#### **指导**

以爬取网易云歌单为例，练习多线程的使用。

1. 获取一个分类下的所有歌单的id。观察url可以发现其页码规律：https://music.163.com/#/discover/playlist/?order=hot&cat=%E8%AF%B4%E5%94%B1&limit=35&offset=35。offset是本页开始的数据位置，第1页是0，第2页是35。分类参数是utf-8编码后的汉字"说唱"，使用str.encode可以处理。

2. 对每个id，获取歌单的详细信息，至少包括：歌单的封面图片（需把图片保存到本地）、歌单标题、创建者id、创建者昵称、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。可以自行实现其他信息的获取。基本信息汇总到同一张表中，以csv文件保存。https://music.163.com/playlist?id=3037221581

1. 3. 要求使用生产者-消费者模式实现，要求1作为生产者，每次请求后产生新的任务交给消费者，消费者执行要求2。

 4.（附加）爬虫程序往往需要稳定运行较长的时间，因此如果你的程序突然中 断或异常（比如网络或被封），如何能够快速从断点重启？

 5.（附加）爬虫程序往往需要比较友好的状态输出，因此可否专门有一个线程 动态地进行输出更新，来显示当前的状态，比如程序连续运行的时长，要完成的 总页面数，其中有多少已被爬取，已收集的文件占用了多少空间，大概还需要多 少时间才能完成，预计需要耗费多少硬盘空间等。