部分网站有反爬虫限制，会根据IP地址来判断是否为爬虫访问，这时就需要使用代理来避开这些限制。

代理池的维护：代理不能保证一定可用，需要剔除不可用的代理，保留可用的代理。

Flask测试：flask\_test/flask\_test.py

如果不使用虚拟环境，在命令行下测试：



代理池分为4个模块：

1. 获取模块：定时在各个代理网站抓取代理，可以是免费的，也可以是付费的。
2. 存储模块：负责存储抓取下来的代理，要保证代理不重复，还要标明代理的可用状况，并能够动态实时处理代理。
3. 检测模块：定时检测数据库中的代理，设置检测链接（最好是爬取那个网站，就将其设置为检测链接），标识每个代理的可用状态。
4. 接口模块：通过Web API获取可用代理。