**技能训练2：爬取社招及电影网数据**

## 1 训练目的

1. 熟悉解析数据的技术
2. 掌握正则表达式的使用，会使用re模块解析网页数据
3. 掌握XPath语法的使用，会使用lxml库解析网页数据
4. 掌握BeautifulSoup的使用，会使用bs4库解析网页数据

## 2 技能训练题目

1. **爬取腾讯社招网数据**

1.1页面地址： https://careers.tencent.com ——》查看工作岗位

（地址通过Request URL获取）（参考5.7.3）

获取json数据如下图所示：



1.2 通过jsonpath解析出所有的工作岗位



1. **爬取电影网数据**

https://ssr1.scrape.center/page/1

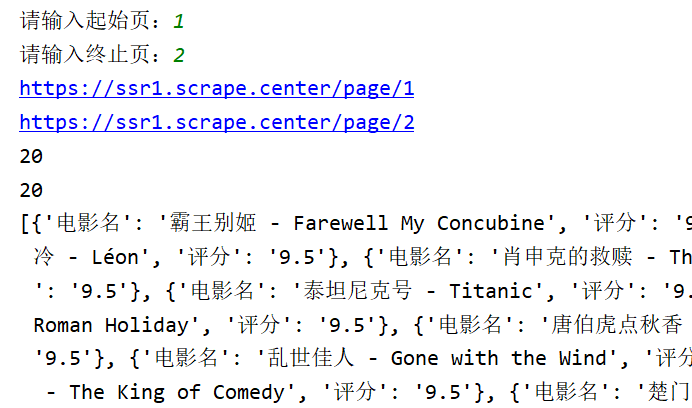
2.1、**获取页面源代码**如下图所示：（参考5.7.3）



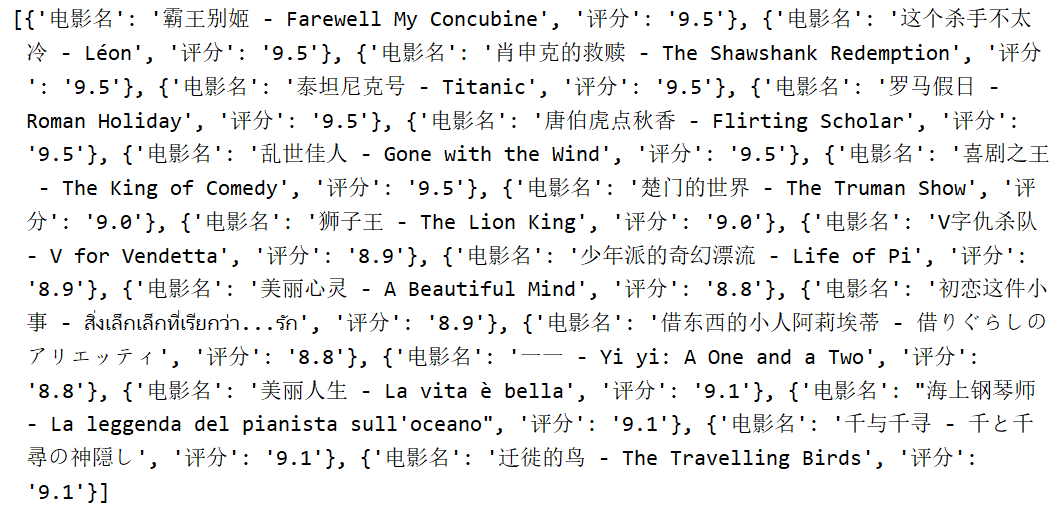
2.2、（选做）**使用正则表达式解析出电影名和评分**，如下图所示：

names\_list = re.findall(r'class="m-b-sm">(.\*?)</h2>',html)

score\_list= re.findall(r'class="score m-t-md m-b-n-sm">\n(.\*?)</p>',html)



2.3、**使用lxml库解析出电影名和评分，要求保存在json文件中，控制台如下图所示：**



json文件参考下图：



2.4、**使用bs4库解析出电影名和评分，要求保存在json文件中**

## 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。