表单交互：Request中的POST方法模拟登陆。

某些网站如果不登陆，无法获取到有效的信息。

例：程序crawler\_test/crawler\_test14.py

import requests  
import json  
import time  
import pymongo  
from mongodb\_test import mongo\_host  
from mongodb\_test import mongo\_port  
  
client = pymongo.MongoClient(mongo\_host, mongo\_port)  
mydb = client['mydb']  
lagou = mydb['lagou']  
  
# 拉勾网有反爬虫机制，表头需要尽可能贴近实际浏览器操作  
headers = {  
 'Host': 'www.lagou.com',  
 'Accept': 'pplication/json, text/javascript, \*/\*; q=0.01',  
 'Accept-Encoding': 'gzip, deflate, br',  
 'Accept-Language': 'zh-CN,zh;q=0.9',  
 'Content-Type': 'application/x-www-form-urlencoded; charset=UTF-8',  
 'X-Requested-With': 'XMLHttpRequest',  
 'X-Anit-Forge-Code': '0',  
 'X-Anit-Forge-Token': 'None',  
 'Content-Length': '25',  
 'Pragma': 'no-cache',  
 'Cache-Control': 'no-cache',  
 'Referer': 'https://www.lagou.com/jobs/list\_python?px=default&city=%E5%85%A8%E5%9B%BD',  
 'Cookie': '\_ga=GA1.2.795752720.1530950987; user\_trace\_token=20180707160948-1dbd13fc-81bd-11e8-993c-5254005c3644; LGSID=20180707160948-1dbd1658-81bd-11e8-993c-5254005c3644; LGUID=20180707160948-1dbd17f0-81bd-11e8-993c-5254005c3644; \_gid=GA1.2.2018763107.1530950988; LG\_LOGIN\_USER\_ID=4ae2ab36b3b2fac6733377dec13c214e9d7d59cb0769d6d1281cc6d5626eae77; showExpriedIndex=1; showExpriedCompanyHome=1; showExpriedMyPublish=1; hasDeliver=0; index\_location\_city=%E5%8C%97%E4%BA%AC; JSESSIONID=ABAAABAAAFCAAEGC2DA816AB896660A4B210B2358C0B0BF; \_gat=1; Hm\_lvt\_4233e74dff0ae5bd0a3d81c6ccf756e6=1530950987,1530952899; \_putrc=F8AB6C83954FB97E123F89F2B170EADC; login=true; unick=%E5%88%98%E7%8F%85%E7%8F%85; gate\_login\_token=ca9b4558469c71bf3529bfce756b4bba9c1eaeb063ca64d688b5a35836dd8300; TG-TRACK-CODE=index\_search; Hm\_lpvt\_4233e74dff0ae5bd0a3d81c6ccf756e6=1530952920; LGRID=20180707164200-9d6f4a9a-81c1-11e8-993c-5254005c3644; SEARCH\_ID=e07e0e81e6924933849d4fc7f497860c',  
 'User-Agent': 'Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/67.0.3396.87 Safari/537.36',  
 'Connection': 'keep-alive'  
}  
  
  
def get\_page(url, params):  
 html = requests.post(url, data=params, headers=headers)  
 json\_data = json.loads(html.text)  
 state = json\_data['success']  
 if state:  
 total\_count = json\_data['content']['positionResult']['totalCount']  
 page\_number = int(total\_count / 15) if int(total\_count / 15) < 30 else 30  
 get\_info(url, page\_number)  
 else:  
 print('操作过于频繁')  
  
  
def get\_info(url, page\_number):  
 for pn in range(1, page\_number + 1):  
 # Post请求参数  
 params = {  
 'first': 'true',  
 'pn': str(pn),  
 'kd': 'Python'  
 }  
  
 try:  
 html = requests.post(url, data=params, headers=headers)  
 json\_data = json.loads(html.text)  
 results = json\_data['content']['positionResult']['result']  
 for result in results:  
 infos = {  
 'businessZones': result['businessZones'],  
 'city': result['city'],  
 'companyFullName': result['companyFullName'],  
 'companyLabelList': result['companyLabelList'],  
 'companySize': result['companySize'],  
 'district': result['district'],  
 'education': result['education'],  
 'explain': result['explain'],  
 'financeStage': result['financeStage'],  
 'firstType': result['firstType'],  
 'formatCreateTime': result['formatCreateTime'],  
 'gradeDescription': result['gradeDescription'],  
 'imState': result['imState'],  
 'salary': result['salary']  
 }  
  
 lagou.insert\_one(infos)  
 time.sleep(2)  
 except requests.exceptions.ConnectionError:  
 pass  
  
  
if \_\_name\_\_ == '\_\_main\_\_':  
 url = 'https://www.lagou.com/jobs/positionAjax.json'  
  
 # Form data  
 params = {  
 'first': 'true',  
 'pn': '1',  
 'kd': 'Python'  
 }  
 get\_page(url, params)