

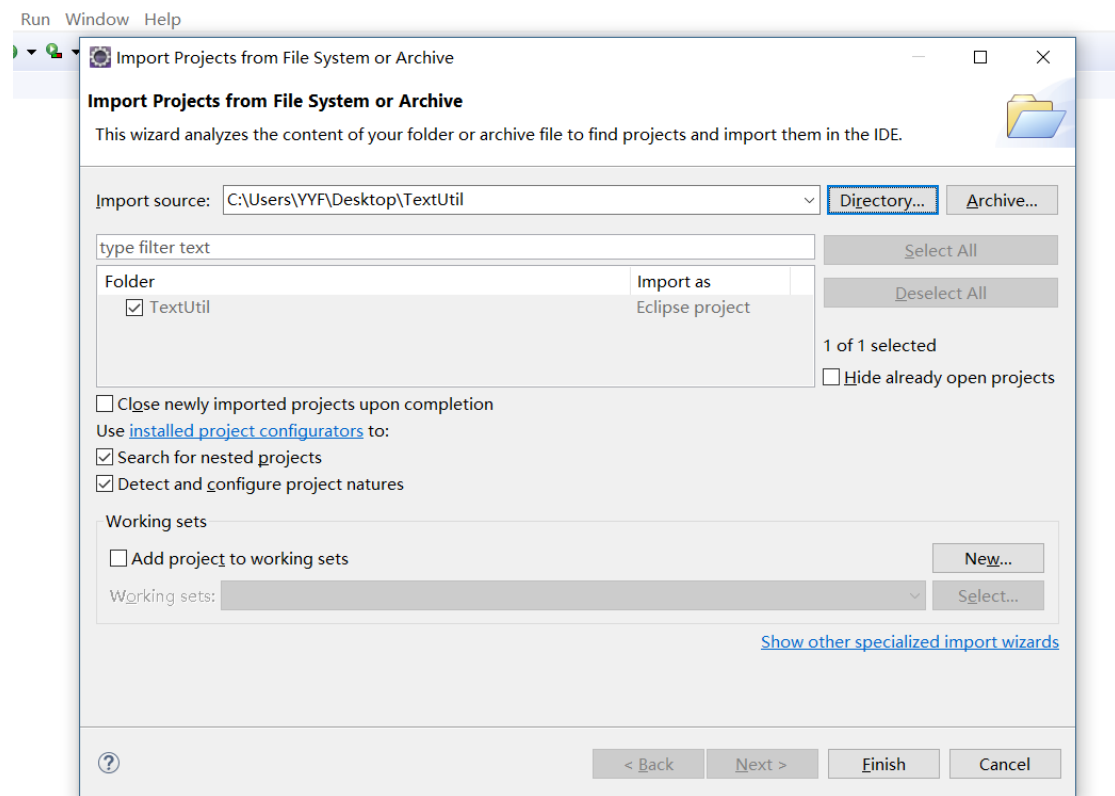
程序说明

附件中程序主要包含了两部分。其中 TextUtil 使用 JAVA 通过 eclipse 编写。Poem 通过 python 编写。

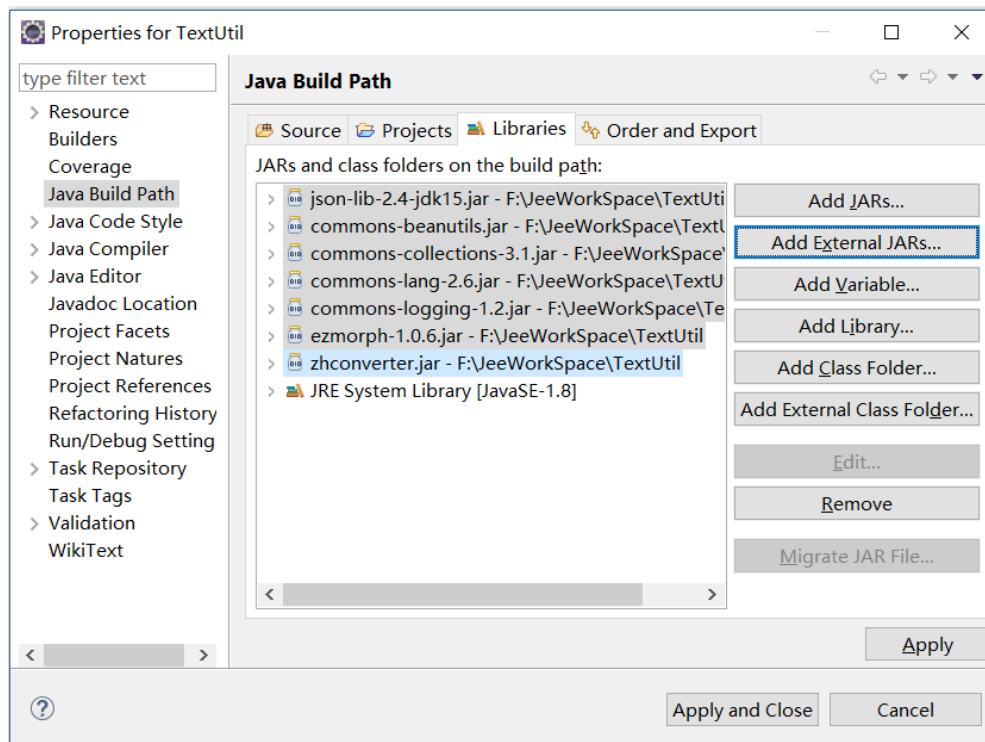
TextUtil:

运行流程:

1. 打开 eclipse, 解压项目包。
2. File->Open Project From File System, 找到项目位置并打开。



3. 导入所需依赖的架包, 配置路径。



3. 以 Java Application 方式运行程序

TextOperation [Java Application] C:\Program Files\Java\jre1.8.0_181\bin\javaw.exe (Jan 15, 2019, 4:52:27 PM)

0号数据集处理成功!

1000号数据集处理成功!

2000号数据集处理成功!

可以看见，程序开始运行处理文档。（注意，此程序会一直运行到处理完全部文件，耗时较长，可自己手动关闭或设置 Constant 类里面的 MAXLENGTH 变量。）

陆游.txt	2019/1/13 14:18	文本文档	1,176 KB
白居易.txt	2019/1/13 14:31	文本文档	493 KB
梅尧臣.txt	2019/1/13 14:11	文本文档	478 KB
杨万里.txt	2019/1/13 14:19	文本文档	475 KB
刘克庄.txt	2019/1/13 14:22	文本文档	475 KB
苏轼.txt	2019/1/13 14:13	文本文档	460 KB
方回.txt	2019/1/13 14:25	文本文档	443 KB
赵蕃.txt	2019/1/13 14:20	文本文档	439 KB
黄庭坚.txt	2019/1/13 14:14	文本文档	338 KB
张耒.txt	2019/1/13 14:14	文本文档	327 KB
苏辙.txt	2019/1/13 14:13	文本文档	305 KB
韩洙.txt	2019/1/13 14:21	文本文档	304 KB
周紫芝.txt	2019/1/13 14:16	文本文档	299 KB
陈造.txt	2019/1/13 14:19	文本文档	297 KB
释德洪.txt	2019/1/13 14:15	文本文档	286 KB
杜甫.txt	2019/1/13 14:31	文本文档	276 KB
李纲.txt	2019/1/13 14:16	文本文档	274 KB
王十朋.txt	2019/1/13 14:17	文本文档	272 KB
刘敞.txt	2019/1/13 14:12	文本文档	266 KB
郭祥正.txt	2019/1/13 14:13	文本文档	261 KB
范成大.txt	2019/1/13 14:18	文本文档	235 KB
王安石.txt	2019/1/13 14:12	文本文档	234 KB
李白.txt	2019/1/13 14:31	文本文档	231 KB
楼钥.txt	2019/1/13 14:20	文本文档	220 KB
项安世.txt	2019/1/13 14:19	文本文档	202 KB
刘放.txt	2019/1/13 14:12	文本文档	202 KB
吕本中.txt	2019/1/13 14:16	文本文档	198 KB
释居简.txt	2019/1/13 14:21	文本文档	197 KB
宋祁.txt	2019/1/13 14:11	文本文档	197 KB

至此，文件处理完成。

PoemCloud

直接使用可以运行 python 的程序 (IDLE、Pycharm) 运行 word_cloud.py 程序即可, 该程序会根据前面得到的 dataset_Pretreatment 下的文件生成词云 (注意, 由于全部生成耗时、耗空间, 因此附件中的程序将 dataset_Pretreatment 改为 dataset_Pretreatment1, 即只生成 dataset_Pretreatment1 下的 10 个文件, 如有需要, 可将程序中的 dataset_Pretreatment1 改回 dataset_Pretreatment, 注意区别!!!)

