

温州大学瓯江学院

WENZHOU UNIVERSITY OUJIANG
COLLEGE



综合实验报告

学 期 大三第二学期

课 程 python 爬虫

专 业 计算机科学与技术 班 级 16 计算机四班

学生姓名 方正波 学 号 16219111433

爬虫项目说明

项目名称:HelloWord 运行项目

py manage.py runserver

完成功能

1. 豆瓣 top250 电影爬取
2. 七日天气爬取
3. 京东手机信息爬取
4. 使用 scrapy 爬取美剧信息
5. 12306 验证码与登陆处理



界面展示:



爬虫数据显示:

天气爬取

七日天气预报

- 北京 3日（今天）晴 9°C
- 北京 4日（明天）多云 28°C/9°C
- 北京 5日（后天）晴转多云 20°C/8°C
- 北京 6日（周六）多云 21°C/9°C
- 北京 7日（周日）多云转晴 18°C/7°C
- 北京 8日（周一）多云 17°C/8°C
- 北京 9日（周二）阴转多云 15°C/6°C
- 南宁 3日（今天）小雨 17°C
- 南宁 4日（明天）小雨转阴 21°C/17°C
- 南宁 5日（后天）多云转阴 25°C/17°C
- 南宁 6日（周六）阴 27°C/20°C

美剧(scrapy爬虫)

- 杰克绅士第一季
- 三人行第四季
- 切尔诺贝利第一季

京东手机



品牌： 小辣椒 价格： 689.00元

简介： 小辣椒 红辣椒7X 4+64GB 学生智能手机 美颜双摄 微Q多开 人脸识别 移动联

京东手机

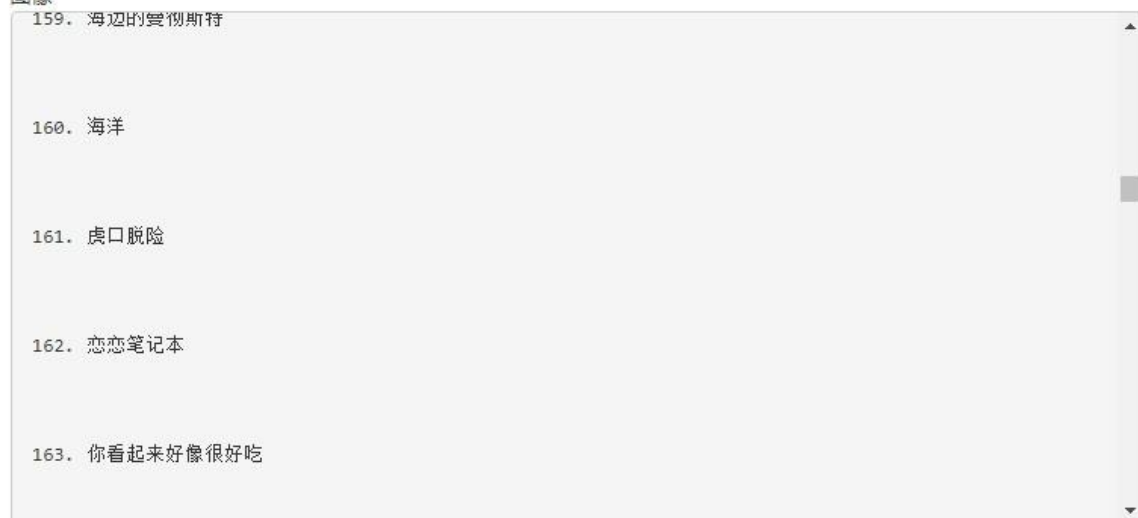


品牌： 华为 价格： 1499.00元

简介： 华为 HUAWEI 畅享 9S 4GB+64GB 极光蓝 全网通 2400万超广角三摄珍珠屏大

豆瓣Top250电影

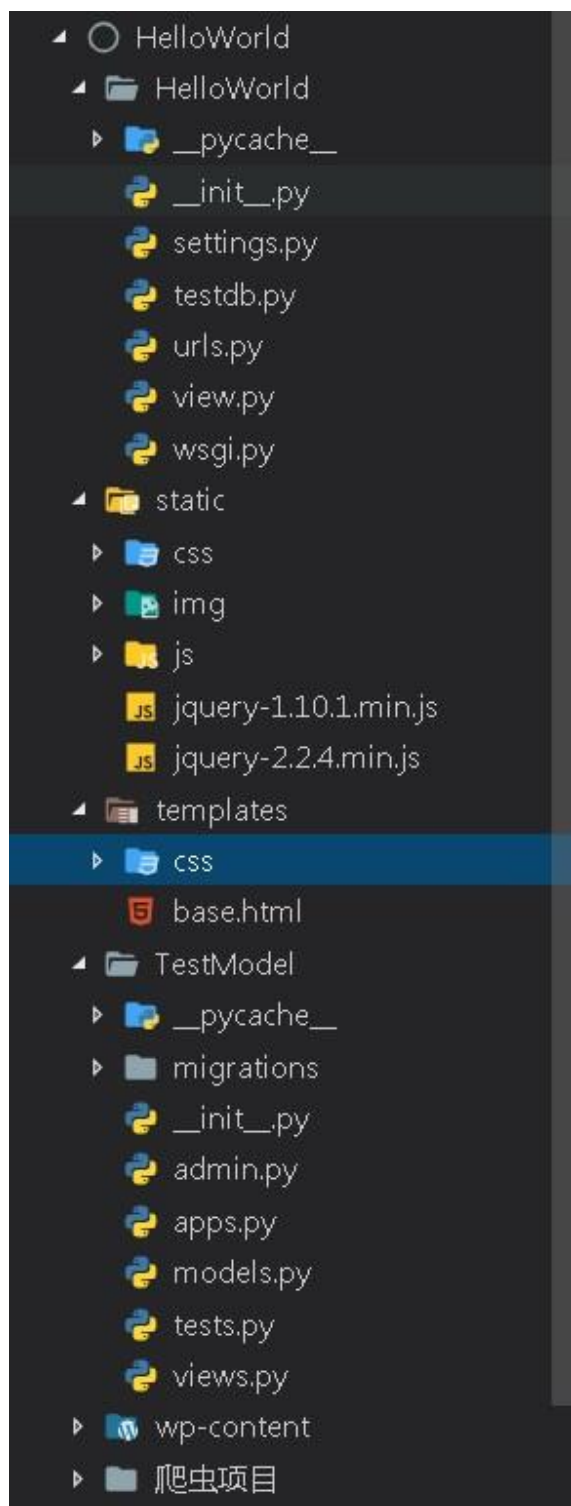
图像



6. 12306 验证码与登陆处理

```
C:\Users\Administrator\AppData\Local\Programs\Python
项目\验证码.py"
位置: 6
answer: 120,100
['验证码校验成功']
['登录成功']
```

项目组成:



TestModel 创建的 app

Models: 创建的数据表

Templates: 显示页面

Testdb.py:接收数据库内的数据

```
#数据库操作
def testdb(request):
    '''response = ""
    response1 = ""
    ...

    allList = Test.objects.all()#获取top250电影
    weather = weathers.objects.all()#获取天气
    phones = phone.objects.all()#获取手机信息

    '''response3 = Test.objects.get(id=1)'''

    '''Test.objects.order_by('name')[0:2]'''

    Test.objects.order_by("id")

    '''Test.objects.filter(name="runoob").order_by("id")'''

    '''for var in list:
    |     response1 += var.content + " "
    response = response1
    return HttpResponse("<p>" + response + "</p>")
    context
    = {}
    context['top250'] = response'''
    return render(request,'base.html',{'allList':allList,'weather':weather,'phones':phones})
```

url.py:向视图传递参数

```
from django.urls import path,include
from . import view,testdb

urlpatterns = [
    |     path('testdb/',testdb.testdb),
    ]
```

Static: 引入本地 bootstrap 文件, 本地图片

爬虫项目	2
top250电影爬取.py	2
京东手机爬取.py	2
处理12306验证码及登入.py	2
天气爬取.py	2

movie	
movie	
__pycache__	
spiders	
__init__.py	
items.py	
meiju.csv	
middlewares.py	
pipelines.py	
settings.py	