Python爬虫入门 丨如何在豆瓣中获取自己喜欢的TOP N电影信息

<http://www.jianshu.com/p/7eceedb39f3b>

**什么是爬虫**

按照一定规则自动的获取互联网上的信息(如何快速有效的利用互联网上的大量信息

**爬虫的应用**

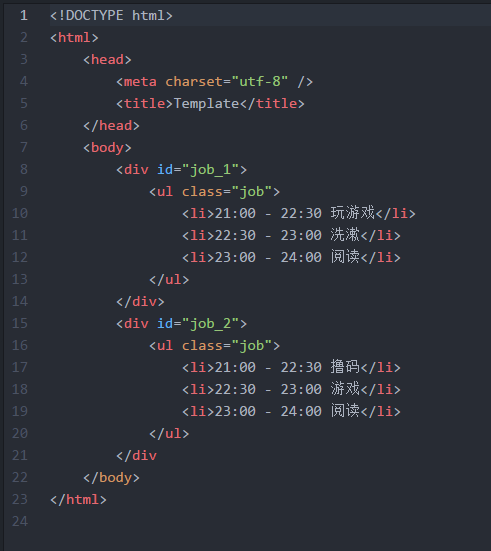
* 搜索引擎（Google、百度、Bing等搜索引擎，辅助人们检索信息）
* 股票软件（爬取股票数据，帮助人们分析决策，进行金融交易）
* Web扫描（需要对网站所有的网页进行漏洞扫描）
* 获取某网站最新文章收藏
* 爬取天气预报
* 爬取漂亮mm照片
* …

**基础知识**

**1.HTTP 协议**

客户端发起请求，服务器接收到请求后返回格式化的数据，客户端接收数据，并进行解析和处理

**2.HTML（超文本标记语言）**



**3.Python**

* 基础语法&常用系统模块
* 第三方模块requests,pyquery使用

安装：

pip install requests

pip install pyquery

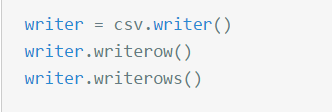
requests模块使用：



pyquery模块使用：

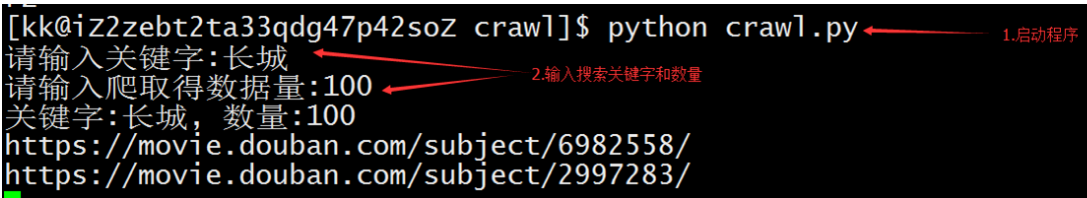


csv模块使用：

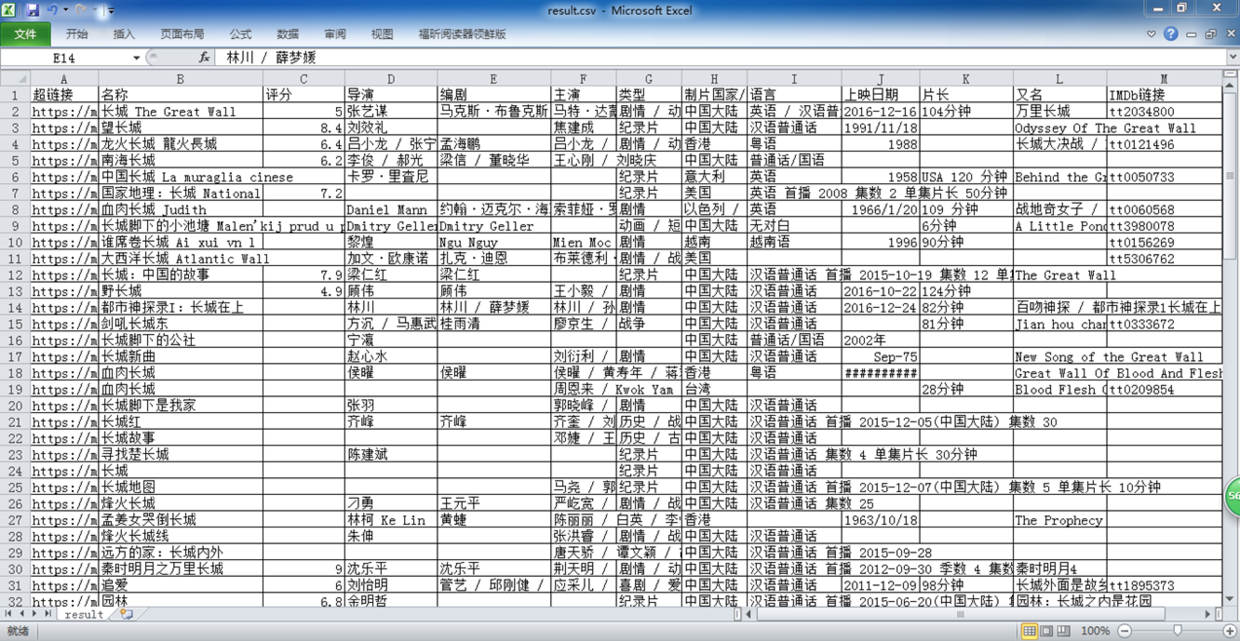


**程序运行**

1.程序启动



2.运行结果



**手动搜索TOP N电影信息**

1.获取电影列表



2.获取电影详情超链接

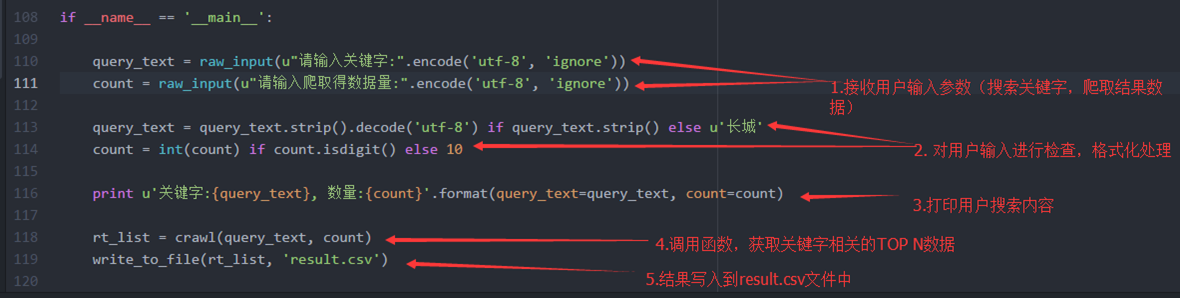


3.获取电影详情

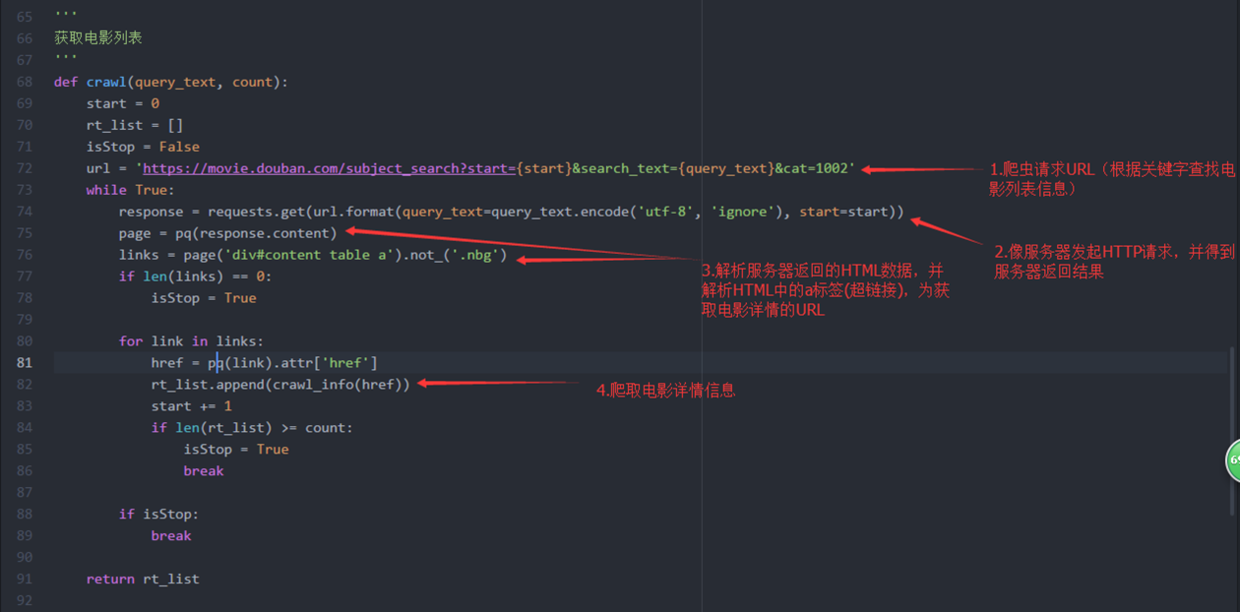


**代码走读**

1.程序启动



2.查找电影列表



3.查找电影详情



4.写入csv文件

