

Setting.py 主要配置

```
INSTALLED_APPS = [  
    'django.contrib.admin',  
    'django.contrib.auth',  
    'django.contrib.contenttypes',  
    'django.contrib.sessions',  
    'django.contrib.messages',  
    'django.contrib.staticfiles',  
    'jd',  
]
```

```
DATABASES = {  
    'default': {  
        'ENGINE': 'django.db.backends.mysql',  
        'NAME': 'python_wzl',  
        'USER': 'root',  
        'PASSWORD': '123456',  
        'HOST': 'localhost',  
        'PORT': '3306',  
    }  
}
```

Urls.py 连接 jd.urls

```
1  from django.contrib import admin  
2  from django.urls import path, include  
3  
4  
5  
6  urlpatterns = [  
7      path('', include('jd.urls')),  
8      path('admin/', admin.site.urls),  
9  
10 ]
```

Jd.urls

```
from django.urls import path  
from . import views  
  
urlpatterns = [  
    path('', views.order, name='order'),  
]
```

Models.py 重要映射

```
# -*- coding: utf-8 -*-
from django.db import models

class jingdong(models.Model):
    name = models.CharField( max_length=255, blank=True, null=True)
    phoneno = models.CharField( max_length=255, blank=True, null=True)
    address = models.CharField( max_length=255, blank=True, null=True)
    price = models.CharField( max_length=255, blank=True, null=True)

    class Meta:
        managed = True
        db_table = 'jingdong'
```

Views.py 获取数据库信息并输出

```
# -*- coding: utf-8 -*-
from django.shortcuts import render
from jd import models
import MySQLdb

def get_data(sql):#获取数据库的数据
    conn = MySQLdb.connect('127.0.0.1','root','123456','python_wz1',port=3306)
    cur = conn.cursor()
    cur.execute(sql)
    results = cur.fetchall()
    cur.close()
    conn.close()
    return results

def order(request):# 向页面输出订单
    sql = "select * from jingdong"
    m_data = get_data(sql)
    return render(request, 'index.html', {'order':m_data})
```

京东信息爬虫代码:

```
# coding:utf-8
from lxml import etree
import sys
import time
from multiprocessing.dummy import Pool as Threadpool
import requests
import json
from pymongo import MongoClient
```

```

import datetime
import pymysql

import importlib
import sys
importlib.reload(sys)

def get_sku_id(url):
    html= requests.get(url,headers = headers)
    selector = etree.HTML(html.text)
    product_list = selector.xpath('//*[@id="J_goodsList"]/ul/li')
    for product in product_list:
        product_sku = product.xpath("@data-sku")[0]
        get_phont_content(product_sku)
        get_phone_price(product_sku)

def get_phont_content(sku):
    content_url = 'https://item.jd.com/{}.html'.format(sku)
    html = requests.get(content_url)
    selector = etree.HTML(html.text)
    product_detail = selector.xpath('//ul[@class="parameter2 p-parameter-list"]')
    for product_li in product_detail:
        product_dict["手机名称"] = product_li.xpath("li[1]/@title")[0]
        product_dict["手机编号"] = product_li.xpath("li[2]/@title")[0]
        product_dict["手机产地"] = product_li.xpath("li[4]/@title")[0]

def get_phone_price(sku):
    phone_url = 'https://p.3.cn/prices/mgets?callback=jQuery1346298&type=1&area=1_72_2799_0&pdtk=&pdui=1713961549&pdpin=&pin=null&pdbp=0&skulds=J_{}'.format(str(sku))
    phone_detail = requests.get(phone_url,headers = headers).content
    ss = phone_detail.decode('utf-8')
    ss = ss.split("(")[1]
    ss = ss.split(")")[0]
    phone_json = json.loads(ss)
    for info in phone_json:
        product_dict["手机价格"] = info.get("p")
    save_mysql(product_dict)

def save_mysql(product_dict):

```

```

db
pymysql.connect(host="localhost",user="root",passwd="123456",db="python_wzl",charset="utf
8")

cursor = db.cursor()

sql = "insert into jingdong(name,phoneno,address,price)
values(%s,%s,%s,%f)"%( ""+product_dict["手机名称"]+"",""+product_dict["手机编号
"]+"",""+product_dict["手机产地"]+"",float(product_dict["手机价格"]))

print(sql)
cursor.execute(sql)
db.commit()

product_dict = {}
i = 1111

if __name__ == '__main__':
    #save_mysql()
    headers = {'User-Agent':'Mozilla/5.0 (Macintosh; Intel Mac OS X 10_13_3)
AppleWebKit/537.36 (KHTML, like Gecko) Chrome/65.0.3325.181 Safari/537.36'}

    html

    ['https://search.jd.com/Search?keyword=%E6%89%8B%E6%9C%BA&enc=utf-8&page={}'.format(
page) for page in range(1,10)]

    #html

    get_sku_id('https://search.jd.com/Search?keyword=%E6%89%8B%E6%9C%BA&enc=utf-8' )
    pool = ThreadPool(2)
    pool.map(get_sku_id,html)
    pool.close()
    pool.join()

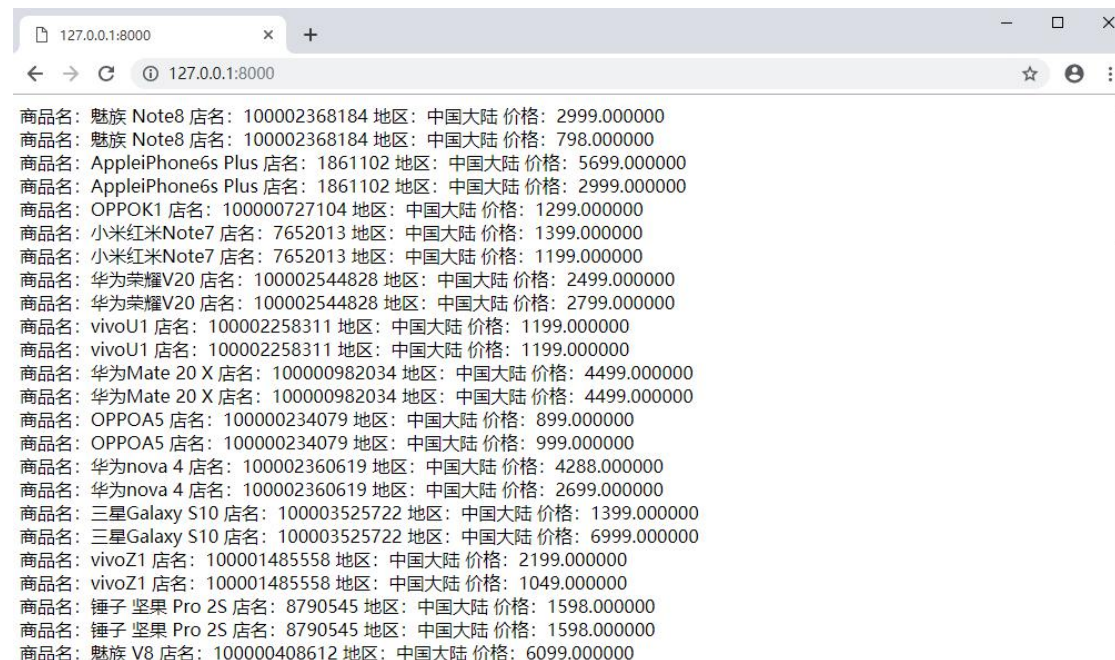
```

运行结果

jingdong @python_wzl (localhost) - 表 - Navicat for MySQL

id	name	phoneno	address	price
1	魅族 Note8	100002368184	中国大陆	2999.000000
2	魅族 Note8	100002368184	中国大陆	798.000000
3	AppleiPhone6s Plus	1861102	中国大陆	5699.000000
4	AppleiPhone6s Plus	1861102	中国大陆	2999.000000
5	OPPOK1	100000727104	中国大陆	1299.000000
6	小米红米Note7	7652013	中国大陆	1399.000000
7	小米红米Note7	7652013	中国大陆	1199.000000
8	华为荣耀V20	100002544828	中国大陆	2499.000000
9	华为荣耀V20	100002544828	中国大陆	2799.000000
10	vivoU1	100002258311	中国大陆	1199.000000
11	vivoU1	100002258311	中国大陆	1199.000000
12	华为Mate 20 X	100000982034	中国大陆	4499.000000
13	华为Mate 20 X	100000982034	中国大陆	4499.000000
14	OPPOA5	100000234079	中国大陆	899.000000
15	OPPOA5	100000234079	中国大陆	999.000000
16	华为nova 4	100002360619	中国大陆	4288.000000
17	华为nova 4	100002360619	中国大陆	2699.000000
18	三星Galaxy S10	100003525722	中国大陆	1399.000000
19	三星Galaxy S10	100003525722	中国大陆	6999.000000
20	vivoZ1	100001485558	中国大陆	2199.000000

页面展示



豆瓣 top250 原理相同

爬虫代码:

```
import requests
```

```
from bs4 import BeautifulSoup
```

```
import pymysql
```

```
def get_pages_link():
```

```
    # 插入到数据库
```

```
    db
```

```
=
```

```
pymysql.connect(host="localhost",user="root",passwd="123456",db="python_wzl",charset="utf8")
```

```
    cursor = db.cursor()
```

```
    header = {
```

```
        'User-Agent': 'Mozilla/5.0 (Windows NT 10.0; WOW64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/69.0.3497.92 Safari/537.36',
```

```
        'Connection': 'keep-alive'
```

```
    }
```

```
    for item in range(0, 250, 25):
```

```
        url = "https://movie.douban.com/top250?start={}".format(item)
```

```

web_data = requests.get(url, headers=header)
soup = BeautifulSoup(web_data.content, 'lxml')
for movie in soup.select('#wrapper li'):

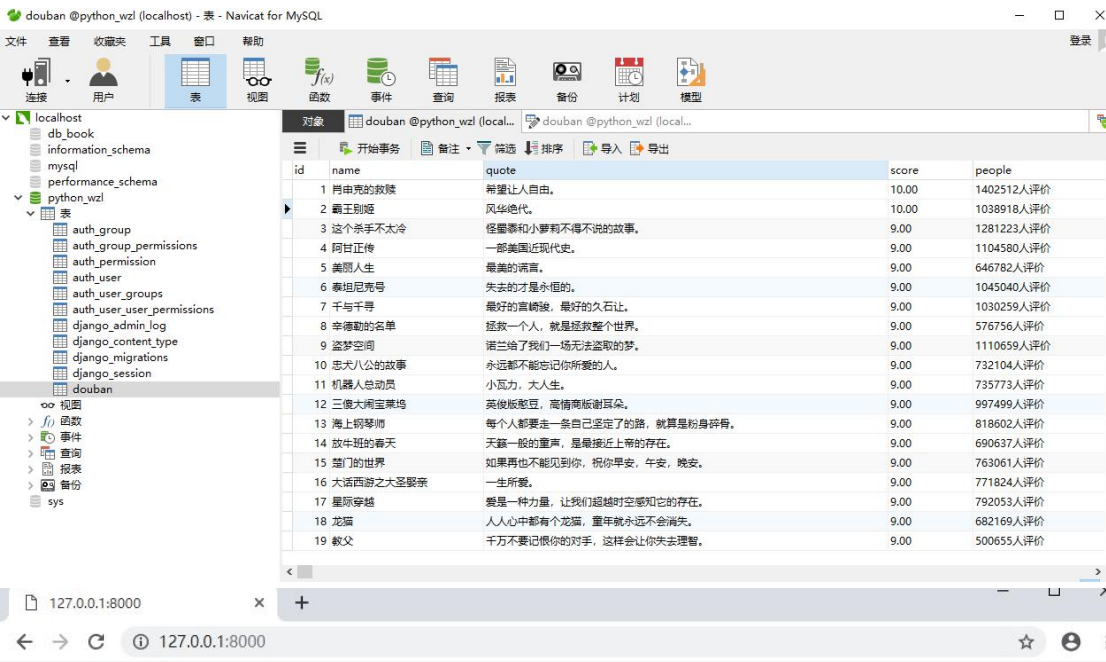
    href = movie.select('.hd > a')[0] # 链接
    href=movie.find('a')['href']
    name = movie.select('.hd > a > span')[0].text # 片名
    star = movie.select('.rating_num')[0].text # 评分
    people = movie.select('.star > span')[3].text # 评价人数
    try:
        quote = movie.select('.inq')[0].text
    except:
        print('没有 quote 哦')
        quote = None
    data = {
        # 'url': href,
        '评分': star,
        '片名': name,
        '名言': quote,
        '评价人数': people
    }
    sql = "insert into douban(score,name,quote,people) values
('%f','%s','%s','%s')"%(float(star), name, quote, people)
    cursor = db.cursor()
    print(sql)
    cursor.execute(sql)
    db.commit()

    print(data)
    # print(movie)
    print('\n' + '-' * 50 + '\n')
# 关闭数据库
cursor.close()
conn.close

if __name__ == '__main__':
    get_pages_link()

```


运行结果：



电影名：肖申克的救赎 影评：希望让人自由。 评价：9.600000 评论人数：1402512人评价
电影名：霸王别姬 影评：风华绝代。 评价：9.600000 评论人数：1038918人评价
电影名：这个杀手不太冷 影评：怪蜀黍和小萝莉不得不说的故事。 评价：9.400000 评论人数：1281223人评价
电影名：阿甘正传 影评：一部美国近现代史。 评价：9.400000 评论人数：1104580人评价
电影名：美丽人生 影评：最美的谎言。 评价：9.500000 评论人数：646782人评价
电影名：泰坦尼克号 影评：失去的才是永恒的。 评价：9.300000 评论人数：1045040人评价
电影名：千与千寻 影评：最好的宫崎骏，最好的久石让。 评价：9.300000 评论人数：1030259人评价
电影名：辛德勒的名单 影评：拯救一个人，就是拯救整个世界。 评价：9.500000 评论人数：576756人评价
电影名：盗梦空间 影评：诺兰给了我们一场无法盗取的梦。 评价：9.300000 评论人数：1110659人评价
电影名：忠犬八公的故事 影评：永远都不能忘记你所爱的人。 评价：9.300000 评论人数：732104人评价
电影名：机器人总动员 影评：小瓦力，大人生。 评价：9.300000 评论人数：735773人评价
电影名：三傻大闹宝莱坞 影评：英俊版憨豆，高情商版谢耳朵。 评价：9.200000 评论人数：997499人评价
电影名：海上钢琴师 影评：每个人都要走一条自己坚定了的路，就算是粉身碎骨。 评价：9.200000 评论人数：818602人评价
电影名：放牛班的春天 影评：天籁一般的童声，是最接近上帝的存在。 评价：9.300000 评论人数：690637人评价
电影名：楚门的世界 影评：如果再也不能见到你，祝你早安，午安，晚安。 评价：9.200000 评论人数：763061人评价
电影名：大话西游之大圣娶亲 影评：一生所爱。 评价：9.200000 评论人数：771824人评价
电影名：星际穿越 影评：爱是一种力量，让我们超越时空感知它的存在。 评价：9.200000 评论人数：792053人评价
电影名：龙猫 影评：人人心中都有个龙猫，童年就永远不会消失。 评价：9.200000 评论人数：682169人评价
电影名：教父 影评：千万不要记恨你的对手，这样会让你失去理智。 评价：9.300000 评论人数：500655人评价
电影名：熔炉 影评：我们一路奋战不是为了改变世界，而是为了不让世界改变我们。 评价：9.300000 评论人数：446500人评价
电影名：无间道 影评：香港电影史上永不过时的杰作。 评价：9.100000 评论人数：633709人评价
电影名：疯狂动物城 影评：迪士尼给我们营造的乌托邦就是这样，永远善良勇敢，永远出乎意料。 评价：9.200000 评论人数：870115人评价
电影名：当幸福来敲门 影评：平民励志片。 评价：9.000000 评论人数：807205人评价
电影名：超越时空的拥抱 影评：真正的幸福来自内心绽放。 评价：9.000000 评论人数：801522人评价