

我是Python小白 完全不知道该如何去运行 但是看到别人的爬虫代码就很开心激动和羡慕

于是我就捉摸着自学与研究 我的电脑是 win10 64位的 安装的是Python3.5

环境变量就是把安装的路径添加到环境变量即可,不会可以搜.网上写的都比我的好.所以就不赘述了

先贴代码,是爬去网易云音乐的热评.

```
import re
import requests
import json
```

```
def get_all_hotsongs():
    """抓热搜榜所有歌曲"""
    url = 'http://music.163.com/discover/toplist?id=3778678'
    headers = {
        'Cookie': '__e_=1515461191756; _ntes_nnid=af802a7dd2cafc9fef605185da6e73fb,1515461190617; _ntes_nuid=af802a7dd2cafc9fef605185da6e73fb; JSESSIONID-WYYY=HMyeRdf98eDm%2Bi%5CRnK9iB%5ChcSODhA%2Bh4jx5t3z20hwwTRsOCWhBS5Cpn%2B5j%5CVf%5Iuqxldmzr_=32; __utma=94650624.61181594.1515583507.1515583507.1515583507.1; __utmc=94650624; __utmz=94650624.1515583507.1.1.utmcsr=(direct)|utmccn=(direct)|utmcmd=(none); __utmb=94650624.4.10.1515583507',
        'Host': 'music.163.com',
        'Refere': 'http://music.163.com/',
        'Upgrade-Insecure-Requests': '1',
        'User-Agent': 'Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/63.0.3239.132 Safari/537.36'
    }
    r = requests.get(url, headers=headers)
```

#使用正则表达式匹配正文响应

```
reg1 = r'<ul class="f-hide"><li><a href="/song?id=\d*>.*</a></li></ul>'
result_contain_songs_ul = re.compile(reg1).findall(r.text)
result_contain_songs_ul = result_contain_songs_ul[0]
```

```
reg2 = r'<li><a href="/song?id=\d*>(.*?)</a></li>'
reg3 = r'<li><a href="/song?id=(\d*>.*?</a></li>'
hot_songs_name = re.compile(reg2).findall(result_contain_songs_ul)
hot_songs_id = re.compile(reg3).findall(result_contain_songs_ul)
```

#返回歌曲名 歌曲id

```
return hot_songs_name, hot_songs_id
```

```
def get_hotcommnets(hot_songs_name, hot_songs_id):
    """抓热搜榜歌曲热评"""
    url = 'http://music.163.com/weapi/v1/resource/comments/R_SO_4_' + hot_songs_id + '?csrf_token='
    headers = {
        'Host': 'music.163.com',
        'Proxy-Connection': 'keep-alive',
```