

20377199 赵芮箐 第14周作业

作业内容： 协程有时被称为“微线程”，因为二者有类似的使用场景。和线程相比，协程占用资源更少、切换迅速，且实现简单（如协程是协作式调用，一般不用考虑资源的抢占，所以大部分情况下不需要通过同步原语来避免冲突）。通常协程被用在高并发的IO场景中，因此本周要求在第十二次作业的基础上，利用协程重写并扩充已有爬虫的功能。
注意：可以使用aiohttp, aiofiles等第三方库。

总体思路：

- 真的真的非常划水的一周了...
- 基本沿用12周的框架，主要是改了主函数，将多线程实现的部分改为了协程实现。
- 协程的实现依赖于gevent库，yysy gevent库真的清晰简洁好写！

协程实现：

```
import os
import csv
import time
import gevent
from PIL import Image
from io import BytesIO
from bs4 import BeautifulSoup
from gevent import monkey, pool
monkey.patch_all()      # 标记IO非阻塞，针对标准库
import requests

if __name__ == '__main__':
    async_time_start = time.time()
    head = ['歌单标题', 'id', '昵称', '介绍', '歌曲数量', '播放量', '添加到播放列表次数', '分享次数', '评论数']
    with open('week14/music_message.csv', 'w', encoding='utf-8') as f:
        csv_writer = csv.writer(f)
        csv_writer.writerow(head)

    p_jobs = []
    c_jobs = []
    p = pool.Pool(20)    # 控制数目
    for i in range(0, 1300, 35):
        url = 'https://music.163.com/discover/playlist/?cat=华语&order=hot&limit=35&offset=' + str(i)
        p_jobs.append(p.spawn(producer, url))
    gevent.joinall(p_jobs)
    for id in id_list:
        c_jobs.append(p.spawn(consumer, id))
    gevent.joinall(c_jobs)

    print("总耗时: %.4f" % (time.time() - async_time_start))
```

爬虫部分:

```
id_list = []
headers = {
    'User-Agent': 'Mozilla/5.0 (Windows NT 6.1; WOW64) AppleWebKit/537.36
(KHTML, like Gecko) Chrome/63.0.3239.132 Safari/537.36'
}

def producer(url):
    response = requests.get(url=url, headers=headers)
    html = response.text
    soup = BeautifulSoup(html, 'html.parser')
    ids = soup.select('.dec a') # 获取包含歌单详情页网址的标签
    id_list.extend(ids)

def consumer(id):
    url = 'https://music.163.com/' + id['href'] # 获取歌单详情页地址
    response = requests.get(url=url, headers=headers)
    html = response.text
    soup = BeautifulSoup(html, 'html.parser')
    # 获取歌单标题
    title = soup.find('h2', "f-ff2 f-brk").text
    # 获取封面图片
    img = soup.find('img', "j-img").get('data-src')
    res = requests.get(img)
    image = Image.open(BytesIO(res.content))
    try:
        image.save('week14/cover/' + title + '.jpg')
    except:
        pass
    # 获取创建者id
    idd = soup.find('a', "s-fc7")['href'].split('=')[-1]
    # 获取创建者昵称
    nickname = soup.find('a', "s-fc7").text
    # 获取歌单介绍
    text = soup.find('p', id="album-desc-more").text.replace('\n', '')
    # 歌单内歌曲数
    songs = soup.find('span', id="playlist-track-count").text
    # 播放量
    plays = soup.find('span', id="playlist-track-count").text
    # 添加进播放列表次数
    adds = soup.find('a', 'u-btni u-btni-fav')['data-count']
    # 分享次数
    shares = soup.find('a', 'u-btni u-btni-share')['data-count']
    # 歌单评论数
    comments = soup.find('span', id="cnt_comment_count").text
    # print(title, idd, nickname, text, songs, plays, adds, shares, comments)
    # 将详情页信息写入CSV文件中
    with open('week14/music_message.csv', 'a+', encoding='utf-8') as f:
        f.write(title + ',' + idd + ',' + nickname + ',' + text + ',' + songs +
        ',' + plays + ',' + adds + ',' + shares + ',' + comments + '\n')
```

结果展示：

- 运行时间：

```
1312 | 民谣 | 世界不小心生下普通的我,1361633352,蛋壳Maxd,介绍：当我立志要改变世界才
1313
问题 输出 调试控制台 终端
PS D:\code\mp2022> python -u "d:\code\mp2022\week14\week14.py"
总耗时： 232.7883
PS D:\code\mp2022>
```

1311个歌单总耗时 232秒

- 歌单信息：

	A	B	C	D	E	F	G	H	I
1	歌单标题	id	昵称	介绍	歌曲数量	播放量	添加到播放	分享次数	评论数
2	先努力让自己	450639585	四季里的懒睡猫	介绍：想	42	3756069	33491	364	668
3	只有几十秒	536588649	學會笑著搖頭拒	介绍：不	893	6666625	59320	865	1308
4	2022超好听的	543197267	、精神领袖V	介绍：有	202	2207	21	0	0
5	民谣 哪一	211868211	月亮藏不住的小	介绍：听	48	1021455	5345	120	11
6	倘若你喜欢	617541172	亲这边建议您不	介绍：你	94	2816124	20509	187	67
7	好听的民谣	554022595	朽念-	介绍：或	19	7631	33	1	1
8	民谣 不	148928651	贝克克y	介绍：春	46	1648143	8565	214	11
9	西域天国	287163860	海峡好歌单	介绍：全	75	885556	6345	64	8
10	摇滚夏日	341030416	十三逆旅Cobair	介绍：总	260	7838879	103469	1804	231
11	【翻唱】	176836719	top1小零	介绍：1	327	8755742	60739	652	317
12	夜色渐晚	450639585	四季里的懒睡猫	介绍：我	45	132512	2465	65	97
13	被迫营业	201586	原创君	介绍：金	28	213887	810	18	13
14	小酒馆里的	359904556	开心疗养院	介绍：民	575	6585937	50313	609	97
15	【电子音乐	252332327	邱小白Time	介绍：这	105	607478	701	35	17
16	隔壁老樊	596499189	Just-7ccc	介绍：愿	83	14918665	144222	1200	381
17	孤独旅人	211868211	月亮藏不住的小	介绍：你	62	35102	203	2	0
18	听完就去	380346613	6D英语专八	介绍：旋	52	855946	3523	335	0
19	你是非常	48115956	杉茶先生	介绍：你	57	14947846	158114	2916	1259
20	华语经典	211868211	月亮藏不住的小	介绍：时	71	557741	3784	28	17
21	说唱诗人	97515301	Bzstrive	介绍：说	51	512167	3031	42	11
22	民谣 我	211868211	月亮藏不住的小	介绍：你	37	335515	1535	13	3
23	赫云 奔向	1	网易云音乐	介绍：赫	23	3484	32	0	3
24	温柔男嗓	565918355	is慎之	介绍：你	268	13101056	110330	1529	450
25	「古风」	129593031	mayuko然	介绍：细	193	10281072	197190	2496	620
26	【民谣】	54278024	零下Young	介绍：如	109	14876397	161722	2805	633
27	经典怀旧	148928651	贝克克y	介绍：老	151	39382	398	1	0
28	华语 我	211868211	月亮藏不住的小	介绍：有	159	854896	4303	48	32
29	民谣 岁月	124272747	夕日向暖	介绍：选	141	8209343	62475	750	254
30	民谣暖调	158057598	迷藏lb	介绍：人	68	2311	4	0	0
31	给过去的自	1	网易云音乐	介绍：每	45	109391	504	20	48
32	【戏腔三	290894929	执剑稳	介绍：戏	108	40823	806	1	5
33	国语经典	148928651	贝克克y	介绍：听	185	2850098	12770	189	173
34	中文跑步	93983257	--苏麻离青--	介绍：跑	99	4033	41	2	2
35	古风 各大	200804198	从不洗裤	介绍：本	71	675556	9945	102	19
36	治愈的风	97515301	Bzstrive	介绍：每	89	4589904	35031	627	105
37	华语回忆	109390499	这条平凡的狗_C	介绍：华	141	9634172	45699	585	123
38	华语民谣	272654740	寒烟入暮戏入骨	介绍：按	22	4022	21	0	1
39	奔向月亮	450639585	四季里的懒睡猫	介绍：太	45	58601	1008	25	36
40	rapper说	345006167	决劣	介绍：评	405	14442805	181888	3294	1501
41	「精选」0	129997080	还是挺难过的	介绍：或	44	868281	10490	282	26
42	暮色正浓	450639585	四季里的懒睡猫	介绍：气	45	27256	95	4	0
43	我愿是月	201586	原创君	介绍：第	35	197016	598	7	6
44	古风精选	140774762	小學响	介绍：喜	341	23152970	320652	1911	797
45	「中文rap	341922132	肆意无人	介绍：不	682	8224187	89148	994	523
46	【古风】	346769611	是你的顾柒柒_o	介绍：11	234	2092991	27836	304	169

- 封面图片：

