

# 前端小结，简洁易懂

原创

[klspre](#) 于 2021-01-09 18:58:16 发布 904 收藏 2

分类专栏: [前端基础](#) 文章标签: [python](#)

版权声明: 本文为博主原创文章, 遵循 [CC 4.0 BY-SA](#) 版权协议, 转载请附上原文出处链接和本声明。

本文链接: [https://blog.csdn.net/weixin\\_41615044/article/details/112391257](https://blog.csdn.net/weixin_41615044/article/details/112391257)

版权



[前端基础](#) 专栏收录该内容

5 篇文章 0 订阅

订阅专栏

## 文章目录

### HTML

[HTML 与 CSS、JavaScript](#)

[CSS](#)

[CSS语法](#)

[JavaScript](#)

[JS语法](#)

### Python HTTP库 - Requests

[Requests 与 re、beautifulsoup](#)

[Requests](#)

[re 正则](#)

[Beautiful Soup](#)

[csv 保存数据](#)

上面是最基础的, 平时使用还可能用到网还可能用到网络API, 请求时需要加上请求头, 获取cookies, 或者使用webdriver、selenium, 多线程等操作

## HTML

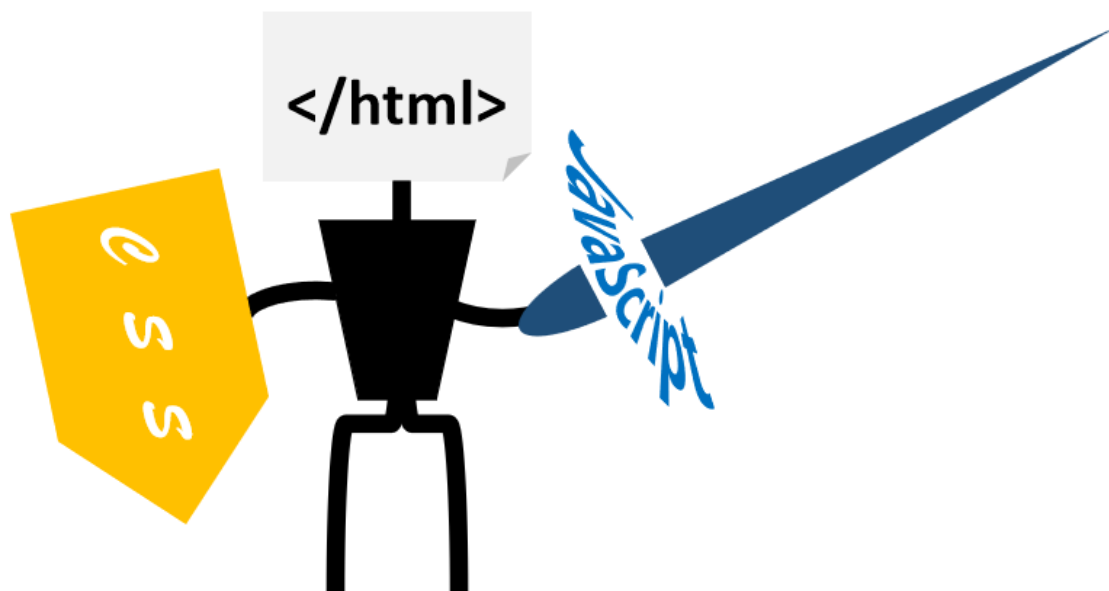
- 超文本标记语言 (Hyper Text Markup Language)
- 一种标记语言
- 不是编程语言
- 使用标记标签来描述网页
- HTML 文档 = 网页
- 标签名来自于英语，很好记

例如：

```
<html> 描述网页的文本 </html>  
<body> 可见的页面内容 </body>  
<h1> 标题 </h1>  
<p> 段落 </p>
```

## HTML 与 CSS、JavaScript

- HTML 定义网页的内容
- CSS 规定网页的布局
- JavaScript 对网页行为进行编程



[https://blog.csdn.net/weixin\\_41615044](https://blog.csdn.net/weixin_41615044)

## CSS

CSS 指层叠样式表 (Cascading Style Sheets)

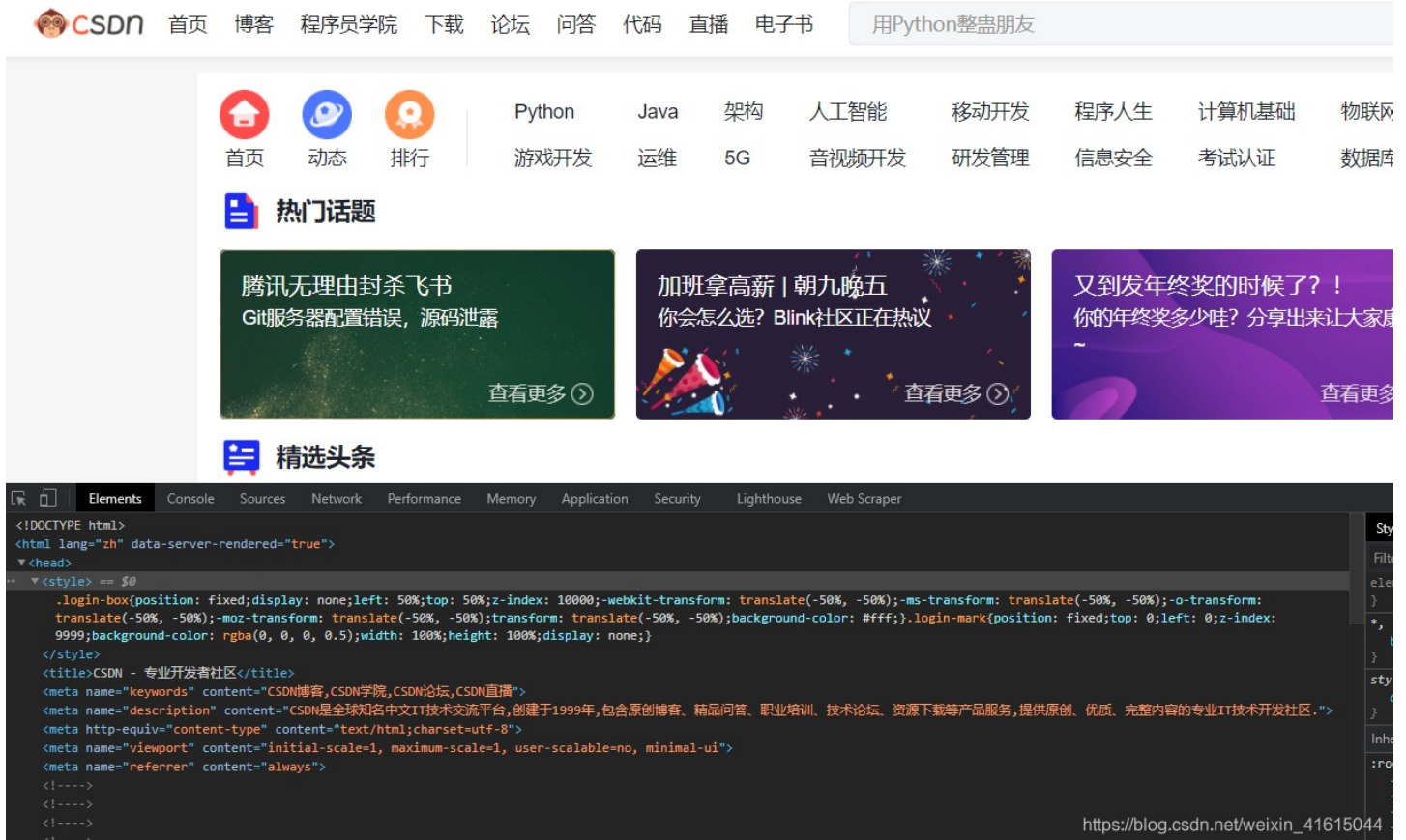
定义如何显示 HTML 元素

HTML 原本只是定义文档内容

CSS 来定义布局和外观

网页的 css一般写在< head> < /head >里面

在页面按 F12或“右键”→“检查”打开开发者界面



在标签 < style> < /style> 里面的就是css层叠样式表

- {}左边的就是选择渲染的页面中的对象
- 选择的方法很多，基本的有根据标签，根据类选择，根据 name和 id选择等等，查看CSS选择器
- {}里面的键值对，在这里，蓝色的部分是属性，白色的部分是值
- 所以页面才会好看

查看更多

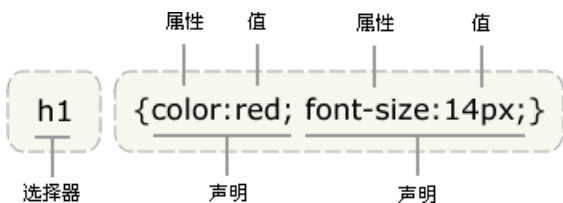
精选头条

```
Elements Console Sources Network Performance Memory Application S...
<!DOCTYPE html>
<html lang="zh" data-server-rendered="true">
  <head>
    <style> == $0
      .login-box{position: fixed;display: none;left: 50%;top: 50%;z-index: 10000;-webkit-t
      translate(-50%, -50%);-moz-transform: translate(-50%, -50%);transform: translate(-50
      9999;background-color: rgba(0, 0, 0, 0.5);width: 100%;height: 100%;display: none;}
    </style>
    <title>CSDN - 专业开发者社区</title>
    <meta name="keywords" content="CSDN博客,CSDN学院,CSDN论坛,CSDN直播">
    <meta name="description" content="CSDN是全球知名中文IT技术交流平台,创建于1999年,包含原创博
    <meta http-equiv="content-type" content="text/html;charset=utf-8">
    <meta name="viewport" content="initial-scale=1, maximum-scale=1, user-scalable=no, mi
    <meta name="referrer" content="always">
    <!-->
    <!-->
    <!-->
    https://blog.csdn.net/weixin_41615044
```

## CSS语法

CSS 规则由两个主要的部分构成：选择器，以及一条或多条声明

- 花括号 {} 里面的是代码块，语句用 分号 隔开
- 如果一个值的中间有空格隔开，那么需要 引号 把这个值括起来



## JavaScript

- JavaScript 是属于 HTML 和 Web 的编程语言
- 对网页行为进行编程
- JavaScript 和 CSS 一样是 HTML 的左膀右臂
- JavaScript 和 Java 是取名蹭热度的关系
- JavaScript 和 Python 一样是动态语言，脚本，语法也非常像
- 通过变量和函数操控、改变页面类容，增加页面的交互功能
- 在 HTML 中，JavaScript 程序由 web 浏览器执行
- 经常写在 <script></script> 标签中

```
</div>
  <div>...</div>
</div>
</div>
</div>
<script> == $0
window._INITIAL_STATE = {"csrf":"9vuhPhU
s1iyup_UAToken=117#95Nm20919RUJE+umd0A9ncy
gc_sid=b943bf1bb1d234c72af570feb1f13d40; c
/m_ct_6bcd52f51e9b3dce32bec4a3997715ac=6523
c_pref=https%3A%u002F%u002Fblog.csdn.net%u0
[{"FloorId":62,"floorComponentName":"floor-
headLines": "componentConfieData":{"report
```

## JS语法

JS 教程 & 案例

# Python HTTP库 - Requests

## Requests 与 re、beautifulsoup

### Requests

官方警告:

1. 非专业使用其他 HTTP 库会导致危险的副作用，包括：安全缺陷症、冗余代码症、重新发明轮子症、啃文档症、抑郁、头疼、甚至死亡
2. Requests 唯一的一个非转基因的 Python HTTP 库，人类可以安全享用

- 安装

```
在 cmd 中输入:
pip install requests
```

发送请求，拿到页面响应

如果使用循环请求多个页面，最好在每次请求时等待一段时间，导入 random和 time，等待一段随机时间，这里先不举例

```
# python
import requests

# 发送请求，把回应放入变量 response 中
response = requests.get('https://www.csdn.net/')
print(response)
print(response.status_code) # 状态码 200表示成功
print(type(response))
```

```
<Response [200]>
200
<class 'requests.models.Response'>
```

状态码	说明	详情
100	继续	请求者应当继续提出请求 服务器已收到请求的一部分，正在等待其余部分
101	切换协议	请求者已要求服务器切换协议，服务器已确认并准备切换
200	成功	服务器已成功处理了请求
201	已创建	请求成功并且服务器创建了新的资源
202	已接受	服务器已接受请求，但尚未处理
203	非授权信息	服务器已成功处理了请求，但返回的信息可能来自另一个源
204	无内容	服务器成功处理了请求，但没有返回任何内容
205	重置内容	服务器成功处理了请求，内容被重置
206	部分内容	服务器成功处理了部分请求
300	多种选择	针对请求，服务器可执行多种操作
301	永久移动	请求的网页已永久移动到新位置，即永久重定向
302	临时移动	请求的网页暂时跳转到其他页面，即暂时重定向
303	查看其他位置	如果原来的请求是 POST，重定向目标文档应该通过 GET 提取
304	未修改	此次请求返回的网页未修改，继续使用上次的资源
305	使用代理	请求者应该使用代理访问该网页
307	临时重定向	请求的资源临时从其他位置响应
400	错误请求	服务器无法解析该请求
401	未授权	请求没有进行身份验证或验证未通过
403	禁止访问	服务器拒绝此请求
404	未找到	服务器找不到请求的网页
405	方法禁用	服务器禁用了请求中指定的方法
406	不接受	无法使用请求的内容响应请求的网页
407	需要代理授权	请求者需要使用代理授权
408	请求超时	服务器请求超时
409	冲突	服务器在完成请求时发生冲突
410	已删除	请求的资源已永久删除
411	需要有效长度	服务器不接受不含有效内容长度标头字段的请求
412	未满足前提条件	服务器未满足请求者在请求中设置的其中一个前提条件
413	请求实体过大	请求实体过大，超出服务器的处理能力
414	请求 URL 过长	请求 URL 过长，服务器无法处理

414	请求URI过长	请求网址过长，服务器无法处理
415	不支持类型	请求格式不被请求页面支持
416	请求范围不符	页面无法提供请求的范围
417	未满足期望值	服务器未满足期望请求标头字段的要求
500	服务器内部错误	服务器遇到错误，无法完成请求
501	未实现	服务器不具备完成请求的功能
502	错误网关	服务器作为网关或代理，从上游服务器收到无效响应
503	服务不可用	服务器目前无法使用
504	网关超时	服务器作为网关或代理，但是没有及时从上游服务器收到请求
505	HTTP 版本不支持	服务器不支持请求中所用的 HTTP 协议版本

[https://blog.csdn.net/weixin\\_41615044](https://blog.csdn.net/weixin_41615044)

- 请求成功后，返回了 Response 对象，有以下属性

```
print(dir(requests.Response))
```

```
['_attrs__', '__bool__', '__class__', '__delattr__', '__dict__', '__dir__', '__doc__', '__enter__', '__eq__', '__exit__', '__format__', '__ge__', '__getattr__', '__getstate__', '__gt__', '__hash__', '__init__', '__init_subclass__', '__iter__', '__le__', '__lt__', '__module__', '__ne__', '__new__', '__nonzero__', '__reduce__', '__reduce_ex__', '__repr__', '__setattr__', '__setstate__', '__sizeof__', '__str__', '__subclasshook__', '__weakref__',  
# 主要看下面的  
'apparent_encoding', 'close', 'content', 'is_permanent_redirect', 'is_redirect', 'iter_content', 'iter_lines', 'json', 'links', 'next', 'ok', 'raise_for_status', 'text']
```

- 输出 Response 对象的文本属性

```
print(response.text)
```

- 就是网页的 html 文档，网页的内容基本都在 <> </> 标签里面

```
<!doctype html><html lang="zh" data-server-rendered="true"><head><title>CSDN - 专业开发者社区</title> <meta name="keywords" content="CSDN博客,CSDN学院,CSDN论坛,CSDN直播"> <meta name="description" content="CSDN是全球知名中文IT技术交流平台,创建于1999年,包含原创博客、精品问答、职业培训、技术论坛、资源下载等产品服务,提供原创、优质、完整内容的专业IT技术开发社区."> <meta http-equiv="content-type" content="text/html;charset=utf-8"> <meta name="viewport" content="initial-scale=1, maximum-scale=1, user-scalable=no, minimal-ui"> <meta name="referrer" content="always"> <!--> <!--> <!--> <!--> <!--> <!--> <link rel="shortcut icon" href="https://g.csdnimg.cn/static/logo/favicon32.ico" type="image/x-icon"> <!-->

    <meta name="toolbar" content={} />

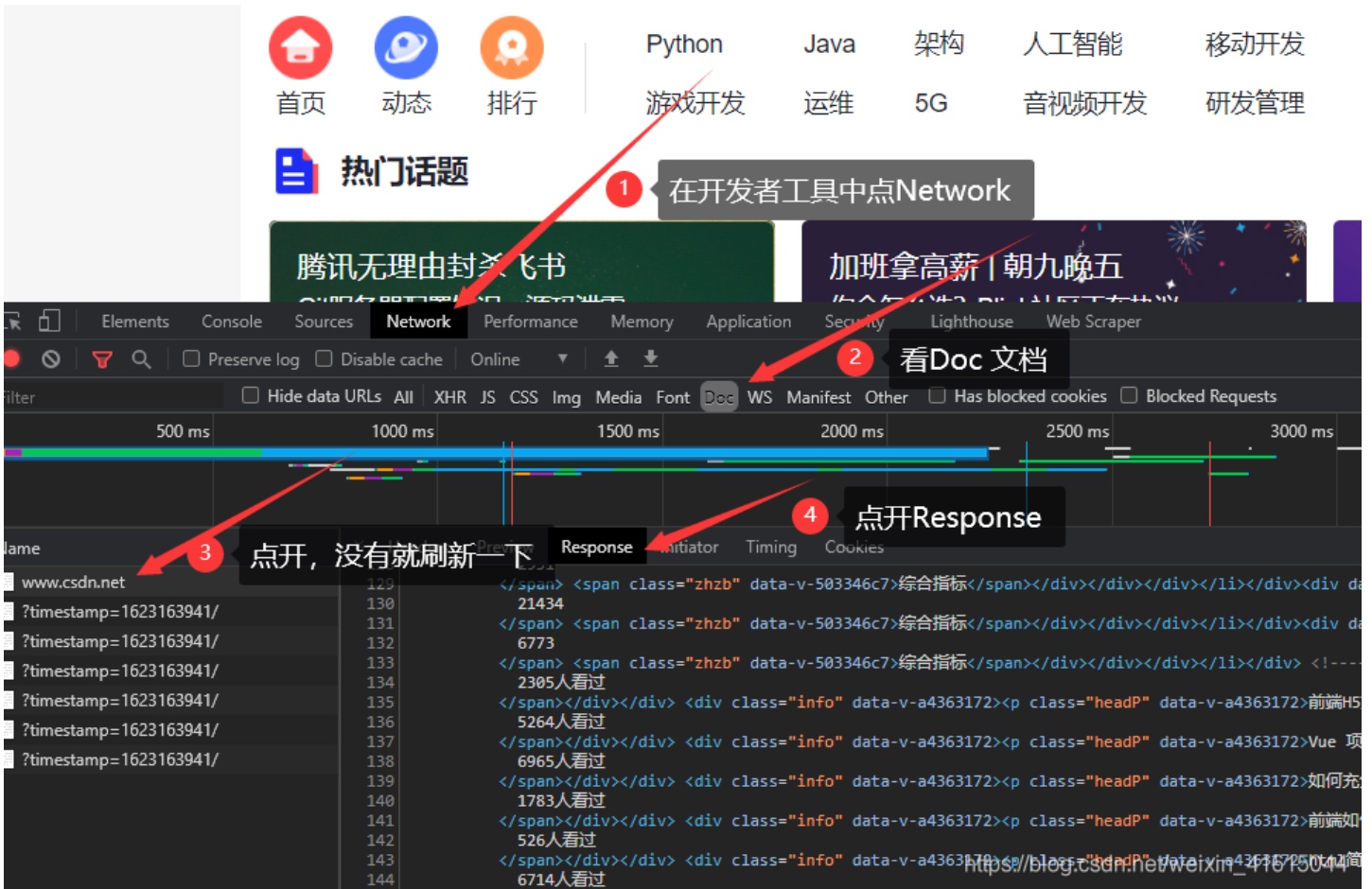
    <meta name="report" content={"spm":"1000.2115"} />
    <script src="https://g.csdnimg.cn/?lib/jquery/1.12.4/jquery.min.js,user-tooltip/2.2/user-tooltip.js,lib/qrcode/1.0.0/qrcode.min.js"></script> <script src='//g.csdnimg.cn/common/csdn-report/report.js' type='text/javascript'></script>
    <script src="https://g.csdnimg.cn/user-ordercart/2.0.0/user-ordercart.js"></script>

    <script src="https://g.csdnimg.cn/login-box/1.1.4/login-box.js"></script>

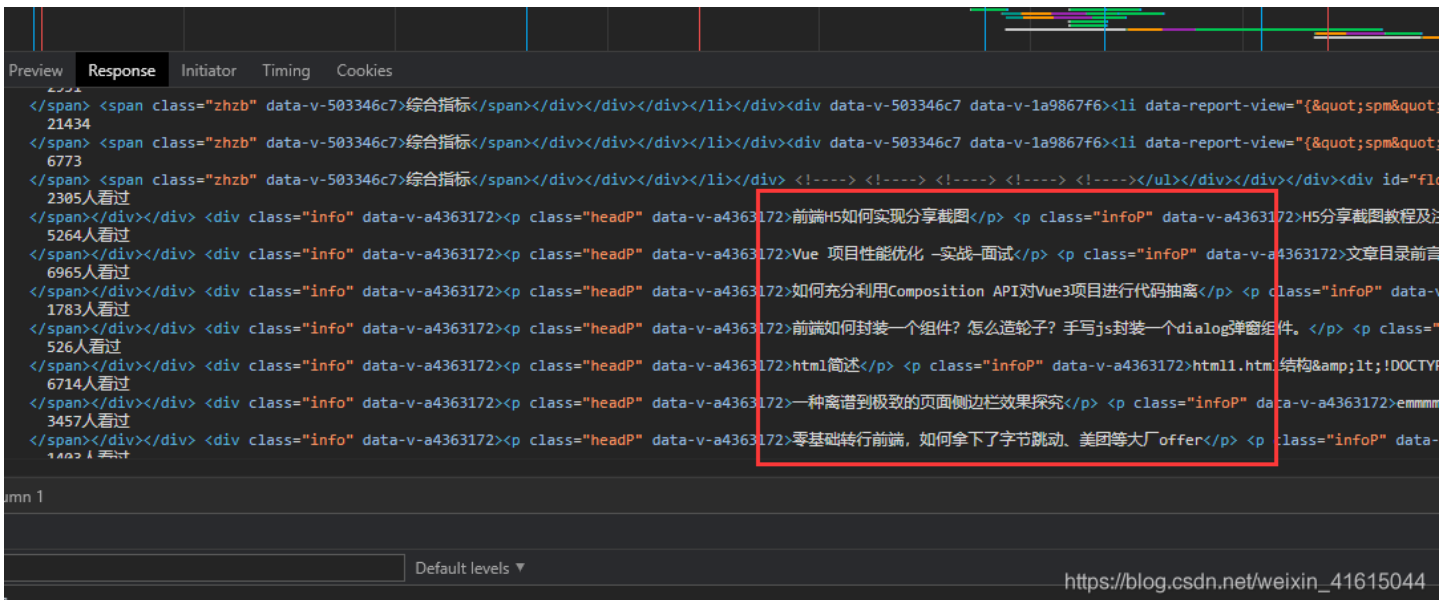
    <script src="https://g.csdnimg.cn/user-ordertip/3.0.1/user-ordertip.js"></script>
<!-->
.....
.....
```

- 网页内容很多很乱
- 可以使用 (re) 正则表达式把想要的内容拿出来
- 也可以 beautiful soup 提取目标内容
- 先来看看开发者工具





- 假设我们需要下方框里的文本内容



## re 正则

```
import re
```

把想要的部分加上周围恒定的部分一起复制下来

```
<p class="headP" data-v-a4363172>前端H5如何实现分享截图</p>
```

```
<p class="headP" data-v-a4363172>(前端H5如何实现分享截图)</p>
```

圆括号里的就是我要的内容

把圆括号里的内容改为 `.*`

`(.*)` 都是英文输入法状态

```
<p class="headP" data-v-a4363172>(.*?)</p>
```

为什么这样写？

[python基础day-16: 正则表达式](#)

上面步骤完成后，接着把这段修改的内容放到以下代码中

```
# 创建一个 Pattern对象，它的作用就像个筛子，过滤文本
pattern = re.compile('<p class="headP" data-v-a4363172>(.*?)</p>')

# 让它来过滤我们之前拿到的网页HTML文档
# 结果保存到变量 result中
result = pattern.findall(response.text)
print(result)
```

看样子就是得到了网页上目标内容的字符串列表

上面举例的“前端H5如何实现分享截图”也肯定在其中

用 python 的列表和字符串方法就可以操作这些数据了

['离散数学总复习精华版(最全 最简单易懂)已完结', 'JavaScript 抢购茅台脚本 仅供学习', 'C/C++指针详解之基础篇(史上最全最易懂指针学习指南!!!)', '高校校园网建设方案【含网络拓扑图+拓扑结构图+配置命令】(详细版)', '如何通过QQ号获取绑定的手机号', '基于LSTM的多变量多步序列预测模型实战「超详细实现说明讲解」', '2020年最火的百场技术直播盘点新鲜出炉!', 'Star国产AI框架MegEngine', '2020中国开发者有奖大调查', 'Unity官方游戏开发课程,第三期招生火热开启!', '华为开发者专区', '英特尔开发人员专区', 'Bitcoin SV开发者专区', '智联万物, 京东IoT技术创新与实践', '华为开发者学院', 'Qualcomm 开发者专区', '微软技术加油栈, 为程序员添“专”加“码”', '旷视天元学习社区', '英特尔® OpenVINO™工具套件初级', 'Intel 工业Edge Insights & IOT系列课程', '对进行股票可视化分析', 'Python项目代码使用过程中遇到的Python问题汇总索引目录【淘宝-天猫超市、京东】', 'Python爬取小姐姐图片', '豆瓣高分电影信息分析(数据分析)', '学点简单的Python之Python生成器', 'python爬虫及数据可视化分析', '天下没有难学的知识 --- 从零讲解DCGAN生成动漫头像', '疫情数据分析与可视化', '基于python实现ZigBee树状拓扑网络建构与动态地址分配及路由仿真', 'Python爬虫之小说信息爬取与数据可视化分析', 'python 列表过滤方法 条件表达式', '用Python写一个天天酷跑', 'js逆向解析, js爬虫', 'Python爬取B站冰冰评论 看看大家说了什么', '前端H5如何实现分享截图', 'Vue 项目性能优化 -实战-面试', '如何充分利用Composition API对Vue3项目进行代码抽离', '前端如何封装一个组件? 怎么造轮子? 手写js封装一个dialog弹窗组件.', 'html简述', '一种离谱到极致的页面侧边栏效果探究', '零基础转行前端, 如何拿下了字节跳动、美团等大厂offer', '八个方面理解seo(搜索引擎)', '送你一朵小红花, CSS实现一朵旋转的小红花', '我女儿说要看雪, 但是我家在南方, 于是我默默的拿起了键盘, 下雪咯.', 'angularjs 搭建项目、使用uirouter配置路由, 封装\$http, 超时退出', '微信小程序抽奖组件', 'React Server Components 介绍 亮点', '关于input框 限制输入长度记录一次', 'DTU工业4G路由器模块实现串口转4G/WiFi以太网的TCP/IP和UDP协议分别是什么', '网络安全行业全领域白皮书', '边缘计算 - 与 5G', 'ESP32-S3和ESP32-C3来了', '关于5G你需要知道', '[5G专题-9]: RRU 数字上变频DUC与数字下变频DDC', '微型嵌入式物联网网关如何应用在无人快递小车', '2021-01-07', '小小模组结束传统汽车百年“移动出行工具”属性?', '未来光通信迈入多通道集成时代, 泰克助力上海交大搭建下一代光通信研发平台', '监控摄像头接入流媒体服务器的几种方式', '减少网络卡顿, 华为HMS Core无线传输服务赋能连接与通信领域', '制造→「智」造, 5G工业网关加速智慧工厂建设', '【werbtc走读】PeerConnectionChannel 1', '全球网速变慢, 运营商: WiFi 和微波炉不要放在一起!', '从实习生到算法专家, 他只用了2年!', '利用python开发app实战', 'llvm 编译器高级用法: 第三方库插桩', '卢华东: 从运动员到程序员, 因兴趣与RT-Thread结缘', '检索20万颗星星只需10秒! 这款AI产品能“擎天”', '吴恩达等AI大佬发出 2021新年寄语', '百度AI的2020', '前端组件化基础知识', '12 大热门事件背后, 藏着你的 2020 年 | 年终盘点', '盘点当下大热的 7 大 Github 机器学习【创新】项目', '云计算正在“抹杀”开源?', '内存分页不就够了? 为什么还要分段? 还有段页式?', '当会打王者荣耀的AI学会踢足球, 一不小心拿下世界冠军!']

## Beautiful Soup

先导入 bs4, 没有就先 install

```
import bs4
```

获取页面

```
import bs4

# 发送请求, 把回应放入变量response中
response = requests.get('https://www.csdn.net/')
print(response.status_code) # 状态码
```

这次页面内容不用正则来筛选, 而是倒进“汤”里

方法里带一个 HTML 解析器参数

```
import bs4

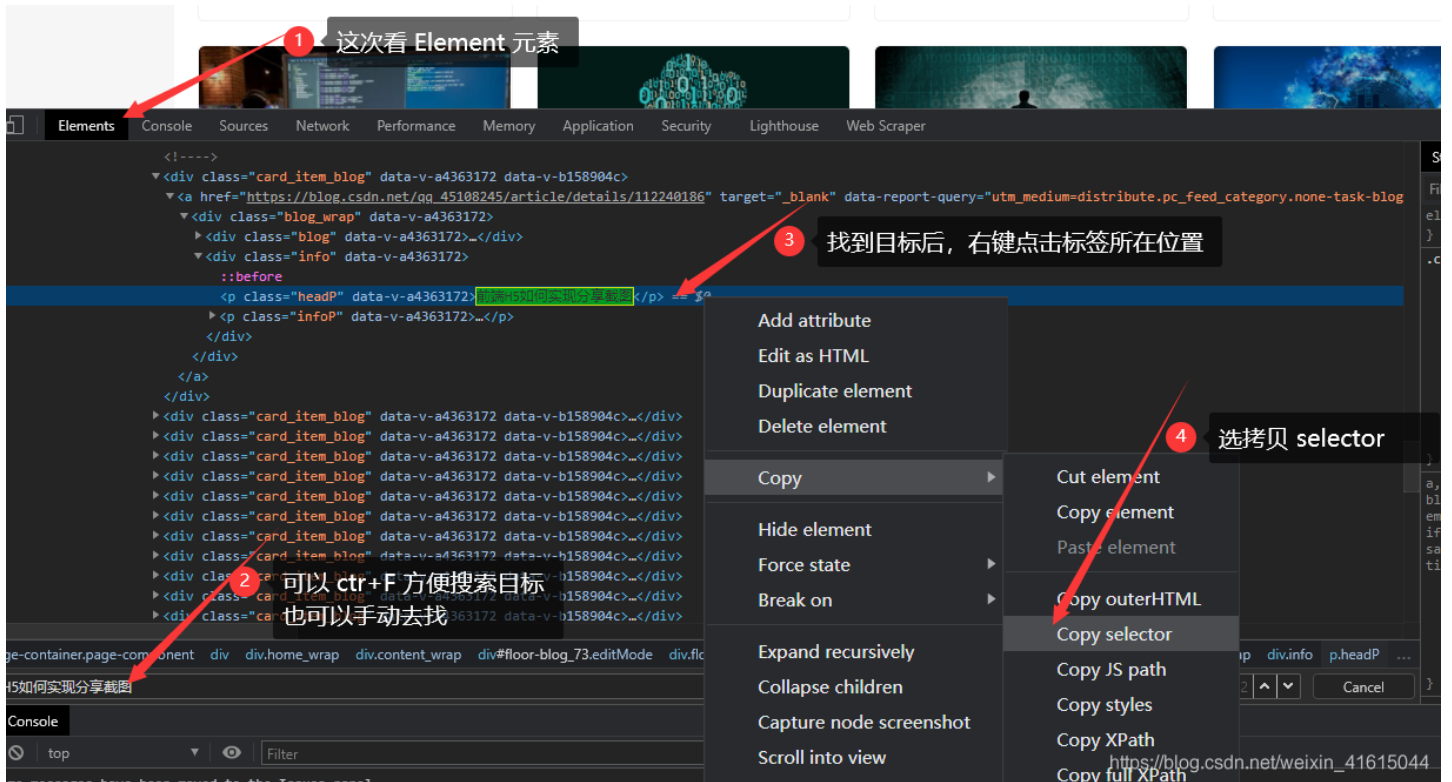
# 发送请求, 把回应放入变量response中
response = requests.get('https://www.csdn.net/')
print(response.status_code) # 状态码

soup = bs4.BeautifulSoup(response.text, 'html.parser')
```

内容倒进汤里后，调用汤的方法来挑选想要的内容

用 HTML 的选择器来挑选

我们再看看网页的开发者工具



复制到了目标所在的精确位置

```
#floor-blog_73 > div > div.card_list.rec_wrap > div.card_list_wrap > div:nth-child(1) > a > div > div.info > p.headP
```

但不需要那么多，我只取后边一小段，交给汤自己去找

```
div.info > p.headP
```

```
import bs4

# 发送请求，把回应放入变量response中
response = requests.get('https://www.csdn.net/')
print(response.status_code) # 状态码

soup = bs4.BeautifulSoup(response.text, 'html.parser')
# 汤.select 方法
targets = soup.select('div.info > p.headP')
print(targets)
```

[<p class="headP" data-v-b98d3800="">Java8编程实战</p>, <p class="headP" data-v-b98d3800="">面试之排序算法</p>, <p class="headP" data-v-b98d3800="">实用数据分析: 数据分析师从小白到精通</p>, <p class="headP" data-v-b98d3800="">Java之路</p>, <p class="headP" data-v-b98d3800="">韦东山嵌入式Linux第一期视频</p>, <p class="headP" data-v-b98d3800="">编程可以这样学</p>, <p class="headP" data-v-a4363172="">离散数学总复习精华版(最全 最简单易懂)已完结</p>, <p class="headP" data-v-a4363172="">JavaScript 抢购茅台脚本 仅供学习</p>, <p class="headP" data-v-a4363172="">C/C++指针详解之基础篇(史上最全最易懂指针学习指南!!!)</p>, <p class="headP" data-v-a4363172="">高校校园网建设方案【含网络拓扑图+拓扑结构图+配置命令】(详细版)</p>, <p class="headP" data-v-a4363172="">如何通过QQ号获取绑定的手机号</p>, <p class="headP" data-v-a4363172="">基于LSTM的多变量多步序列预测模型实战「超详细实现说明讲解」</p>, <p class="headP" data-v-a4363172="">2020年最火的百场技术直播盘点新鲜出炉!</p>, <p class="headP" data-v-a4363172="">Star国产AI框架MegEngine</p>, <p class="headP" data-v-a4363172="">2020中国开发者有奖大调查</p>, <p class="headP" data-v-a4363172="">Unity官方游戏开发课程, 第三期招生火热开启!</p>, <p class="headP" data-v-a4363172="">华为开发者专区</p>, <p class="headP" data-v-a4363172="">英特尔开发人员专区</p>, <p class="headP" data-v-a4363172="">Bitcoin SV开发者专区</p>, <p class="headP" data-v-a4363172="">智联万物, 京东IoT技术创新与实践</p>, <p class="headP" data-v-a4363172="">华为开发者学院</p>, <p class="headP" data-v-a4363172="">Qualcomm 开发者专区</p>, <p class="headP" data-v-a4363172="">微软技术加油栈, 为程序员添“专”加“码”</p>, <p class="headP" data-v-a4363172="">旷视天元学习社区</p>, <p class="headP" data-v-a4363172="">英特尔® OpenVINO™工具套件初级</p>, <p class="headP" data-v-a4363172="">Intel 工业Edge Insights & IOT系列课程</p>, <p class="headP" data-v-a4363172="">对进行股票可视化分析</p>, <p class="headP" data-v-a4363172="">Python项目代码使用过程中遇到的Python问题汇总索引目录【淘宝-天猫超市、京东】</p>, <p class="headP" data-v-a4363172="">Python爬取小姐姐图片</p>, <p class="headP" data-v-a4363172="">豆瓣高分电影信息分析(数据分析)</p>, <p class="headP" data-v-a4363172="">学点简单的Python之Python生成器</p>, <p class="headP" data-v-a4363172="">python爬虫及数据可视化分析</p>, <p class="headP" data-v-a4363172="">天下没有难学的知识 --- 从零讲解DCGAN生成动漫头像</p>, <p class="headP" data-v-a4363172="">疫情数据分析与可视化</p>, <p class="headP" data-v-a4363172="">基于python实现ZigBee树状拓扑网络建构与动态地址分配及路由仿真</p>, <p class="headP" data-v-a4363172="">Python爬虫之小说信息爬取与数据可视化分析</p>, <p class="headP" data-v-a4363172="">python 列表过滤方法 条件表达式</p>, <p class="headP" data-v-a4363172="">用Python写一个天天酷跑</p>, <p class="headP" data-v-a4363172="">js逆向解析, js爬虫</p>, <p class="headP" data-v-a4363172="">Python爬取B站冰冰评论 看看大家说了什么</p>, <p class="headP" data-v-a4363172="">前端H5如何实现分享截图</p>, <p class="headP" data-v-a4363172="">Vue 项目性能优化 -实战-面试</p>, <p class="headP" data-v-a4363172="">如何充分利用Composition API对Vue3项目进行代码抽离</p>, <p class="headP" data-v-a4363172="">前端如何封装一个组件? 怎么造轮子? 手写js封装一个dialog弹窗组件.</p>, <p class="headP" data-v-a4363172="">html简述</p>, <p class="headP" data-v-a4363172="">一种离谱到极致的页面侧边栏效果探究</p>, <p class="headP" data-v-a4363172="">零基础转行前端, 如何拿下了字节跳动、美团等大厂offer</p>, <p class="headP" data-v-a4363172="">八个方面理解seo(搜索引擎)</p>, <p class="headP" data-v-a4363172="">送你一朵小红花, CSS实现一朵旋转的小红花</p>, <p class="headP" data-v-a4363172="">我女儿说要看雪, 但是我家在南方, 于是我默默的拿起了键盘, 下雪咯.</p>, <p class="headP" data-v-a4363172="">angularjs 搭建项目、使用uirouter配置路由,封装\$http,超时退出</p>, <p class="headP" data-v-a4363172="">微信小程序抽奖组件</p>, <p class="headP" data-v-a4363172="">React Server Components 介绍 亮点</p>, <p class="headP" data-v-a4363172="">关于input框 限制输入长度记录一次</p>, <p class="headP" data-v-a4363172="">Android购物车效果实现(RecyclerView悬浮头部实现)</p>, <p class="headP" data-v-a4363172="">安卓11 init初始化以及init.rc的解析执行过程详解</p>, <p class="headP" data-v-a4363172="">微信小程序抽奖组件</p>, <p class="headP" data-v-a4363172="">vue搭建项目: 配置路由; 封装axios; 引入mockjs; 设置bus跨层级通信\*\*</p>, <p class="headP" data-v-a4363172="">安卓基础学习 Day01 | 第一个安卓应用程序: Hello Word!</p>, <p class="headP" data-v-a4363172="">安卓开发学习—day2</p>, <p class="headP" data-v-a4363172="">iOS自定义相机(带拍摄区域边框、半透明遮罩层、点击屏幕对焦、自动裁剪): 1、身份证正反面相机(加一个长方形的框框并裁剪身份证照片) 2、手持证件照相机(含demo源码)</p>, <p class="headP" data-v-a4363172="">安卓学习日志 - Day02</p>, <p class="headP" data-v-a4363172="">uni-app实现实时消息SDK插件</p>, <p class="headP" data-v-a4363172="">Android10以上通过onActivityResult获取图片</p>, <p class="headP" data-v-a4363172="">Android学习笔记</p>, <p class="headP" data-v-a4363172="">Android高版本P/Q/R源码编译指南</p>, <p class="headP" data-v-a4363172="">了解android studio</p>, <p class="headP" data-v-a4363172="">Android7.0修改时间服务器</p>, <p class="headP" data-v-a4363172="">全球网速变慢, 运营商: WiFi 和微波炉不要放在一起!</p>, <p class="headP" data-v-a4363172="">从实习生到算法专家, 他只用了2年!</p>, <p class="headP" data-v-a4363172="">利用python开发app实战</p>, <p class="headP" data-v-a4363172="">llvm 编译器高级用法: 第三方库插桩</p>, <p class="headP" data-v-a4363172="">卢华东: 从运动员到程序员, 因兴趣与RT-Thread结缘</p>, <p class="headP" data-v-a4363172="">检索20万颗星星只需10秒! 这款AI产品能“擎天”</p>, <p class="headP" data-v-a4363172="">吴恩达等AI大佬发出 2021新年寄语</p>, <p class="headP" data-v-a4363172="">百度AI的2020</p>, <p class="headP" data-v-a4363172="">前端组件化基础知识</p>, <p class="headP" data-v-a4363172="">12 大热门事件背后, 藏着你的 2020 年 | 年终盘点</p>, <p class="headP" data-v-a4363172="">盘点当下大热的 7 大 Github 机器学习【创新】项目</p>, <p class="headP" data-v-a4363172="">云计算正在“抹杀”开源?</p>, <p class="headP" data-v-a4363172="">内存分页不就够了? 为什么还要分段? 还有段页式?</p>, <p class="headP" data-v-a4363172="">当会打王者荣耀的AI学会踢足球, 一不小心拿下世界冠军!</p>]

可见拿到的结果不是最后的文字信息, 而是目标所在的标签

根据标签所具有的属性，就可以拿到我们想要的东西了

```
# 查看对象属性
print(dir(targets[0]))
```

```
['_bool__', '__call__', '__class__', '__contains__', '__copy__', '__delattr__', '__delitem__', '__dict__', '__dir__', '__doc__', '__eq__', '__format__', '__ge__', '__getattr__', '__getattribute__', '__getitem__', '__gt__', '__hash__', '__init__', '__init_subclass__', '__iter__', '__le__', '__len__', '__lt__', '__module__', '__ne__', '__new__', '__reduce__', '__reduce_ex__', '__repr__', '__setattr__', '__setitem__', '__sizeof__', '__str__', '__subclasshook__', '__unicode__', '__weakref__', '_all_strings', '_find_all', '_find_one', '_is_xml', '_lastRecursiveChild', '_last_descendant', '_should_pretty_print', 'append', 'attrs', 'can_be_empty_element', 'cdata_list_attributes', 'childGenerator', 'children', 'clear', 'contents', 'decode', 'decode_contents', 'decompose', 'decomposed', 'descendants', 'encode', 'encode_contents', 'extend', 'extract', 'fetchNextSiblings', 'fetchParents', 'fetchPrevious', 'fetchPreviousSiblings', 'find', 'findAll', 'findAllNext', 'findAllPrevious', 'findChild', 'findChildren', 'findNext', 'findNextSibling', 'findNextSiblings', 'findParent', 'findParents', 'findPrevious', 'findPreviousSibling', 'findPreviousSiblings', 'find_all', 'find_all_next', 'find_all_previous', 'find_next', 'find_next_sibling', 'find_next_siblings', 'find_parent', 'find_parents', 'find_previous', 'find_previous_sibling', 'find_previous_siblings', 'format_string', 'formatter_for_name', 'get', 'getText', 'get_attribute_list', 'get_text', 'has_attr', 'has_key', 'hidden', 'index', 'insert', 'insert_after', 'insert_before', 'isSelfClosing', 'is_empty_element', 'known_xml', 'name', 'namespace', 'next', 'nextGenerator', 'nextSibling', 'nextSiblingGenerator', 'next_element', 'next_elements', 'next_sibling', 'next_siblings', 'parent', 'parentGenerator', 'parents', 'parserClass', 'parser_class', 'prefix', 'preserve_whitespace_tags', 'prettify', 'previous', 'previousGenerator', 'previousSibling', 'previousSiblingGenerator', 'previous_element', 'previous_elements', 'previous_sibling', 'previous_siblings', 'recursiveChildGenerator', 'renderContents', 'replaceWith', 'replaceWithChildren', 'replace_with', 'replace_with_children', 'select', 'select_one', 'setup', 'smooth', 'sourceline', 'sourcepos', 'string', 'strings', 'stripped_strings', 'text', 'unwrap', 'wrap']
```

我们这里只看其中的文本内容，即 text 属性

```
print(targets[0].text)
```

## Java8编程实战

是想要拿取的东西

写个循环把所有标签的文本信息拿出来就行了

## csv 保存数据

- 逗号分隔值Comma-Separated Values, CSV, 有时也称为字符分隔值，因为分隔字符也可以不是逗号)

```
import csv
```

一行一行地写进去

```
# 以 w (write) 的方式打开或创建一个文件，执行完会自动关闭
# ./ 在
with open('./data.csv', 'w', encoding='utf-8') as f:
    writer = csv.writer(f)
    for item in targets:
        writer.writerow([item.text])
```

结果

这里一行只有一个元素，每行间没有看见分隔

Java8编程实战

面试之排序算法

实用数据分析：数据分析师从小白到精通

Java之路

韦东山嵌入式Linux第一期视频

编程可以这样学

离散数学总复习精华版(最全 最简单易懂)已完结

JavaScript 抢购茅台脚本 仅供学习

C/C++指针详解之基础篇（史上最全最易懂指针学习指南！！！！）

高校校园网建设方案【含网络拓扑图+拓扑结构图+配置命令】（详细版）

在这里插入代码片

...  
...

如果把 `writerow([item.text])` 里面的中括号[]去掉，`item.text` 是一个字符串，字符串里面的每个字符会被当作一个元素，就可以明显的看到逗号了

```
with open('./data.csv', 'w', encoding='utf-8') as f:  
    writer = csv.writer(f)  
    for item in targets:  
        writer.writerow(item.text)
```

J,a,v,a,8,编,程,实,战

面,试,之,排,序,算,法

实,用,数,据,分,析,：,数,据,分,析,师,从,小,白,到,精,通

J,a,v,a,之,路

韦,东,山,嵌,入,式,L,i,n,u,x,第,一,期,视,频

编,程,可,以,这,样,学

高,校,校,园,网,建,设,方,案,【,含,网,络,拓,扑,图,+ ,拓,扑,结,构,图,+ ,配,置,命,令,】,（,详,细,版,）

基,于,L,S,T,M,的,多,变,量,多,步,序,列,预,测,模,型,实,战,「,超,详,细,实,现,说,明,讲,解,」

基,于,深,度,学,习,的,人,脸,识,别,系,统,实,战,【,从,零,开,始,搭,建,你,的,人,脸,识,别,系,统,】

【,超,详,细,】 ,计,算,机,操,作,系,统,总,结,及,思,维,导,图,（,汤,子,瀛,版,）

...  
...

如果数据量小，还可以可以一次性地通过 pandas 保存

[列表推导式](#)

```
import pandas as pd
```

```
# 用列表推导式把标签里的文字放入一个列表中  
content = [item.text for item in targets]  
# DataFrame 是一个二维的表格结构  
data = pd.DataFrame(content, columns=['文字内容'])  
data.to_csv('./data.csv')
```

DataFrame 是一个二维的表格结构

有从 0 开始自动生成的索引，还有每列的标签 columns

在 python 里可以看作这样一个字典

```
{'column1': [item01, item11, item21],  
'column2': [item02, item12, item22]...}
```

```
,文字内容  
0,Java8编程实战  
1,面试之排序算法  
2,实用数据分析：数据分析师从小白到精通  
3,Java之路  
4,韦东山嵌入式Linux第一期视频  
5,编程可以这样学  
6,离散数学总复习精华版(最全 最简单易懂)已完结  
7,JavaScript 抢购茅台脚本 仅供学习  
8,C/C++指针详解之基础篇(史上最全最易懂指针学习指南!!!)  
9,高校校园网建设方案【含网络拓扑图+拓扑结构图+配置命令】(详细版)  
10,如何通过QQ号获取绑定的手机号  
11,win10产品密钥 win10专业版激活码key  
12,免费8节课让你成为Flink专家!  
13,Star国产AI框架MegEngine  
14,2020中国开发者有奖大调查  
15,Unity官方游戏开发课程,第三期招生火热开启!  
16,华为开发者专区  
17,英特尔开发人员专区  
18,Bitcoin SV开发者专区  
...  
...
```

上面是最基础的，平时使用还可能用到网还可能用到网络API，请求时需要加上请求头，获取cookies，或者使用webdriver、selenium，多线程等操作