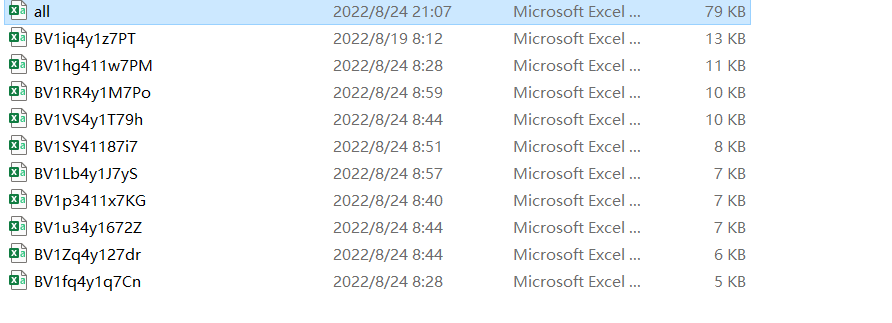
**运行文档**

**姓名：张毅轩 学号：1120211228 班级：软件工程2021研**

本文档分三个部分，分别对应爬虫程序，词云图程序，以及LDA主题模型程序。

备注：本文档中，为了节约时间，爬虫程序只爬取一个视频。词云图和LDA程序，使用的数据为3700条弹幕的总和。 图形用户界面, 应用程序, 表格, Excel

描述已自动生成

1. 爬虫程序

如图1所示，在程序中的源数文件中，存放一个EXCEL文件，打开后，将需要进行爬虫的视频的BV号和相应网站地址，填入对应栏目中。运行B站弹幕爬取程序即可，运行结果如图2。爬取的弹幕存放在弹幕文件夹中，名称是写入的BV号，如图3，具体内容如图4。

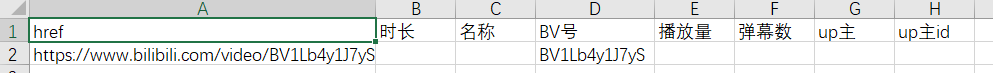


图1，填入爬取视频的BV号和网址

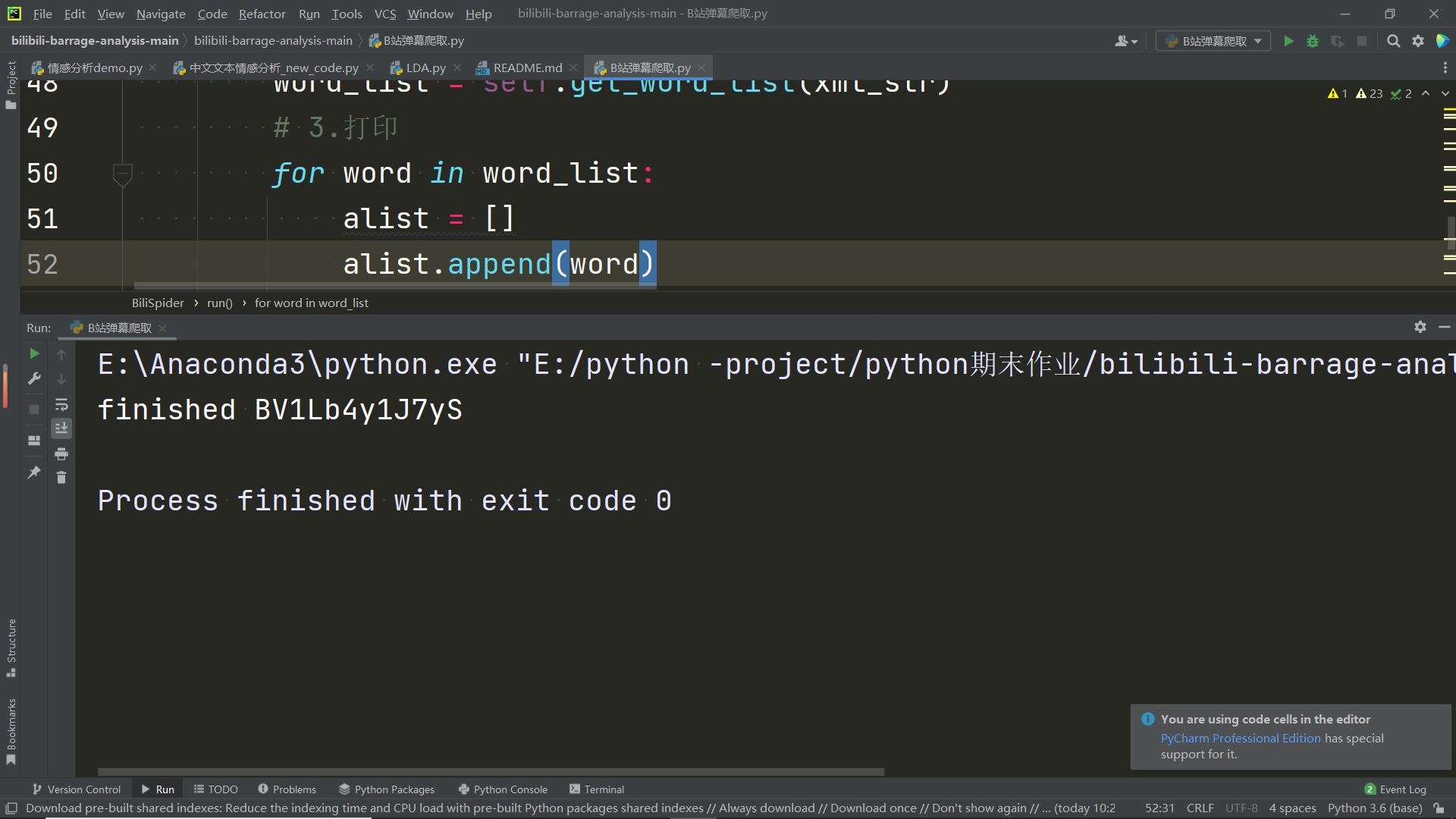


图2，爬虫程序运行结果

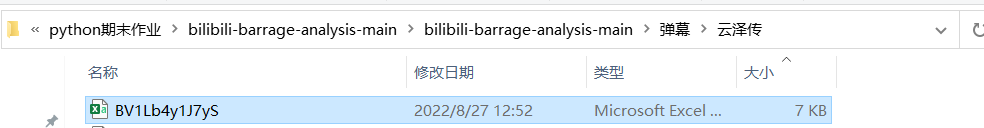


图3，爬取的弹幕图形用户界面, 应用程序, 表格, Excel

描述已自动生成

图4，爬取数据的具体内容

1. 词云图程序

如图5所示，运行词云图程序。词云图如图6。

在词云图中，词语的大小和词频成正比。

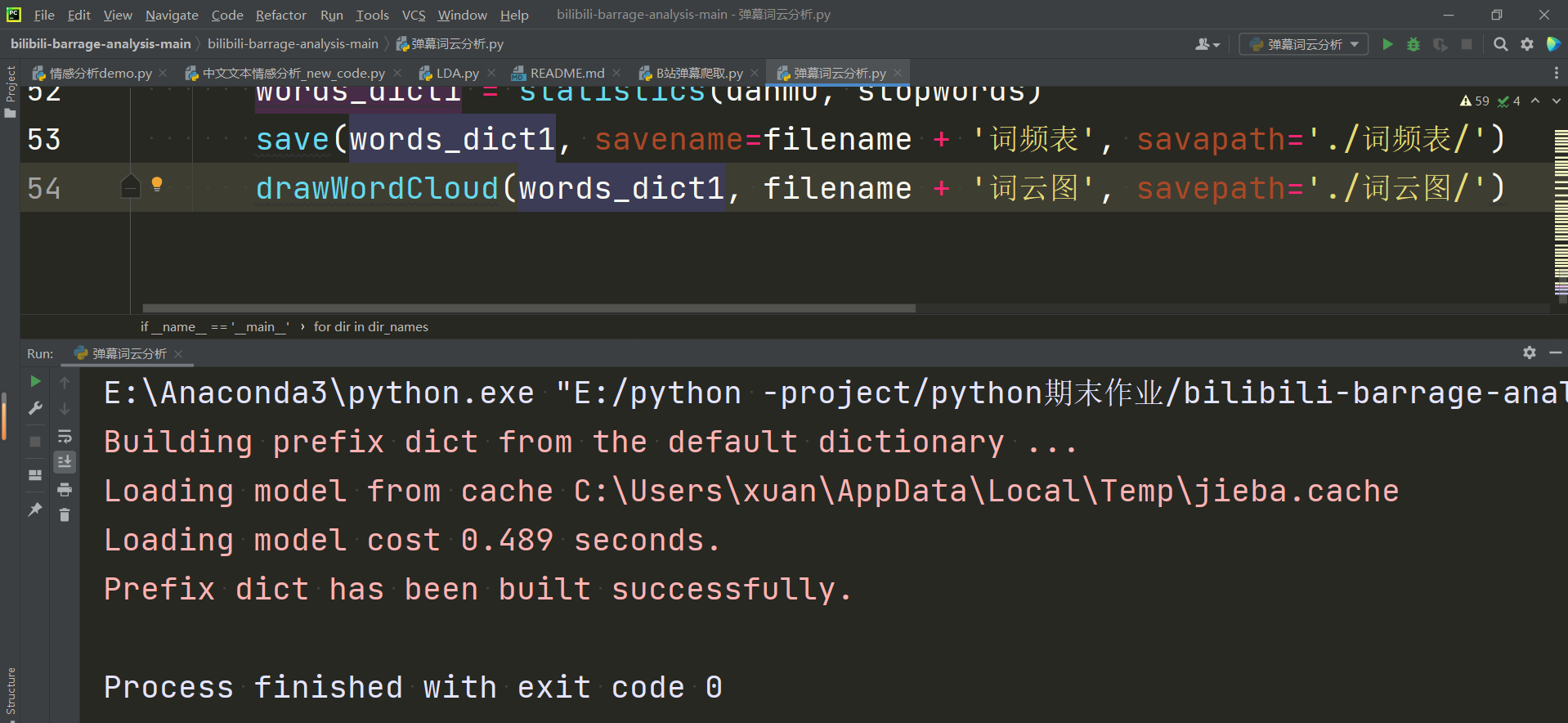


图5，词云图程序运行结果

电脑萤幕画面

描述已自动生成

图6，词云图

1. LDA主题模型程序

LDA主题模型程序运行如图6，结果如图7。

经过多次运行程序，最终选择的主题数量为5时，LDA主题模型运行效果最好。通过主题词语可推断，主题0讨论主角的演技，相关的词语为演技，冷漠，表情等。主题1讨论感情，对应的词语为行动，表情，感觉。主题2为对角色的感受。主题3为讨论剧情走向，主题4对应剧中感情戏的结局。

由图7可知，主题数量为5时候，划分效果最佳，5个主题对应的蓝色圆圈没有重叠，表明主题间的差异性很大。

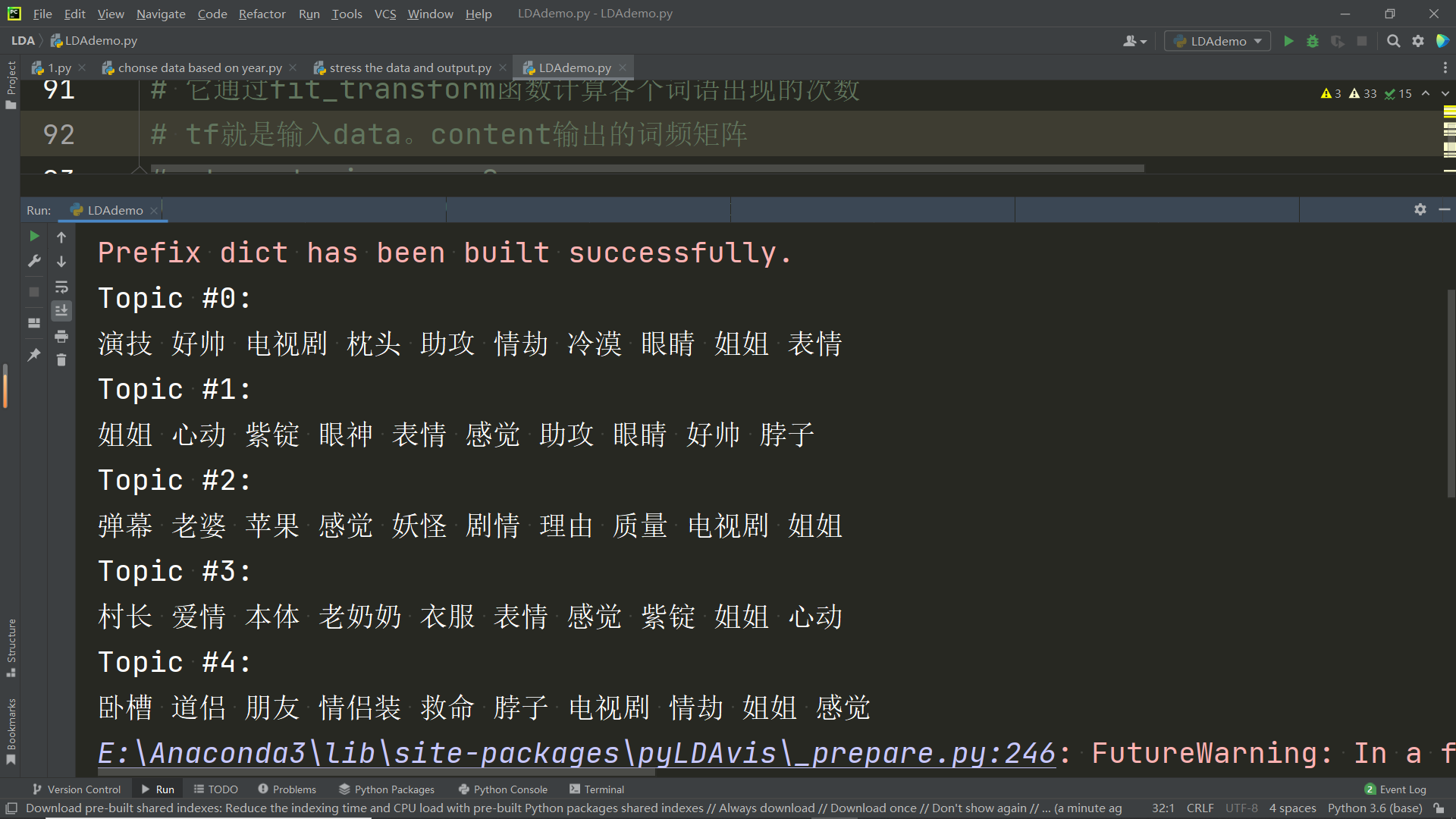


图6，LDA主题模型程序运行结果

图表

描述已自动生成

图7，LDA结果