反爬虫技术总结

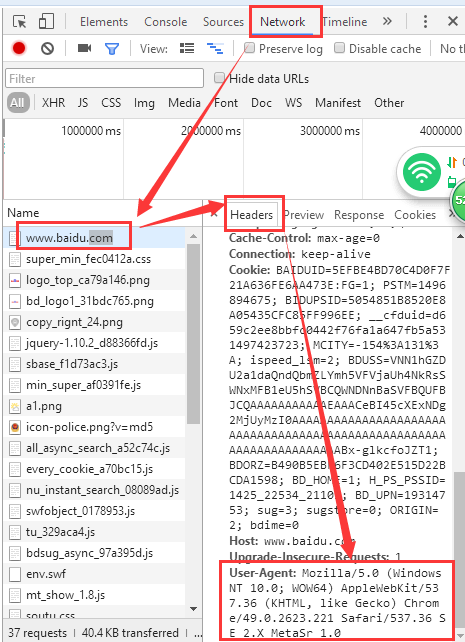
# 修改User-Agent

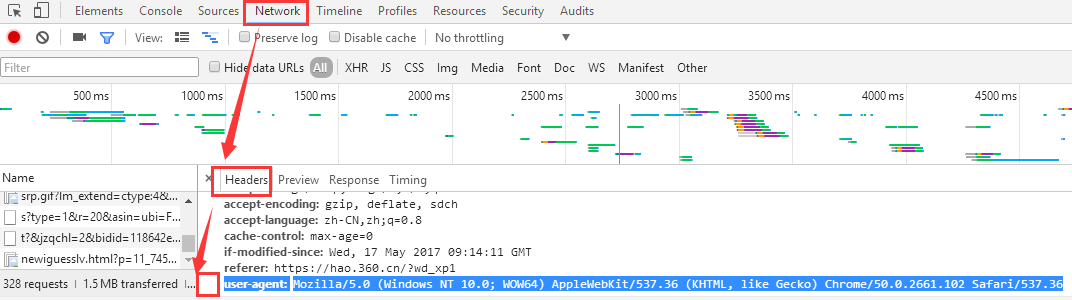
## 获取当前浏览器的user-agent的方法

利用浏览器的开发者工具，在Network中随便点击一次请求，找到Headers，即可查看到User-Agent内容。此时在复制到爬虫代码中，利用request的请求参数headers，即可伪装成一个**浏览器**。

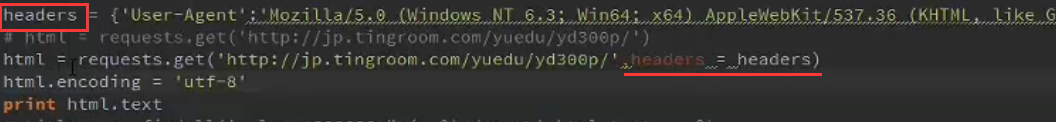
headers={“**User-Agent**”:”**Mozilla/5.0** (Windows NT 10.0; WOW64) AppleWebKit/537.36 (KHTML, like Gecko) **Chrome**/49.0.2623.221 Safari/537.36 SE 2.X MetaSr 1.0”}

headers={“user-agent”:”Mozilla/5.0 (Windows NT 10.0; WOW64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/50.0.2661.102 Safari/537.36”}





## 伪装方法



## 对于User-Agent 可以直接在settings.py中设置

在**settings.py**文件中添加一句：

USER\_AGENT ='Mozilla/5.0 (Windows NT 10.0; WOW64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/49.0.2623.221 Safari/537.36 SE 2.X MetaSr 1.0'

**这样**就是在Request的头部自动添加了该头信息，等价于

header = {'user-agent':'Mozilla/5.0 (Windows NT 10.0; WOW64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/49.0.2623.221 Safari/537.36 SE 2.X MetaSr 1.0'}

yield scrapy.Request(url = url,callback=self.parse,headers = header)