

1 代码结构

1.1 协程函数定义

```
1 import csv, re
2 from lxml import etree
3 import asyncio, aiohttp, aiofiles
4
5
6
7 async def request(url):
8     conn=aiohttp.TCPConnector(ssl=False)
9     async with aiohttp.ClientSession(connector=conn) as session:
10         response = await session.get(url)
11         return await response.text()
12
13 async def download(url, i):
14     async with aiohttp.ClientSession() as session:
15         async with session.get(url) as response:
16             with open('week14_网易云歌单封面/'+f'{i+1}'.jpg', 'wb') as image:
17                 image.write(await response.content.read())
18
19 async def write_row(row): #line = ','.join(line) + '\n'
20     async with aiofiles.open(r'week14_网易云歌单.csv', 'a', encoding = 'utf-8') as f:
21         await f.write(row)
22
23 async def a_write(response, i):
24     html = etree.HTML(response)
25     title = html.xpath('//h2[@class="f-ff2 f-brk"]')[0].text
26     cover = html.xpath('//img[@class="j-img"]/@src')[0]
27     await download(cover, i)
28     creator_name = html.xpath('//a[@class="s-fc7"]')[0].text
29     creator_id = re.search('(?!<=(:?)id=).*', html.xpath('//a[@class="s-fc7"]/@href')[0]).group()
30     intro_result = html.xpath('//p[@class="intr f-brk"]')
31     if intro_result == []:
32         introduction = ''
33     else:
34         introduction = intro_result[0].xpath('string(.)')[4:] # 删去'介绍\n'
35     num_songs = html.xpath('//span[@id="playlist-track-count"]')[0].text
36     num_play = html.xpath('//strong[@id="play-count"]')[0].text
37     creator_add = html.xpath('//a[@class="u-btni u-btni-fav "] /i/text()')[0][1:-1]
38     num_share = html.xpath('//a[@class="u-btni u-btni-share "] /i/text()')[0][1:-1]
39     num_comment = html.xpath('//span[@id="cnt_comment_count"] /text()')[0]
40     row = [title, cover, creator_name, creator_id, introduction,
41           num_songs, num_play, num_add, num_share, num_comment]
42     # 法1: 逐步添加
43     row = ','.join(row) + '\n'
44     await write_row(row)
45     print(f'>>>--第{i+1}个歌单写入成功-->>>')
```

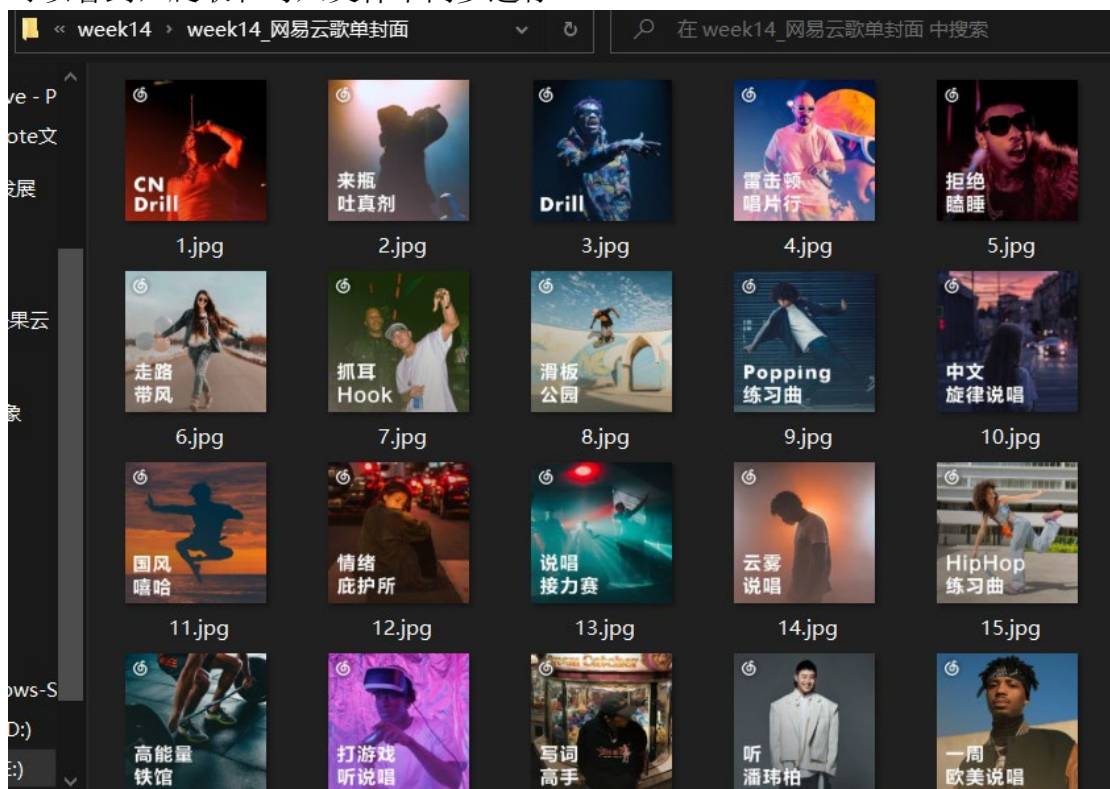
1.2 主函数

```
48 if __name__ == '__main__':
49     with open(r'week14_网易云歌单.csv', 'a', encoding = 'utf-8', newline = '') as f:
50         csv.writer(f).writerow(['title', 'cover', 'creator_name', 'creator_id', 'introduction',
51                                 'num_songs', 'num_play', 'num_add', 'num_share', 'num_comment'])
52     with open('week14_pl_urls.csv', 'r', encoding = 'utf-8', newline = '') as f:
53         urls = [i[0] for i in csv.reader(f)]
54     loop = asyncio.get_event_loop()
55     tasks = []
56     for i in range(len(urls)):
57         print(f'---开始爬取第{i+1}个歌单--->>>')
58         url = urls[i]
59         task_request = loop.create_task(request(url))
60         loop.run_until_complete(task_request)
61         tasks.append(loop.create_task(a_write(task_request.result(), i)))
62     loop.run_until_complete(asyncio.wait(tasks))
```

2 运行结果

```
E:\北航\课业\大三上\现代程序设计技术\week14>python week14_homework.py
--开始爬取第1个歌单-->>>
--开始爬取第2个歌单-->>>
>>>--第1个歌单写入成功--✓
--开始爬取第3个歌单-->>>
>>>--第2个歌单写入成功--✓
--开始爬取第4个歌单-->>>
>>>--第3个歌单写入成功--✓
--开始爬取第5个歌单-->>>
>>>--第4个歌单写入成功--✓
--开始爬取第6个歌单-->>>
>>>--第5个歌单写入成功--✓
--开始爬取第7个歌单-->>>
>>>--第6个歌单写入成功--✓
--开始爬取第8个歌单-->>>
>>>--第7个歌单写入成功--✓
```

可以看到，爬取和写入文件不同步进行。



```
week14_网易云歌单.csv - 记事本
文件(F) 编辑(E) 格式(O) 查看(V) 帮助(H)
title,cover,creator_name,creator_id,introduction,num_songs,num_play,num_add,num_share,num_comment
[CN Drill] 中文钻头节奏，听就完了,http://p4.music.126.net/kmcRWBitc6vvyC-2sE0wvA==/10995116804691733
,67,168960,714,9,9
[来瓶吐真剂] 让歌里的故事诉说心事,http://p3.music.126.net/KuhEhth3Dzwu1yYa--Nx6w==/1099511678985430
,71,390615,1046,14,11
[Drill] 燥起来！接受钻头说唱的低音轰炸吧,http://p4.music.126.net/M9XVRxHkBFYsAZmOchusPA==/1099511679
,60,255010,1797,16,9
[雷击顿唱片行] 让你扭动身体的异域节奏,http://p4.music.126.net/BSs2iYeiFUJclFSiAPP8Bg==/109951165870075
,30,2504023,12457,127,25
[拒绝瞌睡] 跟着节拍抖腿的嘻哈音乐,http://p4.music.126.net/AEjvfbLi56fDGcSXQnkxrg==/109951165529653802
戴上耳机 我就是这条gai最酷的崽
,30,12269693,82238,659,163
[走路带风] 用音乐给今天充满电,http://p3.music.126.net/hFMcFqf2BPCxVer_ehx0FA==/109951167600118285.jp
无聊的一天，如坐针毡，深陷996苦海...
按下播放键，一秒带你找回状态！
,100,1263952,5380,45,18
```
