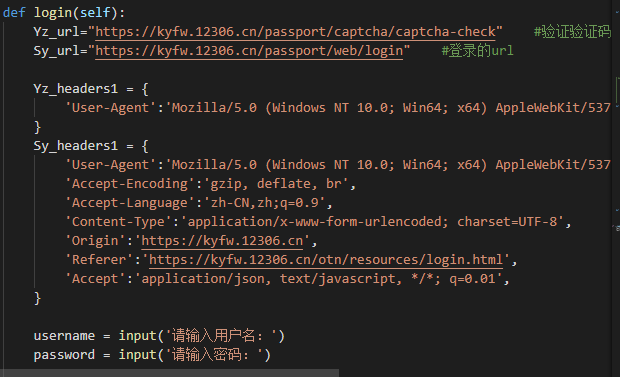
1. 下载验证码图片：



1. 用http://littlebigluo.qicp.net:47720判断图片：



三：开始进行登录：



四：登录运行结果：

