



# 温州大学瓯江学院

WENZHOU UNIVERSITY OUJIANG COLLEGE

## 《爬虫期中作业》

题    目： 网络爬虫的数据爬取

分    院： 理工分院

班    级： 16 计算机科学与技术 2 班

姓    名： 陈祥牛

学    号： 16219111204

完成日期： 2019 年 4 月 26 日

温州大学瓯江学院教务部

二〇一二年十一月制

# 目录

|                              |   |
|------------------------------|---|
| 目录.....                      | 2 |
| 一. 豆瓣 TOP250 电影网页的静态爬虫 ..... | 3 |
| 二. 动态爬取京东搜索手机信息.....         | 8 |

# 一. 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:

```
def main():
    url='https://movie.douban.com/top250'
    flag=0
    while flag<250:
        html=getHTMLtext(url,flag)
        time.sleep(2)
        ulist=[]
        getText(ulist,html)
        Save(ulist)
        flag=flag+25
```

```
def getHTMLtext(url,flag):
    try:
        if(flag==0):
            kk={}
        else:
            kk={'start':flag,'filter':''}
        r=requests.get(url,params=kk)
        r.raise_for_status()
        r.encoding='utf-8'
        return r.text
    except:
        return ""
```

2. 用 BeautifulSoup 库将相应的内容爬去下来并保存到列表里:

```
def main():
    url='https://movie.douban.com/top250'
    flag=0
    while flag<250:
        html=getHTMLtext(url,flag)
        time.sleep(2)
        ulist=[]
        getText(ulist,html)
        Save(ulist)
        flag=flag+25
```

```
def getText(ulist,html):
    soup=BeautifulSoup(html,"html.parser")
    a=soup.find('ol',attrs={'class':'grid_view'})

    for flag in a.find_all('li'):
        s={}
        hd=flag.find('div',attrs={'class':'hd'})
        name=hd.find('span',attrs={'class':'title'}).getText()
        s['Moviename']=name
        grade=flag.find('span',attrs={'class':'rating_num'}).getText()
        s['Grade']=grade
        star=flag.find('div',attrs={'class':'star'})
        pj=star.find_all('span')[-1].getText()
        s['pj']=pj
        dp=flag.find('span',attrs={'class':'inq'})
        if(dp):
            s['dp']=dp.getText()
        else:
            s['dp']='无短评'

    ulist.append(s)
```

3. 将爬取的内容输入到数据库:

(1) 首先在 Mysql 中创建数据表名为 mymovie:

[表设计] mymovie @test (admin)

文件(F) 编辑(E) 窗口(W)

创建 保存 另存为 创建栏位 插入栏位 删除栏位 主键 上移 下移

| 栏位        | 索引 | 外键 | 触发器 | 选项 | 注记 | SQL 预览   |
|-----------|----|----|-----|----|----|--|
| 名         |    |    |     |    |    |  |
| Moviename |    |    |     |    |    | varchar 100 0 允许空值 <input checked="" type="checkbox"/> |
| Grade     |    |    |     |    |    | varchar 30 0 允许空值 <input checked="" type="checkbox"/>  |
| pj        |    |    |     |    |    | varchar 50 0 允许空值 <input checked="" type="checkbox"/>  |
| dp        |    |    |     |    |    | varchar 100 0 允许空值 <input checked="" type="checkbox"/> |

默认: Empty String

注记:

字符集: utf8

整理: utf8\_bin

键长度:

☐ 二进制

(2) 用代码将列表内数据输入到数据库中:

```
def Save(ulist):
    try:
        db=pymysql.connect(host="localhost",user="root",password="admin",database="test",charset="utf8")
        cursor=db.cursor()
        ls={}
        for ls in ulist:
            cursor.execute("insert into mymovie(Moviename,Grade,pj,dp) values(%s,%s,%s,%s)",(ls['Moviename'],ls['Grade'],ls['pj'],ls['dp']))
            db.commit()
        cursor.close()
        db.close()
    except:
        print("错误！")
```

[表] mymovie @test (admin)

文件(F) 编辑(E) 查看(V) 窗口(W)

导入向导(I) 导出向导(O) 筛选向导 网格视图 表单视图 备注 十六进制 图像 升幂排序

| Moviename | Grade | pj         | dp                          |
|-----------|-------|------------|-----------------------------|
| 肖申克的救赎    | 9.6   | 1399258人评价 | 希望让人自由。                     |
| 霸王别姬      | 9.6   | 1035891人评价 | 风华绝代。                       |
| 这个杀手不太冷   | 9.4   | 1277857人评价 | 怪蜀黍和小萝莉不得不说的爱               |
| 阿甘正传      | 9.4   | 1101890人评价 | 一部美国近现代史。                   |
| 美丽人生      | 9.5   | 645147人评价  | 最美的谎言。                      |
| 泰坦尼克号     | 9.3   | 1042365人评价 | 失去的才是永恒的。                   |
| 千与千寻      | 9.3   | 1027709人评价 | 最好的宫崎骏，最好的久石让               |
| 辛德勒的名单    | 9.5   | 575295人评价  | 拯救一个人，就是拯救整个世界              |
| 盗梦空间      | 9.3   | 1108197人评价 | 诺兰给了我们一场无法盗取的梦              |
| 忠犬八公的故事   | 9.3   | 730369人评价  | 永远都不能忘记你所爱的人                |
| 机器人总动员    | 9.3   | 734203人评价  | 小瓦力，大人生。                    |
| 三傻大闹宝莱坞   | 9.2   | 995323人评价  | 英俊版憨豆，高情商版谢耳朵               |
| 海上钢琴师     | 9.2   | 817015人评价  | 每个人都要走一条自己坚定的路              |
| 放牛班的春天    | 9.3   | 689082人评价  | 天籁一般的童声，是最接近天堂的声音           |
| 楚门的世界     | 9.2   | 761048人评价  | 如果再也不能见到你，祝你一生幸福。           |
| 大话西游之大圣娶亲 | 9.2   | 769689人评价  | 一生所爱。                       |
| 星际穿越      | 9.2   | 790305人评价  | 爱是一种力量，让我们超越时间和空间           |
| 龙猫        | 9.2   | 680577人评价  | 人人心中都有个龙猫，童年最美好的回忆          |
| 教父        | 9.2   | 499545人评价  | 千万不要记恨你的对手，因为你迟早会和他见面       |
| 熔炉        | 9.3   | 445191人评价  | 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 |
| 无间道       | 9.1   | 632073人评价  | 香港电影史上永不过时的杰作               |
| 疯狂动物城     | 9.2   | 867737人评价  | 迪士尼给我们营造的乌托邦国度              |
| 当幸福来敲门    | 9.0   | 805422人评价  | 平民励志片。                      |
| 怦然心动      | 9.0   | 889387人评价  | 真正的幸福是来自内心深处                |
| 触不可及      | 9.2   | 532318人评价  | 满满温情的高雅喜剧。                  |

680577人评价

SELECT \* FROM `mymovie` LIMIT 18 / 250 于页 1

(3)。将数据库内的内容输出并且展示在 django 框架中：

```
def Mysqlfind():
    db=pymysql.connect(host="localhost",user="root",password="admin",database="test",charset="utf8",cursorclass=pymysql.cursors.DictCursor)
    cursor=db.cursor()
    cursor.execute("select * from mymovie")
    ulist=cursor.fetchall()
    cursor.close()
    db.close()
    return ulist
```

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：

```
def hello(request):  
    ulist=Mysqlfind()  
    return render(request, 'hello.html',{'movie':ulist})  
def hello1(request):
```

url.py 文件:

```
from django.urls import path  
  
from . import view  
  
urlpatterns = [  
    path('hello/', view.hello),  
    path('hello1/', view.hello1),  
]
```

Hello.html 文件:

```
view.py  base.html  s1.py  settings.py

1  <!DOCTYPE html>
2  <html>
3  <head>
4  <meta charset="utf-8">
5  <title>豆瓣电影爬取页面</title>
6  </head>
7  <body>
8      <h1>豆瓣电影TOP250</h1>
9      <table>
10         <tr>
11             <td>电影名称</td>
12             <td>电影评分</td>
13             <td>电影评价</td>
14             <td>电影短评</td>
15         </tr>
16         {% for Mymovie in movie %}
17         <tr>
18             <td>{{Mymovie.Moviename}}</td>
19             <td>{{Mymovie.Grade}}</td>
20             <td>{{Mymovie.pj}}</td>
21             <td>{{Mymovie.dp}}</td>
22         </tr>
23         {% endfor %}
24     </table>
25 </body>
26 </html>
```

运行后的效果：

## 豆瓣电影TOP250

| 电影名称      | 电影评分 | 电影评价       | 电影短评                            |
|-----------|------|------------|---------------------------------|
| 肖申克的救赎    | 9.6  | 1399258人评价 | 希望让人自由。                         |
| 霸王别姬      | 9.6  | 1035891人评价 | 风华绝代。                           |
| 这个杀手不太冷   | 9.4  | 1277857人评价 | 怪蜀黍和小萝莉不得不说的故事。                 |
| 阿甘正传      | 9.4  | 1101890人评价 | 一部美国近现代史。                       |
| 美丽人生      | 9.5  | 645147人评价  | 最美的谎言。                          |
| 泰坦尼克号     | 9.3  | 1042365人评价 | 失去的才是永恒的。                       |
| 千与千寻      | 9.3  | 1027709人评价 | 最好的宫崎骏，最好的久石让。                  |
| 辛德勒的名单    | 9.5  | 575295人评价  | 拯救一个人，就是拯救整个世界。                 |
| 盗梦空间      | 9.3  | 1108197人评价 | 诺兰给了我们一场无法盗取的梦。                 |
| 忠犬八公的故事   | 9.3  | 730369人评价  | 永远都不能忘记你所爱的人。                   |
| 机器人总动员    | 9.3  | 734203人评价  | 小瓦力，大人生。                        |
| 三傻大闹宝莱坞   | 9.2  | 995323人评价  | 英俊版憨豆，高情商版谢耳朵。                  |
| 海上钢琴师     | 9.2  | 817015人评价  | 每个人都要走一条自己坚定了的路，就算是粉身碎骨。        |
| 放牛班的春天    | 9.3  | 689082人评价  | 天籁一般的童声，是最接近上帝的存在。              |
| 楚门的世界     | 9.2  | 761048人评价  | 如果再也不能见到你，祝你早安，午安，晚安。           |
| 大话西游之大圣娶亲 | 9.2  | 769689人评价  | 一生所爱。                           |
| 星际穿越      | 9.2  | 790305人评价  | 爱是一种力量，让我们超越时空感知它的存在。           |
| 龙猫        | 9.2  | 680577人评价  | 人人心中都有个龙猫，童年就永远不会消失。            |
| 教父        | 9.2  | 499545人评价  | 千万不要记恨你的对手，这样会让你失去理智。           |
| 熔炉        | 9.3  | 445191人评价  | 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。    |
| 无间道       | 9.1  | 632073人评价  | 香港电影史上永不过时的杰作。                  |
| 疯狂动物城     | 9.2  | 867737人评价  | 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。 |
| 当幸福来敲门    | 9.0  | 805422人评价  | 平民励志片。                          |
| 怦然心动      | 9.0  | 889387人评价  | 真正的幸福是来自内心深处。                   |
| 触不可及      | 9.2  | 532318人评价  | 满满温情的高雅喜剧。                      |
| 蝙蝠侠：黑暗骑士  | 9.1  | 513573人评价  | 无尽的黑暗。                          |
| 乱世佳人      | 9.2  | 370632人评价  | Tomorrow is another day         |

## 二．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：

```
def dgetText(url):
    driver = webdriver.Chrome()
    driver.get(url)
    for i in range(10):
        driver.execute_script("var q=document.documentElement.scrollTop={0}".format(i*1000))
        time.sleep(1)
    searchprice=driver.find_elements_by_xpath('//*[@id="J_goodsList"]/ul/li/div/div[3]/strong/i')
    searchname=driver.find_elements_by_xpath('//*[@id="J_goodsList"]/ul/li/div/div[4]/a/em')
    names=[]
    for a in searchname:
        name=a.text
        names.append(name)
    prices=[]
    for c in searchprice:
        price=c.text
        prices.append(price)
    ulists=[]
    flag=0
    for b in names:
        ulist={}
        ulist['name']=b
        ulist['price']=prices[flag]
        flag=flag+1
        ulists.append(ulist)
    return ulists
```



2. 创建名为 Myphone 的数据表:

[表设计] Myphone @test (admin)

文件(F) 编辑(E) 窗口(W)

创建 保存 另存为 创建栏位 插入栏位 删除栏位 主键 上移 下移

栏位 索引 外键 触发器 选项 笔记 SQL 预览

| 名     | 类型      | 长度  | 十进制 | 允许空值                                |  |
|-------|---------|-----|-----|-------------------------------------|--|
| name  | varchar | 200 | 0   | <input checked="" type="checkbox"/> |  |
| price | varchar | 10  | 0   | <input checked="" type="checkbox"/> |  |

默认: Empty String

笔记:

字符集: utf8

整理: utf8\_bin

键长度:

☐ 二进制

宝怡勒, 2

3. 将数据保存到 Myphone 数据表中:

```
def dSave(ulist):
    try:
        db=pymysql.connect(host="localhost",user="root",password="admin",database="test",charset="utf8")
        cursor=db.cursor()
        ls={}
        for ls in ulist:
            cursor.execute("insert into Myphone(name,price) values(%s,%s)",(ls['name'],ls['price']))
            db.commit()
        cursor.close()
        db.close()
    except:
        print("错误！")

def main():
    key='手机'
    url='https://search.jd.com/Search?keyword='+key+'&enc=utf-8'
    ulists=[]
    ulists=dgetText(url)
    dSave(ulists)
```

[表] Myphone @test (admin)

文件(F) 编辑(E) 查看(V) 窗口(W)

导入向导(I) 导出向导(O) 筛选向导 网络视图 表单视图 备注 十六进制 图像 升幂排序

| name   | price   |
|--|---------|
| 【预售】魅族 16s 骁龙855全面屏拍照游戏手机 6GB+128GB 碳纤维 全网通移动联通电信4G 双卡双待               | 3198.00 |
| 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               | 1298.00 |
| Apple iPhone XR (A2108) 128GB 黑色 移动联通电信4G手机 双卡双待                       | 5699.00 |
| 【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      | 3298.00 |
| 华为 HUAWEI P30 超感光徕卡三摄麒麟980AI智能芯片全面屏屏内指纹版手机8GB+64GB亮黑色全网通双4G双           | 3988.00 |
| 小米 红米Redmi Note7 幻彩渐变AI双摄 4GB+64GB 梦幻蓝 全网通4G 双卡双待 水滴全面屏拍照游戏智能          | 1199.00 |
| 荣耀畅玩8C两天一充 莱茵护眼 刘海屏 全网通版4GB+32GB 幻夜黑 移动联通电信4G全面屏 双卡双待                  | 898.00  |
| vivo U1 水滴全面屏 AI智慧拍照手机 3GB+32GB 极光色 移动联通电信全网通4G                        | 799.00  |
| 联想Z6 Pro 8GB+128GB 黑色 骁龙855 4800万AI四摄 4000mAh大电池 PC级液冷散热 游戏 全网通4G 双卡双待 | 2999.00 |
| 荣耀V20 胡歌同款 麒麟980芯片 魅眼全视屏 4800万深感相机 6GB+128GB 幻夜黑 移动联通电信4G全面屏           | 2798.00 |
| 荣耀10青春版 幻彩渐变 2400万AI自拍 全网通版4GB+64GB 渐变蓝 移动联通电信4G全面屏 双卡双待               | 1299.00 |
| 小米9 4800万超广角三摄 8GB+128GB全息幻彩蓝 骁龙855 全网通4G 双卡双待 水滴全面屏拍照游戏智能             | 3299.00 |
| 小米8SE 全面屏智能游戏拍照手机 6GB+64GB 灰色 骁龙710处理器 全网通4G 双卡双待                      | 1399.00 |
| 小米8青春版 镜面渐变AI双摄 6GB+64GB 梦幻蓝 骁龙 全网通4G 双卡双待 全面屏拍照游戏智能                   | 1499.00 |
| 小米 红米Redmi 7 AI双摄 3GB+32GB 亮黑色 全网通4G 双卡双待 水滴全面屏拍照游戏智能                  | 799.00  |
| vivo Z3 6GB+64GB 极光蓝 性能实力派 全面屏游戏手机 移动联通电信全网通4G手机                       | 1598.00 |
| 三星 Galaxy S10+ 8GB+128GB炭晶黑 (SM-G9750) 3D超声波屏下指纹超感官全视屏骁龙855双卡双待全网通4    | 6999.00 |
| 黑鲨游戏手机2 8GB+128GB 暗影黑 骁龙855 Magic Press立体操控 塔式全域液冷 全面屏 全网通4G 双卡双待      | 3499.00 |
| 荣耀10 GT游戏加速 AIS手持夜景 6GB+64GB 幻夜黑 全网通 移动联通电信4G 双卡双待 游戏                  | 2198.00 |
| Apple iPhone X (A1865) 64GB 深空灰色 移动联通电信4G手机                            | 6099.00 |
| vivo S1 6GB+128GB 冰湖蓝 2480万AI高清自拍 超广角后置三摄拍照手机 移动联通电信全网通4G              | 2298.00 |
| vivo X27 8GB+256GB大内存 雀羽蓝 4800万AI三摄全面屏拍照手机 移动联通电信全网通4G                 | 3598.00 |
| 小米 红米6A AI美颜 3GB+32GB 流沙金 全网通4G手机 双卡双待                                 | 649.00  |
| 小米 红米Redmi Note7Pro AI双摄 6GB+128GB 梦幻蓝 全网通4G 双卡双待 水滴屏拍照游戏              | 1599.00 |
| 三星 Galaxy S10e 6GB+128GB 沁柠黄 (SM-G9700) 超感官全视屏 骁龙855 双卡双待 全网通4G        | 4999.00 |
| Apple iPhone 8 (A1863) 64GB 深空灰色 移动联通电信4G手机                            | 3799.00 |
| 荣耀20i 3200万AI自拍 超广角三摄 全网通版6GB+64GB 幻夜黑 移动联通电信4G全面屏 双卡双待                | 1599.00 |
| Apple iPhone 8 Plus (A1864) 64GB 深空灰色 移动联通电信4G手机                       | 4699.00 |

4. 将数据表内的内容输出到 django 框架:

```
def hello1(request):
    '''key='手机'
    url='https://search.jd.com/Search?keyword='+key+'&enc=utf-8'
    ulists=[]
    ulists=dgetText(url)
    dsave(ulists)'''
    ulist=dMysqlfind()
    return render(request, 'base.html',{'phone':ulist})

def Mysqlfind():
```

5. urls.py 文件:

```

from django.urls import path

from . import view

urlpatterns = [
    path('hello/', view.hello),
    path('hello1/', view.hello1),
]

```

Base.html 文件:

```

view.py  base.html x  s1.py  settings.py
1  <!DOCTYPE html>
2  <html>
3  <head>
4  <meta charset="utf-8">
5  <title>动态爬取京东手机页面</title>
6  </head>
7  <body>
8      <h1>京东手机</h1>
9      <table>
10         <tr>
11             <td>手机名称</td>
12             <td>手机价格</td>
13         </tr>
14         {% for Myphone in phone %}
15         <tr>
16             <td>{{Myphone.name}}</td>
17             <td>{{Myphone.price}}</td>
18         </tr>
19         {% endfor %}
20     </table>
21 </body>
22 </html>

```

运行后输出的网页:

## 京东手机

| 手机名称   | 手机价格    |
|--|---------|
| 【预售】魅族 16s 骁龙855全面屏拍照游戏手机 6GB+128GB 碳纤维黑 全网通移动联通电信4G 双卡双待              | 3198.00 |
| 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               | 1298.00 |
| Apple iPhone XR (A2108) 128GB 黑色 移动联通电信4G手机 双卡双待                       | 5699.00 |
| 【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      | 3298.00 |
| 华为 HUAWEI P30 超感光徇卡三摄麒麟980AI智能芯片全面屏屏内指纹版手机8GB+64GB亮黑色全网通双4G双           | 3988.00 |
| 小米 红米Redmi Note7 幻彩渐变AI双摄 4GB+64GB 梦幻蓝 全网通4G 双卡双待 水滴全面屏拍照游戏智能          | 1199.00 |
| 荣耀畅玩8C两天一充 莱茵护眼 刘海屏 全网通版4GB+32GB 幻夜黑 移动联通电信4G全面屏 双卡双待                  | 898.00  |
| vivo U1 水滴全面屏 AI智慧拍照手机 3GB+32GB 极光色 移动联通电信全网通4G                        | 799.00  |
| 联想Z6 Pro 8GB+128GB 黑色 骁龙855 4800万AI四摄 4000mAh大电池 PC级液冷散热 游戏 全网通4G 双卡双待 | 2999.00 |
| 荣耀V20 胡歌同款 麒麟980芯片 魅眼全视屏 4800万深感相机 6GB+128GB 幻夜黑 移动联通电信4G全面屏           | 2798.00 |
| 荣耀10青春版 幻彩渐变 2400万AI自拍 全网通版4GB+64GB 渐变蓝 移动联通电信4G全面屏 双卡双待               | 1299.00 |
| 小米9 4800万超广角三摄 8GB+128GB全息幻彩蓝 骁龙855 全网通4G 双卡双待 水滴全面屏拍照游戏智能             | 3299.00 |
| 小米8SE 全面屏智能游戏拍照手机 6GB+64GB 灰色 骁龙710处理器 全网通4G 双卡双待                      | 1399.00 |
| 小米8青春版 镜面渐变AI双摄 6GB+64GB 梦幻蓝 骁龙 全网通4G 双卡双待 全面屏拍照游戏智能                   | 1499.00 |
| 小米 红米Redmi 7 AI双摄 3GB+32GB 亮黑色 全网通4G 双卡双待 水滴全面屏拍照游戏智能                  | 799.00  |
| vivo Z3 6GB+64GB 极光蓝 性能实力派 全面屏游戏手机 移动联通电信全网通4G手机                       | 1598.00 |
| 三星 Galaxy S10+ 8GB+128GB炭晶黑 (SM-G9750) 3D超声波屏下指纹超感官全视屏骁龙855双卡双待全网通4G   | 6999.00 |
| 黑鲨游戏手机2 8GB+128GB 暗影黑 骁龙855 Magic Press立体操控 塔式全域液冷 全面屏 全网通4G 双卡双待      | 3499.00 |
| 荣耀10 GT游戏加速 AIS手持夜景 6GB+64GB 幻夜黑 全网通 移动联通电信4G 双卡双待 游戏                  | 2198.00 |
| Apple iPhone X (A1865) 64GB 深空灰色 移动联通电信4G手机                            | 6099.00 |
| vivo S1 6GB+128GB 冰湖蓝 2480万AI高清自拍 超广角后置三摄拍照手机 移动联通电信全网通4G              | 2298.00 |
| vivo X27 8GB+256GB大内存 雀羽蓝 4800万AI三摄全面屏拍照手机 移动联通电信全网通4G                 | 3598.00 |
| 小米 红米6A AI美颜 3GB+32GB 流沙金 全网通4G手机 双卡双待                                 | 649.00  |
| 小米 红米Redmi Note7Pro AI双摄 6GB+128GB 梦幻蓝 全网通4G 双卡双待 水滴屏拍照游戏              | 1599.00 |
| 三星 Galaxy S10e 6GB+128GB 沁柠黄 (SM-G9700) 超感官全视屏 骁龙855 双卡双待 全网通4G        | 4999.00 |
| Apple iPhone 8 (A1863) 64GB 深空灰色 移动联通电信4G手机                            | 3799.00 |

## 期末作业

自动登入 12306:

1.

超级鹰网站（帮助识别验证码）:

```
class Chaojiying_Client(object):

    def __init__(self, username, password, soft_id):
        self.username = username
        password = password.encode('utf8')
        self.password = md5(password).hexdigest()
        self.soft_id = soft_id
        self.base_params = {
```

```

        'user': self.username,
        'pass2': self.password,
        'softid': self.soft_id,
    }
    self.headers = {
        'Connection': 'Keep-Alive',
        'User-Agent': 'Mozilla/4.0 (compatible; MSIE 8.0; Windows NT 5.1;
Trident/4.0)',
    }

    def PostPic(self, im, codetype):
        params = {
            'codetype': codetype,
        }
        params.update(self.base_params)
        files = {'userfile': ('12306.jpg', im)}
        r =
requests.post('http://upload.chaojiying.net/Upload/Processing.php',
data=params, files=files, headers=self.headers)
        return r.json()

    def ReportError(self, im_id):
        params = {
            'id': im_id,
        }
        params.update(self.base_params)
        r =
requests.post('http://upload.chaojiying.net/Upload/ReportError.php',
data=params, headers=self.headers)
        return r.json()

```

用户登录:

```

def login(self):
    self.browser = webdriver.Chrome()
    self.browser.get(self.url)
    time.sleep(2)
    login_select = self.browser.find_element_by_class_name('login-hd-account')
    login_select.click()
    user = self.browser.find_element_by_id('J-username')
    word = self.browser.find_element_by_id('J-password')
    user.send_keys(self.username)
    word.send_keys(self.password)

```

图片识别:

```
def get_pic(self):
    tag = self.browser.find_element_by_class_name('imgCode')
    temp = tag.get_attribute('src')
    b64_pic = re.sub(r'data:image/jpg;base64,', '', temp)
    pic = base64.b64decode(b64_pic)
    return pic
```

自动点击正确验证码:

```
def click(self, j):
    temp = j.get('pic_str')
    locations = [list(map(int, i.split(','))) for i in temp.split('|')]
    for location in locations:
        ActionChains(self.browser).move_to_element_with_offset(self.browser.find_element_by_class_name('imgCode'),
                                                                location[0], location[1]).click().perform()
        time.sleep(5)
    self.browser.find_element_by_id('j-login').click()

def get_cookies(self):
    return self.browser.get_cookies()
```

主函数:

```
if __name__ == '__main__':
    a = Login('https://kyfw.12306.cn/otn/resources/login.html', '15858772795', 'chenxiangniu')
    a.login()
    time.sleep(2)
    chaojiying = Chaojiying_Client('15858772795', 'chenxiangniu110.', '900049')
    im = a.get_pic()
    print(im)
    z = chaojiying.PostPic(im, 9004)
    print(z)
    a.click(z)
    print(a.get_cookies())
```

效果:

扫码登录

账号登录

15858772795

\*\*\*\*\*

请点击下图中的所有 **跑步机**

刷新





立即登录

注册12306账号 | 忘记密码?

1、12306.cn网站每日06:00~23:00提供服务。  
2、在12306.cn网站购票、改签和退票须不晚于开车前30分钟；办理“变更到站”业务时，请不晚于开车前48小时。

餐饮·特产

带有温度的旅途的配餐





| 图片  | 类型   | 识别结果           | 状态 | 题分 | 软件ID   | 时间   | IP地址         |
|---|------|----------------|----|----|--------|--|--------------|
| <p>请点击下图中的所有 <b>跑步机</b></p>  <p>图片ID:2071420102525700004</p> | 9004 | 186,82         | OK | 25 | 900049 | 上传:2019-06-18<br>20:10:54<br>识别:2019-06-18<br>20:11:00 | 218.75.17.77 |
| <p>请点击下图中的所有 <b>口哨</b></p>  <p>图片ID:2071419502525700003</p>  | 9004 | 29,153 246,144 | OK | 25 | 900049 | 上传:2019-06-18<br>19:50:43<br>识别:2019-06-18<br>19:50:52 | 218.75.17.77 |
| <p>请点击下图中的所有 <b>印章</b></p>  <p>图片ID:9071412512525700002</p> | 9004 | 29,71          | OK | 25 | 900049 | 上传:2019-06-18<br>12:51:07<br>识别:2019-06-18<br>12:51:12 | 218.75.17.74 |
| <p>请点击下图中的所有 <b>印章</b></p>                                 | 9004 | 265,130        | OK | 25 | 900049 | 上传:2019-06-17<br>20:17:38<br>识别:2019-06-17<br>20:17:42 | 218.75.17.78 |