**技能训练3：爬取斗鱼直播平台数据**

## 1 训练目的

1. 知道什么是动态网页
2. 掌握爬取动态网页的selenium和webdriver技术
3. 能够与网页上的元素进行交互，或执行一些操作后爬取页面数据

## 2 技能训练题目

爬取斗鱼直播平台（https://www.douyu.com/directory/all）的所有房间信息，包括房间标题和观众人数。

必做：要求如下：

1. 定义一个用于爬取斗鱼网站的类DouyuSelenium；
2. 在DouyuSelenium类中定义3个方法，分别是\_\_init\_\_()、start\_work()、tear\_down()。
3. \_\_init\_\_()方法用来创建浏览器
4. 用于获取数据的方法start\_work()，用于爬取和解析网页数据，并将最终的数据以“观众人数：\*\* 房间标题：\*\*”的形式输出；（只要求爬取前5页数据，在点击“下一页“的循环中获取页面源代码解析数据）
5. 声明一个用于退出webdriver的方法tear\_down()；
6. 在main中创建DouyuSelenium类实例，依次调用每个方法进行爬取。

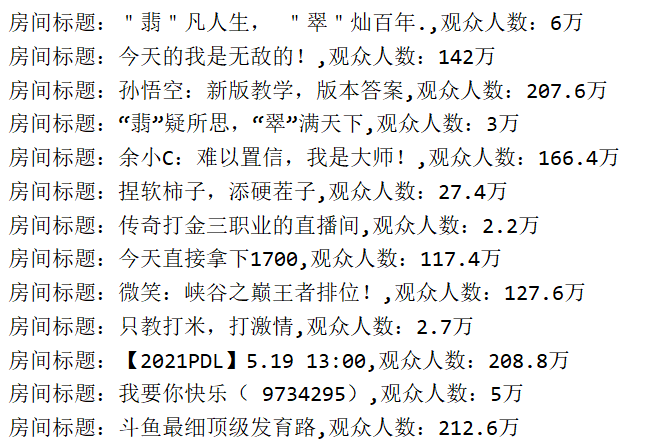


图 控制台打印效果参考

选做：爬取的数据存入json文件，文件名：douyu\_姓名首字母.json

## 3 提交的作业文档中应该包含代码文件、代码截图和运行效果截图。