**作业：抓取豆瓣电影短评**

**要求：**

1. 选择一部你喜欢的电影，从豆瓣电影上抓取短评数据

示例：

流浪地球短评链接

<https://movie.douban.com/subject/26266893/comments>

1. 基于Scrapy构建（已给样例代码 RUCNews.zip），Scrapy教程可以参考：  
   <https://scrapy-chs.readthedocs.io/zh_CN/0.24/intro/tutorial.html#spider>
2. 网页爬取和解析方法可以参考：  
   <https://www.cnblogs.com/smallSevens/p/10427597.html>
3. 至少抓取500条，并对抓取到的评论数据进行分词
4. 输出文件格式：文本文件，每一行一条评论，包括“[作者]:分词后的文本”。

如：

|  |
| --- |
| [乌鸦乌鸦]:优点 和 缺点 都 很 好莱坞 ， 为什么 至今 没 看到 “ 魔改 ” 这个 词 ？ 好莱坞 大片 已经 四平八稳 到 ， 所有 反派 都 是 虚构 人物 ， 且 都 是 个人 品质恶劣 。 本片 的 坏人 竟然 是 木星 ！ 没有 一个 人类 反派 ， 这是 求生欲 多强 啊 ？  [杨迪]:说实话 我 有些 震撼 一 开场 的 地下城 我 看着 就 非常 有 与 之前 中国 电影 不 一样 的 感觉 还有 “ 饱和 式 营救 ” 非常 合情合理 值得 打 五星……  [sweet1002008]:楼上 那位 说 什么 煽动 妄想 的 怕 是 个 外国人 吧 中国 的 神话 本来 就是 大禹治水 精卫填海 夸父追日 后羿 射日 拍个 科幻片 难道 不 应该 遵循 自己 的 传统 文化 ？ ？ ？ |

1. 根据抓取的评论构建词云，可以选取自己喜欢的图片作为背景。

**提交：**

提交zip包，命名为“学号\_姓名”，其中包括：

1. 含有你的代码的scrapy工程（目录）
2. 爬取到的评论文件
3. 一个word文档对你的代码进行说明，并包含你绘制的词云。