

## 静态网站数据获取

### 作业

1. 抓取腾讯首页 <http://news.qq.com/> 或腾讯社会新闻 <http://society.qq.com/> 中一页的新闻

- ① 需要标题、来源、时间、具体新闻文本内容
- ② 在文件中存储
- ③ 一条新闻一行(段落尾部的换行符请处理)
- ④ 格式为 标题#####来源#####时间#####具体新闻文本内容

2. 模仿链家网例子，爬取北京链家网租房信息 <https://bj.lianjia.com/zufang/>

### 资料说明

- 1. Beautiful Soup 说明文件 打印版.pdf 如文件名
- 2. requests-docs-cn.epub epub 格式文档
- 3. requests-docs-cn.zip 解压后放类似 tomcat 中运行，访问 index.html
- 4. Xpath 介绍.pdf
- 5. lesson2 讲师课件