020 多进程爬取在线课程并存入MySQL数据库

* 核心技术

使用多进程的方式可以提高爬取数的效率。在Python中,multiprocessing模块可是实现多进程。该模块实现多进程有几种方式，例如Process多进程类和Pool进程池等。本例使用Pool进程池的方式来实现多进程的工能。主要流程如下：

* 创建进程池

首先实例化Pool()类，然后调用pool实例对象的map方法，这里需要传递2个参数，第一个参数main是进程要执行的函数，第二个参数index是当前页码索引。示例代码如下：

1. pool = Pool() # 开启进程
2. index = ([x **for** x **in** range(totlePageCount)]) # 生成当前页码元组
3. pool.map(main, index) # 执行多进程
4. pool.close() # 关闭进程
5. pool.join() # 等待子进程结束

* 创建进程执行函数

pool.map()方法中定义了一个要执行的方法参数main，接下来是需要定义main()函数。示例代码如下：

1. **def** main(index):
2. # 获取JSON格式数据
3. content\_json = get\_json(index)
4. # 获取课程内容
5. content = get\_content(content\_json)
6. # 保存到数据库
7. save\_mysql(content)