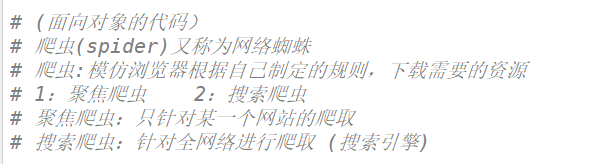
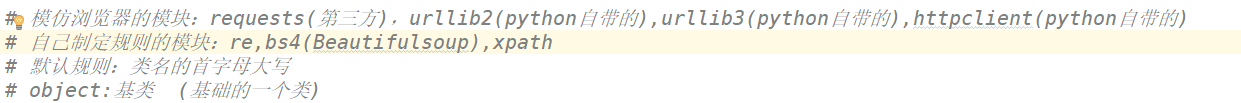
1. **爬虫：**

模仿浏览器根据字的制定的规则，下载需要的资源

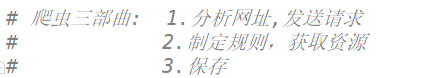
爬虫（spider）又被称为网络蜘蛛

分为：聚焦爬虫和搜索爬虫



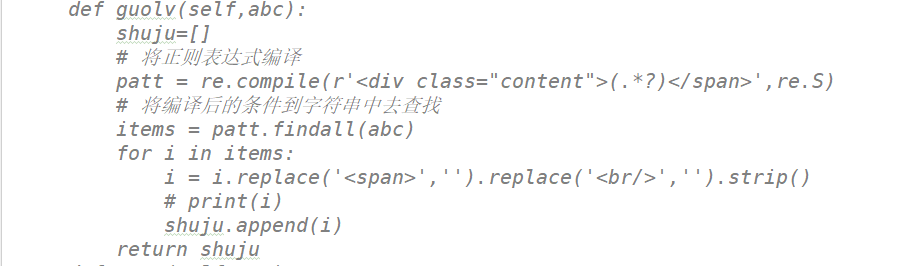


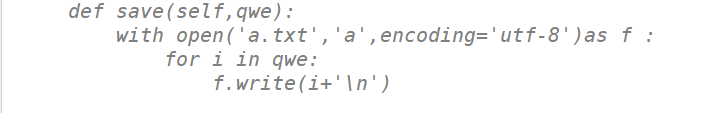
1. **爬虫三部曲：**

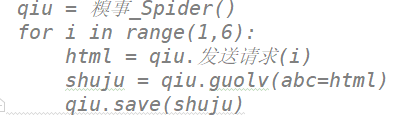


1. **爬取资源**

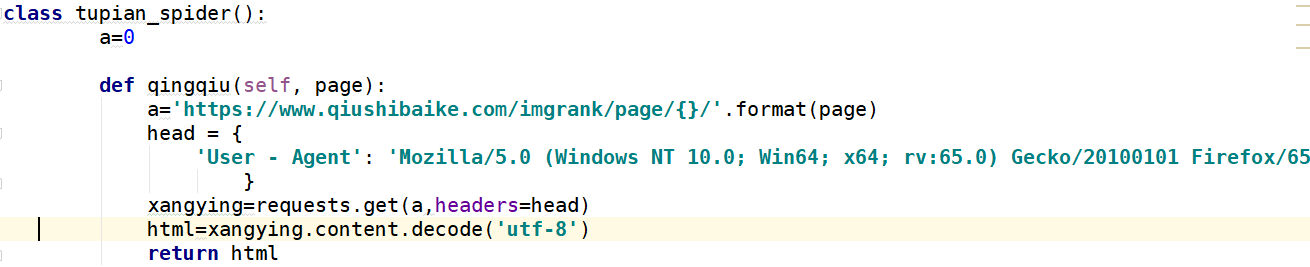


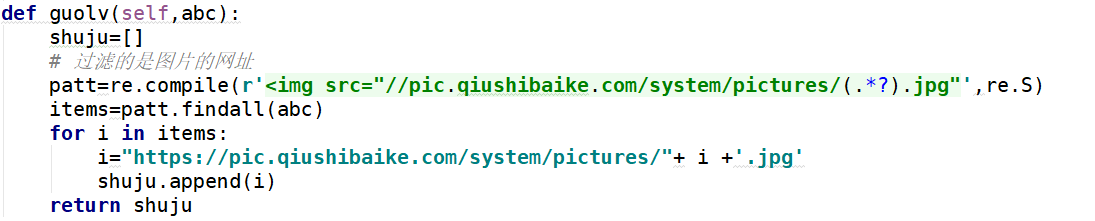


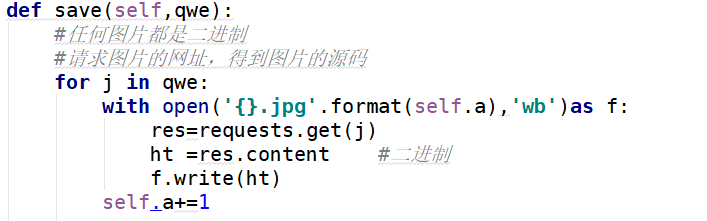


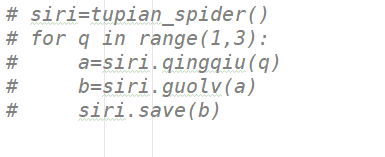


1. **爬取图片**









1. **反爬技术：**

1：通过headers：use-Agent

refere（防盗链）

解决方法，伪装成windows浏览器访问

2：验证码：登录（cookie）

误别

3：拉黑IP地址：检测访问的频率（设置代码IP）

4：数据混淆

5：行为分析

1. **爬取豆瓣网**

爬取影名，导演，多少人评价，介绍，并保存excel表格



