16219111203 沈阳阳

1

静态

通过 beautisoup 库爬取豆瓣电影 top250 的内容。存入 sqlserver。再对数据库进行读取,通过 render 方法,以字典列表形式回传到页面展示

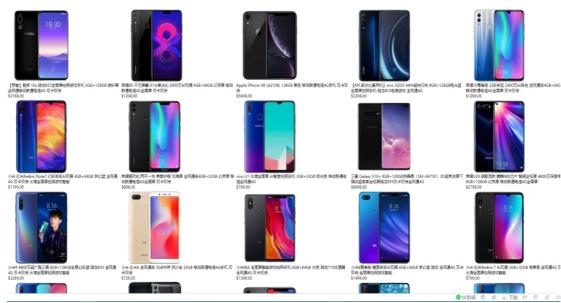


2

动态

用 selenium 通过谷歌浏览器对京东手机爬取,用到了页面下拉加载,进行动态爬取。

京东手机



3 自动登录 12306

```
from selenium import webdriver
import requests
from hashlib import md5
import base64
from selenium.webdriver.common.action_chains import ActionChains
class Chaojiying_Client(object):
        self.username = username
        password = password.encode('utf8')
        self.password = md5(password).hexdigest()
        self.soft id = soft id
        self.base_params = {
            'pass2': self.password,
            'softid': self.soft id,
            '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)
requests.post('http://upload.chaojiying.net/Upload/ReportError.php',
data=params, headers=self.headers)
        return r. json()
class Login:
    def init (self, url, username, password):
        self.url = url
        self.password = password
    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-
        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)
    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)
    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_eleme nt_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()

if __name__ == '__main__':

a = Login('https://kyfw.12306.cn/otn/resources/login.html', '账号', '密码')

a.login()

time.sleep(2)

chaojiying = Chaojiying_Client('s674560845', 's13958920775', '900033'))

im = a.get_pic()

print(im)

z = chaojiying.PostPic(im, 9004)

print(z)

a.click(z)

print(a.get_cookies())
```

