**技能训练5：爬取阳光热线问政平台或豆瓣Top250数据**

## 1 训练目的

1. 掌握Scrapy框架的架构和运作流程
2. 掌握Scrapy框架的基本操作

## 2 技能训练题目（任选一题）

（一）、创建一个Scrapy项目SunHot，用于爬取阳光问政平台的部分信息，主要包括投诉帖子的编号、帖子的URL、帖子的标题及帖子的内容（选做），其网址为<https://wz.sun0769.com/political/index/politicsNewest> 。具体要求如下：

1. 在项目的/spiders目录下，新建用作爬虫的文件sun.py
2. 在sun.py文件中，使用parse()方法取出每个页面中帖子的编号、URL及链接，再从中迭代获取每个帖子，并交给回调函数parse\_item()处理。
3. 在parse\_item()方法中，提取帖子内容。
4. 将爬取到的数据使用管道item pipeline存储到json文件或者数据库中。
5. 要求爬出所有页面信息。

（二）、将豆瓣T250排行榜（https://movie.douban.com/top250）爬取到的数据使用管道item pipeline存储到json文件或者数据库中。要求爬出所有页面信息。（爬取哪些数据自定）

## 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。