

# API爬虫案例

基于API爬虫的一般步骤：

- 1.在网站注册开发者账户用户名，获得相应的开发者密钥
- 2.在网站的API帮助说明文档中找到自己需要使用的API,确认API请求的限制次数,确认调用API需要使用的参数
- 3.在联网状态下，编写正确代码调用API
- 4.从API返回的内容（JSON格式）获取正确的属性
- 5.将获取的内容存储到本地（文件或数据库）

## 案例一：通过网易云音乐的API，获取相关音乐的评论

API说明文档：<https://api.imjad.cn/cloudmusic.md>

调用方法：GET <https://api.imjad.cn/cloudmusic/>

比如：<https://api.imjad.cn/cloudmusic/?type=comments&id=347230> 这是Beyond-海阔天空的评论

### 第一步，利用python获取内容

```
import requests

#id是必须要的参数，可以在网易云音乐的网站上搜索相应的歌曲，进入歌曲的页面后，可以在网址的后面看到id=xxxxxxx
#type参数是获取内容的类型，默认是song，而我们要获取评论，所有type要改成comments
#offset是偏移量，在这里可以认为是多少页
#其他参数可以去说明文档中查看
response = requests.get('https://api.imjad.cn/cloudmusic/?
type=comments&id=347230&offset=0')    #使用request获取内容
# print(response.text)
```

### 第二步，解析json

```
import json
data = json.loads(response.text)           #将返回的内容加载成json格式
print(data['hotComments'][0]['user']['nickname'])    #根据key取出value，如果看不清json的数
据，可以百度一个在线解析json的网站，将json格式的数据放进去解析一下
print(data['hotComments'][0]['content'])
```

代价慢慢  
绝对影响世界的乐队，词曲都是和平，奋斗，鼓励。可惜啊

### 第三步，保存数据

```
f = open('comment.txt', 'w') #打开文件
for i in data["hotComments"]:
    f.write('{} {} \n'.format(i['user']['nickname'], i['content'])) #逐条写入文件
f.close()
```

## 练习：获取更多的评论

可以通过改变offset的值来获取更多的内容

```
import requests
import time
import json
f = open('musicComments.txt', 'w', encoding= 'utf-8')
for i in range(10):
    response = requests.get('https://api.imjad.cn/cloudmusic/?
type=comments&id=347230&offset={}'.format(i))
    data = json.loads(response.text)
    #通过观察数据可以发下，第一页有精彩评论，需要单独处理
    if i == 0:
        for i in data["hotComments"]:
            f.write('{} {} \n'.format(i['user']['nickname'], i['content']))
    #除了精彩评论，其余的评论都是comments
    for i in data["comments"]:
        f.write('{} {} \n'.format(i['user']['nickname'], i['content']))
    time.sleep(3) #设置延时，避免给服务器造成太大压力
f.close()
```

## 案例二：通过聚合数据的API进行数据采集

首先，在聚合数据的网站上注册账号，申请需要的API。注意：免费的调用次数是有限制的，并且只能用部分API  
申请好后就可以去查看API的文档。在这里使用的是作文大全的API。

接口的地址为：<http://zuowen.api.juhe.cn/zuowen/typeList>

与不需要key的API相比，就多传入了一个key参数。

### 第一步，利用python获取内容

```
import requests
#注意你爬的API，根据你的需求进行修改
appkey = '你的key'
url = 'http://zuowen.api.juhe.cn/zuowen/content?id=11655&key=' + appkey #多传入了一个key参
数
response = requests.get(url)
# print(response.text)
```

```
{ "error_code" : 0 , "reason" : "success" , "result" : { "comment" : "这篇文章，作者用细腻的语言写出了自己的成长。由开始的桀骜，叛逆，渐渐的明白了生活的内涵，用心去体会生活中的事物，这就是成长。全文语言真实质朴，字里行间流露出真情。" , "content" : "<p>每一颗星星都闪烁着无尽的光辉，每一缕阳光都珍藏着如歌往事。匆匆的岁月一去不复返，留给我们的是无尽的回忆和许许多多的遗憾。走过了天真无邪的童年，走过了如诗如画的花季和潇洒的雨季，意气风发的我好像变得懂事、成熟、稳重了许多。我意识到，我长大了。</p><p>曾经只是浮想联翩，华而不实，现在才发现自己是多么的幼稚可笑；曾经任性、傲慢、倔强，现在却为自己的无知而深感不安；曾经懒惰，缺乏韧性和毅力，依赖思想严重，现在才懂得什么叫不畏艰难，坚持到底，享受到了从未有过的拼搏后的喜悦。一时间，我确实感觉到我在成长，在不断地成长。</p><p>曾经无忧无虑的我，现在忽然发现母亲的脸上不知什么时候增添了这么多的皱纹，满头黑发中也露出缕缕白发来。看着略显苍老的母亲，我怔住了，这就是我那个容光焕发、有着健康体魄的母亲吗？我突然感到一丝陌生。待我清醒过来，我才明白，母亲为我操劳过度，付出了太多太多，而我却.....我的心如刀绞一般，不住地流着泪，流着血，顿时，一种莫名的感动涌上心头。这时，我想起了考试失败时母亲微笑着安慰我、鼓励我的情形。其实，母亲才是最需要安慰、最需要理解的。我平生第一次感到母亲的伟大。十九岁的我，开始懂事了。</p><p>一向桀骜不驯、有些蛮横的我，总喜欢以取笑自己的朋友为乐，事事都爱压倒别人，现在才发现，其实朋友都是在让着我。这时，我感到自己是如此的渺小，我也为我能拥有这样的朋友而庆幸，而自豪。</p><p>一向逆反心理严重的我，从不接受老师的正确批评，甚至还要顶上几句嘴，并为此沾沾自喜，自鸣得意，现在却突然发现，老师时常在关注自己，经常叮嘱同学们多帮助自己。我突然感觉到，我遇到了一位敬业的、责任心强的伯乐，他是在用心热爱着这份事业，用心来教育学生。我也只有用心，才能够体会到这种心灵的感应。我又一次感动了。</p><p>成长告诉我，我不是只为自己而活着，因为关注我的人实在太多，对我寄予的厚望也太多，这使我感到一种压力。不！这不是压力，而是动力，是能够催人奋进的催化剂，这不正是我所急需的“氧气”吗？</p><p>我开始懂得生命的意义。带着这纯洁的感情，我进出生活的角角落落，或痛苦或快乐或感恩戴德，珍藏我所珍视的，遗忘给我伤害的，并努力挣脱一切烦琐的羁绊，向生命追溯最真实的本源。</p><p>我试着把自己融入到生活的潮水之中，没有先前的刻意，只是用心悄悄地去体会，深深地呼吸一下生活的气息，我突然觉得有了一些从未有过的而又不经意的充实。尝试着，经历一次灵魂的挣扎，良知的洗礼，自己在思考中成长，在“绿色”中飞扬。</p><p>聆听着周杰伦的那首旋律优美而又唱出我心声的歌——《我们在成长》，不经意间，我进入了梦乡，并且梦见了我想梦见的一切。</p>" , "school" : "高思教育" , "teacher" : "高思名师: 李美微" , "id" : 11655}}
```

## 第二步，解析json

```
import json
data = json.loads(response.text)
print(data['result']['content'])
```

每一颗星星都闪烁着无尽的光辉，每一缕阳光都珍藏着如歌往事。匆匆的岁月一去不复返，留给我们的的是无尽的回忆和许许多多的遗憾。走过了天真无邪的童年，走过了如诗如画的花季和潇洒的雨季，意气风发的我好像变得懂事、成熟、稳重了许多。我意识到，我长大了。

曾经只是浮想联翩，华而不实，现在才发现自己是多么的幼稚可笑；曾经任性、傲慢、倔强，现在却为自己的无知而深感不安；曾经懒惰，缺乏韧性和毅力，依赖思想严重，现在才懂得什么叫不畏艰难，坚持到底，享受到了从未有过的拼搏后的喜悦。一时间，我确实感觉到我在成长，在不断地成长。

曾经无忧无虑的我，现在忽然发现母亲的脸上不知什么时候增添了这么多的皱纹，满头黑发中也露出缕缕白发来。看着略显苍老的母亲，我怔住了，这就是我那个容光焕发、有着健康体魄的母亲吗？我突然感到一丝陌生。待我清醒过来，我才明白，母亲为我操劳过度，付出了太多太多，而我却.....我的心如刀绞一般，不住地流着泪，流着血，顿时，一种莫名的感动涌上心头。这时，我想起了考试失败时母亲微笑着安慰我、鼓励我的情形。其实，母亲才是最需要安慰、最需要理解的。我平生第一次感到母亲的伟大。十九岁的我，开始懂事了。

一向桀骜不驯、有些蛮横的我，总喜欢以取笑自己的朋友为乐，事事都爱压倒别人，现在才发现，其实朋友都是在让着我。这时，我感到自己是如此的渺小，我也为我能拥有这样的朋友而庆幸，而自豪。

一向逆反心理严重的我，从不接受老师的正确批评，甚至还要顶上几句嘴，并为此沾沾自喜，自鸣得意，现在却突然发现，老师时常在关注自己，经常叮嘱同学们多帮助自己。我突然感觉到，我遇到了一位敬业的、责任心强的伯乐，他是在用心热爱着这份事业，用心来教育学生。我也只有用心，才能够体会得到这种心灵的感应。我又一次感动了。

成长告诉我，我不是只为自己而活着，因为关注我的人实在太多，对我寄予的厚望也太多，这使我感到一种压力。不！这不是压力，而是动力，是能够催人奋进的催化剂，这不正是我所急需的“氧气”吗？

我开始懂得生命的意义。带着这纯洁的感情，我进出生活的角角落落，或痛苦或快乐或感恩戴德，珍藏我所珍视的，遗忘给我伤害的，并努力挣脱一切烦琐的羁绊，向生命追溯最真实的本源。

我试着把自己融入到生活的潮水之中，没有先前的刻意，只是用心悄悄地去体会，深深地呼吸一下生活的气息，我突然觉得有了一些从未有过的而又不经意的充实。尝试着，经历一次灵魂的挣扎，良知的洗礼，自己在思考中成长，在“绿色”中飞扬。

聆听着周杰伦的那首旋律优美而又唱出我心声的歌——《我们在成长》，不经意间，我进入了梦乡，并且梦见了我梦见的一切。

### 第三步，保存数据

```
f = open('zuowen.txt', 'w') #打开文件
f.write('{} {} \n'.format(data['result']['teacher'], data['result']['content']))
f.close()
```