json的解析：

例：程序crawler\_test/json\_test.py

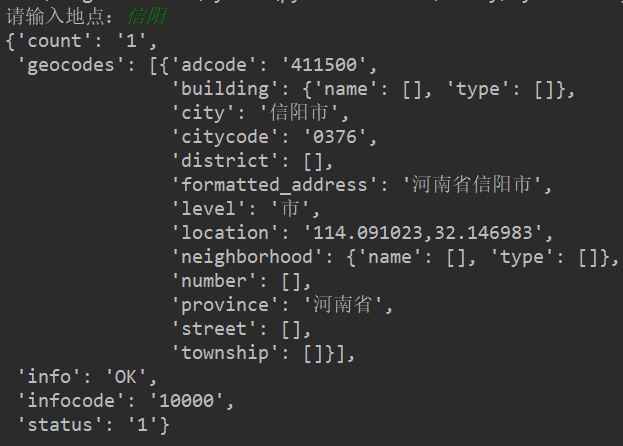
import json  
  
if \_\_name\_\_ == '\_\_main\_\_':  
 json\_str = '{"user\_man":[{"name":"Peter"},{"name":"xiaoming"}],"user\_woman":[{"name":"Anni"},{"name":"zhangsan"}]}'  
 json\_data = json.loads(json\_str)  
 print(json\_data.get("user\_man"))  
 print(json\_data.get("user\_man")[0].get("name"))

API调用：

例：程序crawler\_test/crawler\_test7.py

import requests  
import json  
import pprint  
  
if \_\_name\_\_ == '\_\_main\_\_':  
 address = input('请输入地点：')  
 par = {'address': address, 'key': 'cb649a25c1f81c1451adbeca73623251'}  
 url = 'http://restapi.amap.com/v3/geocode/geo'  
 res = requests.get(url, par)  
 json\_data = json.loads(res.text)  
 pprint.pprint(json\_data)

输出为：



爬取糗事百科地理信息

import requests  
from lxml import etree  
import csv  
import crawler\_test1  
import json  
  
fp = open('map.csv', 'wt', newline='', encoding='utf-8')  
writer = csv.writer(fp)  
writer.writerow(('address', 'longitude', 'latitude'))  
  
  
def get\_user\_url(url):  
 url\_part = 'http://www.qiushibaike.com'  
 html = requests.get(url)  
 selector = etree.HTML(html.text)  
 url\_infos = selector.xpath('//div[contains(@class,"article block untagged mb15")]') # 模糊查询  
 for url\_info in url\_infos:  
 user\_part\_urls = url\_info.xpath('div[1]/a[1]/@href')  
 if len(user\_part\_urls) == 1:  
 user\_part\_url = user\_part\_urls[0]  
 get\_user\_address(url\_part + user\_part\_url)  
 else:  
 pass  
  
  
# 获取用户地址  
def get\_user\_address(url):  
 res = requests.get(url, headers=crawler\_test1.headers)  
 selector = etree.HTML(res.text)  
 user\_info = selector.xpath('//div[@class="user-statis user-block"]/ul/li[4]/text()')  
  
 if len(user\_info) > 1:  
 address = user\_info[1].split('·')  
 if len(address) > 1:  
 get\_geo(address[1])  
 elif len(address) == 1:  
 get\_geo(address[0])  
 else:  
 pass  
  
  
# 利用API获取用户经纬度信息  
def get\_geo(address):  
 par = {'address': address, 'key': 'cb649a25c1f81c1451adbeca73623251'}  
 api = 'http://restapi.amap.com/v3/geocode/geo'  
 res = requests.get(api, par)  
 json\_data = json.loads(res.text)  
  
 try:  
 geo = json\_data['geocodes'][0]['location']  
 longitude = geo.split(',')[0]  
 latitude = geo.split(',')[1]  
 writer.writerow((address, longitude, latitude))  
 except IndexError:  
 pass  
  
  
if \_\_name\_\_ == '\_\_main\_\_':  
 urls = ['http://www.qiushibaike.com/8hr/page/{0}/'.format(str(i)) for i in range(1, 5)]  
 for url in urls:  
 get\_user\_url(url)  
  
 fp.close()