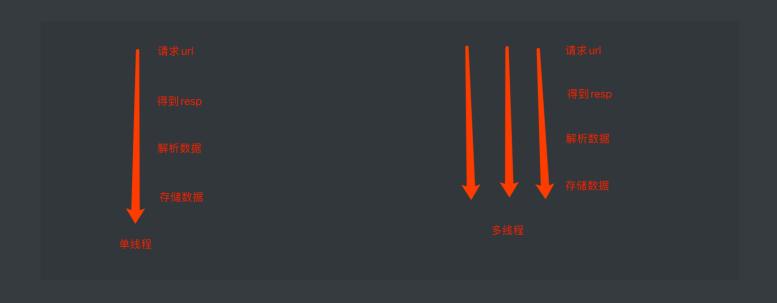
本章内容梗概

到目前为止,我们可以解决爬虫的基本抓取流程了.但是抓取效率还是不够高.如何提高抓取效率呢?我们可以选择多线程,多进程,协程等操作完成异步爬虫.

何为异步?这里我们不讨论蹩脚的概念性问题.直接说效果.

打个比方, 我们目前写的爬虫可以理解为单线程, 比喻为单车道公路. 如何提高效率呢? 很简单, 搞成多车道就OK了啊. 异步爬虫你就可以理解为多车道同时进行爬取.



原来一股道跑,现在三股道跑.哪个快?

在这里要特殊说明一下. 多线程异步爬虫中每一步都可以设立成多线程的. 具体操作得实际去分析. 当然, 也可以像我画图这样, 每一个url一个单独线程. 还是那句话. 技术是死的. 人是活的. 怎么干还得看你自己~

本章涉及内容:

- 1. 快速学会多线程
- 2. 快速学会多进程
- 3. 线程池和进程池
- 4. 扒光新发地
- 5. 协程
- 6. 多任务异步协程实现
- 7. aiohttp模块详解
- 8. 扒光一本小说
- 9. 综合训练-抓取一部电影

