

温州大学瓯江学院

WENZHOU UNIVERSITY OUJIANG COLLEGE



综合实验报告

学 期 2018-2019-2

课 程 Python 网络爬虫期中实验

专 业 计算机科学与技术 班 级 16 计算机三

学生姓名 应苗苗 学 号 16219111302

目 录

1 项目需求.....	- 1 -
2 相关技术简介.....	2
2.1Python.....	2
2.2Django.....	2
2.3Bootstrap.....	2
2.4Selenium.....	2
3 软件设计.....	3
3.1 设计方案概述.....	3
3.2 界面设计.....	3
3.3 数据库设计.....	8
4 编码实现.....	13
4.1 用例的实现思路.....	13
4.2 关键代码说明.....	13
5 结论.....	15
5.1 优缺点.....	15
5.2 个人感想.....	15

1 项目需求

实验要求:

- 1、基本 可运行的 Django 项目
- 2、进阶 展示爬虫的内容的 Django 项目
- 3、高级 基于 Bootstrap 或其他 Django 爬虫项目
- 4、代码上传到 Github

Github 账号:mlgmya

项目地址: <https://github.com/mlgmya/Python.git>

2 相关技术简介

2.1 Python

Python 是一种计算机程序设计语言。是一种面向对象的动态类型语言，最初被设计用于编写自动化脚本(shell)，随着版本的不断更新和语言新功能的添加，越来越多被用于独立的、大型项目的开发。

2.2 Django

Django 是一个开放源代码的 Web 应用框架，由 Python 写成。采用了 MVC 的框架模式，即模型 M，视图 V 和控制器 C。它最初是被开发来用于管理劳伦斯出版集团旗下的一些以新闻内容为主的网站的，即是 CMS（内容管理系统）软件。并于 2005 年 7 月在 BSD 许可证下发布。这套框架是以比利时的吉普赛爵士吉他手 Django Reinhardt 来命名的。

2.3 Bootstrap

Bootstrap 是美国 Twitter 公司的设计师 Mark Otto 和 Jacob Thornton 合作基于 HTML、CSS、JavaScript 开发的简洁、直观、强悍的前端开发框架，使得 Web 开发更加快捷。Bootstrap 提供了优雅的 HTML 和 CSS 规范，它即是由动态 CSS 语言 Less 写成。Bootstrap 一经推出后颇受欢迎，一直是 GitHub 上的热门开源项目，包括 NASA 的 MSNBC（微软全国广播公司）的 Breaking News 都使用了该项目。国内一些移动开发者较为熟悉的框架，如 WeX5 前端开源框架等，也是基于 Bootstrap 源码进行性能优化而来

2.4 Selenium

Selenium 是一个用于测试网站的自动化测试工具，支持各种浏览器包括 Chrome、Firefox、Safari 等主流界面浏览器，同时也支持 phantomJS 无界面浏览器。

3 软件设计

3.1 设计方案概述

导航栏：采用 bootstrap 框架响应式的导航栏，当页面缩小到一定程度时将导航栏折叠

首页：采用 bootstrap 框架的图片轮播，点击所有箭头可切换图片

豆瓣电影 TOP250：爬取豆瓣排行榜 250 的电影，并按每页 25 分页展示。其中，分页为 bootstrap 框架

天气预报：爬取中国天气网北京、上海、广州、温州的七天天气，并使用 bootstrap 折叠框架展示

手机：采用 selenium 爬取京东手机搜索结果，并按照仿淘宝页面分页展示手机信息

书包：爬取淘宝搜索书包页面前两页结果并展示

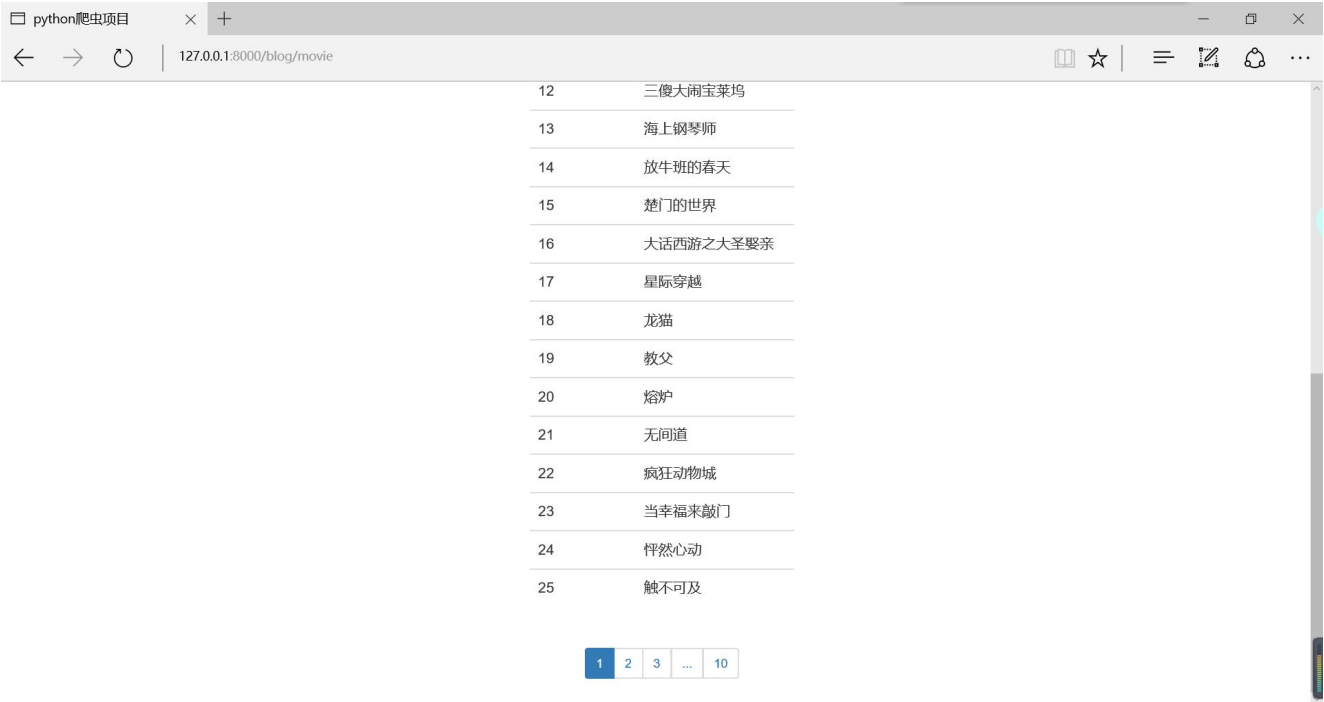
关于：个人信息介绍

3.2 界面设计





排名	电影
1	肖申克的救赎
2	霸王别姬
3	这个杀手不太冷
4	阿甘正传
5	美丽人生
6	泰坦尼克号
7	千与千寻
8	辛德勒的名单
9	盗梦空间
10	忠犬八公的故事
11	机器人总动员
12	三傻大闹宝莱坞
13	海上钢琴师



python爬虫项目

127.0.0.1:8000/blog/weather

爬虫项目 首页 豆瓣电影TOP250 天气预报 手机 书包 关于

北京

上海

广州


温州

城市	日期	天气	温度
温州	24日(今天)	小雨	30/21°C
温州	25日(明天)	小到中雨转小雨	26/20°C
温州	26日(后天)	小雨转阴	22/16°C
温州	27日(周六)	多云	21/14°C
温州	28日(周日)	小雨转多云	23/18°C
温州	29日(周一)	小雨转阴	30/20°C
温州	30日(周二)	小雨转大雨	25/19°C

python爬虫项目


127.0.0.1:8000/blog/phones

爬虫项目 首页 豆瓣电影TOP250 天气预报 手机 书包 关于




¥ 1448.00

三星 Galaxy A6s 6GB+64GB 花木蓝(SM-G6200) 全面屏 渐变色 全网通 4G 双卡双待




¥ 5899.00

Apple iPhone XR (A2108) 128GB 黑色 移动联通电信4G手机 双卡双待




¥ 3298.00


【KPL官方比赛用机】vivo iQOO 44W超快闪充 8GB+128GB电光蓝 全面屏拍照手机 骁龙855电竞游戏 全网通4G





¥ 1299.00

荣耀8X 千元屏霸 91%屏占比 2000万AI双摄 4GB+64GB 幻夜黑 移动联通电信4G全面屏 双卡双待









python爬虫项目

127.0.0.1:8000/blog/phones


☆

≡

...

¥ 2599.00

荣耀Note10 全网通6G+64G 幻夜黑 移动联通电信4G全面屏手机 双卡双待 游戏手机




¥ 3598.00

vivo 【新品上市】 X27 4800万广角夜景二摄 零界全面屏拍照手机 移动联通电信全网通4G 雀羽蓝 8GB+256GB

¥ 4588.00

华为 HUAWEI P20 Pro 全面屏徕卡三摄游戏手机 6GB+128GB 亮黑色 全网通移动联通电信4G 双卡双待




¥ 949.00

魅族 Note8 全面屏手机 4GB+64GB 曜黑 全网通移动联通电信4G手机 双卡双待

¥ 4499.00

华为 HUAWEI Mate20X 麒麟980 芯片全面屏超微距影像超大广角徕卡三摄6GB+128GB宝石蓝全网通版双4G游戏




¥ 3099.00

Apple iPhone 6s Plus (A1699) 128G 玫瑰金色 移动联通电信4G手机

¥ 2999.00

OPPO R17 2500万美颜拍照 6.4英寸水滴屏 光感屏指纹 8G+128G 流光蓝 全网通 移动联通电信4G 双卡双待



¥ 599.00

酷派 (Coolpad) 酷玩8 Lite 6"高清全面屏 1300万双摄 私密双系统 梦幻紫 3GB+32GB 双卡双待全网通

1

2

3

...

100

python爬虫项目

127.0.0.1:8000/blog/bags

☆

≡

...

爬虫项目 [首页](#) [豆瓣电影TOP250](#) [天气预报](#) [手机](#) [书包](#) [关于](#)

序号	价格	商品名称
1	45.80	小学生书包男生1-3-4-6年级6-12周岁儿童
2	69.90	迪卡侬双肩包运动背包男女健身包书包学生户外旅行足球篮球KIPSTA
3	88.00	迪士尼书包小学生女童1-3-6三年级公主女孩轻便儿童减负双肩背包5
4	499.00	Fjallraven/北极狐双肩包kanken classic 书包女户外旅行背包23510
5	129.00	小米双肩包简约休闲多功能书包男女笔记本电脑包时尚潮流旅行背包
6	258.00	电视剧款JanSport旗舰店官网杰斯伯双肩包时尚女书包背包男大容量
7	348.00	爆款anello官方旗舰店日本ins潮风双肩女背包男离家出走书包包
8	199.00	小米 米兔儿童书包 6-12岁男女小学生潮双肩背包幼儿园大容量背包
9	79.00	双肩包男士背包大容量旅行包电脑休闲包



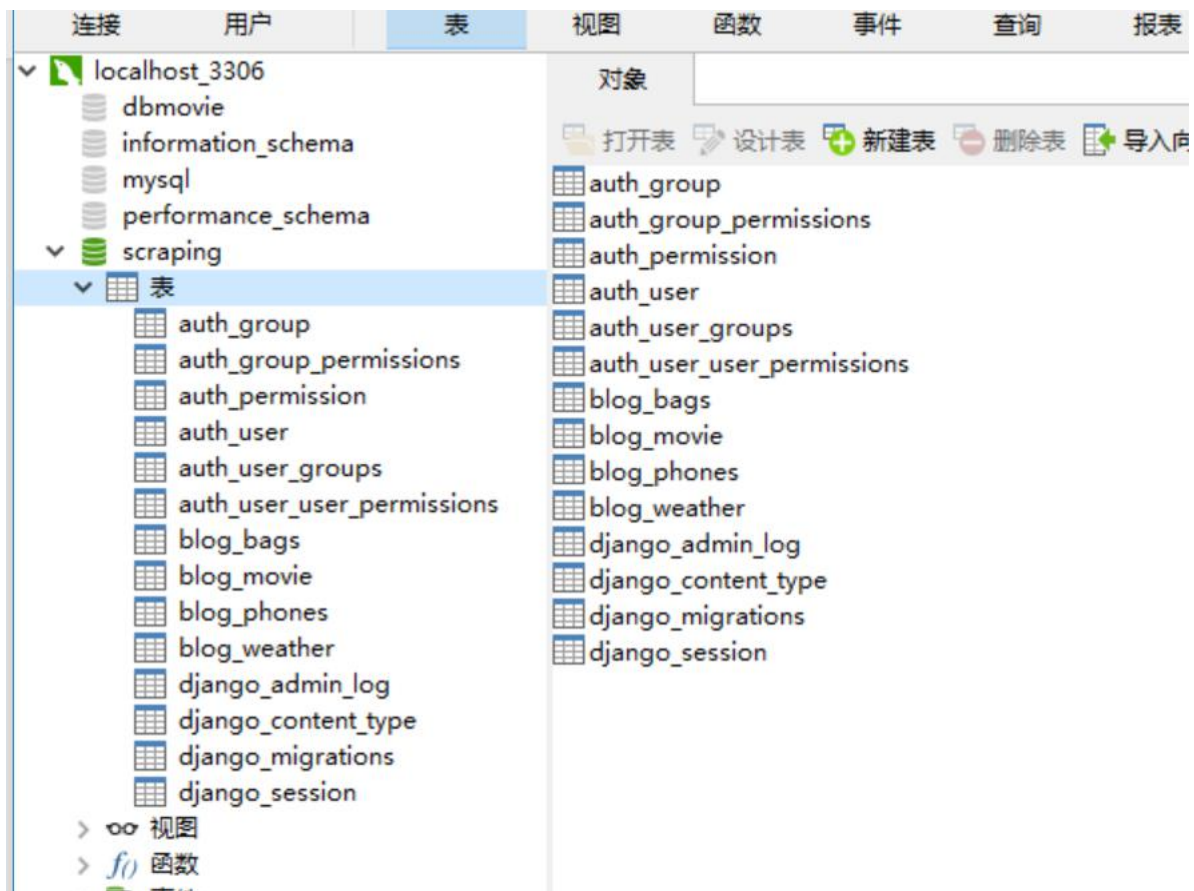
个人信息

学号 16219111302

班级 16计算机科学与技术三班

姓名 应苗苗

3.3 数据库设计



对象

blog_movie @scraping (loc...

开始事务

备注

筛选

排序

导入

导出

id	name
1	肖申克的救
2	霸王别姬
3	这个杀手不
4	阿甘正传
5	美丽人生
6	泰坦尼克号
7	千与千寻
8	辛德勒的名
9	盗梦空间
10	忠犬八公的
11	机器人总动
12	三傻大闹宝
13	海上钢琴师
14	放牛班的春
15	楚门的世界
16	大话西游之
17	星际穿越
18	龙猫
19	教父

对象

blog_weather @scraping (lo...

开始事务

备注

筛选

排序

导入

导出

id	wCity	wDate	wWeather	wTemp
29	北京	24日 (今天)	小雨	19/9℃
30	北京	25日 (明天)	多云	19/8℃
31	北京	26日 (后天)	晴转多云	21/12℃
32	北京	27日 (周六)	阴转小雨	18/8℃
33	北京	28日 (周日)	多云	22/10℃
34	北京	29日 (周一)	多云转小雨	25/13℃
35	北京	30日 (周二)	多云	26/14℃
36	上海	24日 (今天)	小雨转阴	25/18℃
37	上海	25日 (明天)	多云	25/16℃
38	上海	26日 (后天)	多云	20/13℃
39	上海	27日 (周六)	晴转多云	19/16℃
40	上海	28日 (周日)	多云	24/18℃
41	上海	29日 (周一)	中雨转大雨	26/18℃
42	上海	30日 (周二)	小雨转中雨	22/18℃
43	广州	24日 (今天)	多云	32/24℃
44	广州	25日 (明天)	雷阵雨	29/22℃
45	广州	26日 (后天)	大雨转大到暴雨	26/23℃
46	广州	27日 (周六)	大雨转雷阵雨	27/24℃
47	广州	28日 (周日)	雷阵雨	29/24℃

对象 blog_phones @scraping (lo...					
开始事务 备注 筛选 排序 导入 导出					
id	mNo	mMark	mPrice	mNote	mFile
1	000001	三星	1448.00	三星 Galaxy	000001.jpg
2	000002	Apple	5899.00	Apple iPhor	000002.jpg
3	000003	【KPL官方比	3298.00	【KPL官方比	000003.jpg
4	000004	荣耀8X	1299.00	荣耀8X 千元	000004.jpg
5	000005	荣耀10青春版	1299.00	荣耀10青春版	000005.jpg
6	000006	vivo	799.00	vivo U1 水滴	000006.jpg
7	000007	荣耀V20	2799.00	荣耀V20 胡	000007.jpg
8	000008	vivo	3598.00	vivo X27 8G	000008.jpg
9	000009	OPPO	2999.00	OPPO Reno	000009.jpg
10	000010	小米	1199.00	小米 红米Re	000010.jpg
11	000011	荣耀畅玩8C	899.00	荣耀畅玩8C	000011.jpg
12	000012	小米	799.00	小米 红米6 4	000012.jpg
13	000013	黑鲨游戏手机	3499.00	黑鲨游戏手机	000013.jpg
14	000014	Apple	6199.00	Apple iPhor	000014.jpg
15	000015	小米	799.00	小米 红米Re	000015.jpg
16	000016	小米8SE	1599.00	小米8SE 全	000016.jpg
17	000017	OPPO	2999.00	OPPO Reno	000017.jpg
18	000018	Apple	9699.00	Apple iPhor	000018.jpg
19	000019	Apple	3799.00	Apple iPhor	000019.jpg

对象 blog_bags @scraping (local...		
<div> <div>开始事务</div> <div>备注</div> <div>筛选</div> <div>排序</div> <div>导入</div> <div>导出</div> </div>		
id	price	name
1	45.80	小学生书包男生1-3-4-6年级6-12周岁儿童
2	69.90	迪卡侬双肩包运动背包男女健身包包学生户外旅行足球篮球KIPSTA
3	88.00	迪士尼书包小学生女童1-3-6三年级公主女孩轻便儿童减负双肩背包5
4	499.00	Fjallraven/北极狐双肩包kanken classic书包女户外旅行背包23510
5	129.00	小米双肩包简约休闲多功能书包男女笔记本电脑包时尚潮流旅行背包
6	258.00	电视剧款JanSport旗舰店官网杰斯伯双肩包时尚女书包背包男大容量
7	348.00	爆款anello官方旗舰店日本ins潮风双肩女背包男离家出走包包
8	199.00	小米 米兔儿童书包 6-12岁男女小学生潮双肩背包幼儿园大容量背包
9	79.00	双肩包男士背包大容量旅行包电脑休闲女时尚潮流高中初中学生书包
10	109.00	七匹狼商务双肩包男书包中学生女电脑包旅行包休闲男士背包大容量
11	148.00	佑一良品男士背包双肩包男韩版大学生书包男时尚潮流大容量旅行包
12	69.00	巴布豆旗舰店书包1-3年级护脊减负儿童书包男4-6小学生书包轻便
13	299.00	BOPAI博牌电脑背包男户外旅行休闲双肩包商务书包出差多功能男包
14	49.00	小学生书包6-12周岁 女儿童双肩包 3-5年级女童背包 1-3年级女孩
15	45.80	儿童书包小学生男童1-3年级6-12周岁4-6年级男孩双肩背包轻便减负
16	59.80	商务背包男士双肩包韩版潮流旅行包休闲女学生书包简约时尚电脑包
17	168.00	双肩包男书包男士时尚潮流青年休闲简约潮牌旅行背包大学生电脑包
18	69.00	迪士尼书包小学生男女1-3-4-6年级米奇减负背包儿童书包8-10-12岁
19	119.00	巴朗商务双肩包休闲时尚潮流大学生书包15.6寸电脑包男士背包男潮

4 编码实现

4.1 用例的实现思路

1. 创建实验项目 myblog
2. 创建 APP blog
3. 使用 models 创建模板数据库 movie, weather, phones, bags
4. 爬取电影, 天气, 手机及书包并存入数据库
5. 创建 html 页面展示数据库数据
6. 通过 urls 设置各个页面地址
7. 采用 bootstrap 框架设计各个页面

4.2 关键代码说明

models.py//数据表设计

```
from django.db import models
```

```
class Movie(models.Model):
```

```
name=models.CharField(max_length=4000)
```

```
objects=models.Manager()
```

```
class Weather(models.Model):
```

```
wCity=models.CharField(max_length=20)
```

```
wDate=models.CharField(max_length=20)
```

```
wWeather=models.CharField(max_length=20)
```

```
wTemp=models.CharField(max_length=20)
```

```
objects=models.Manager()
```

```
class Phones(models.Model):
```

```
mNo=models.CharField(max_length=32)
```

```
mMark=models.CharField(max_length=256)
```

```
mPrice=models.CharField(max_length=32)
```

```
mNote=models.CharField(max_length=1024)
```

```
mFile=models.CharField(max_length=256)
```

```
objects=models.Manager()
```

```
class Bags(models.Model):
```

```
price=models.CharField(max_length=50)
```

```
name=models.CharField(max_length=50)
```

```
objects=models.Manager()
```

Myblog.urls.py

```
from django.contrib import admin
```

```
from django.conf.urls import url,include
```

```
from . import view

urlpatterns = [
    url(r'^admin/', admin.site.urls),
    url(r'^blog/', include('blog.urls', namespace='blog')),
]
```

Blog.urls.py

```
from django.contrib import admin
from django.conf.urls import url
from . import views

app_name='blog'

urlpatterns = [
    url(r'^$', views.index, name='index'),
    url(r'^movie/$', views.movie, name='movie'),
    url(r'^weather/$', views.weather, name='weather'),
    url(r'^phones/$', views.phones, name='phones'),
    url(r'^bags/$', views.bags, name='bags'),
    url(r'^info/$', views.info, name='info'),
]
```


5 结论

5.1 优缺点

5.1.1 优点

完成老师布置的任务

页面展示比较完整

5.1.2 缺点

需要使用线程优化爬取速度

页面美化不够

5.2 个人感想

Python 世界博大精深，通过这次的实验学习了解了 `django`, `bootstrap`, `selenium` 等，学习网上的课程初步完成了一个网站，还有更多的框架或库需要我们学习。