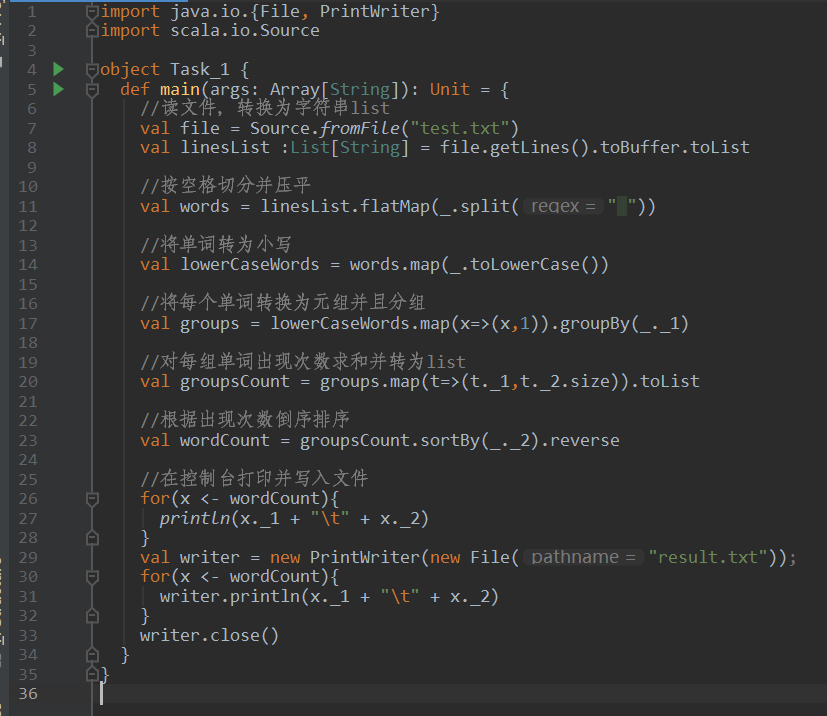
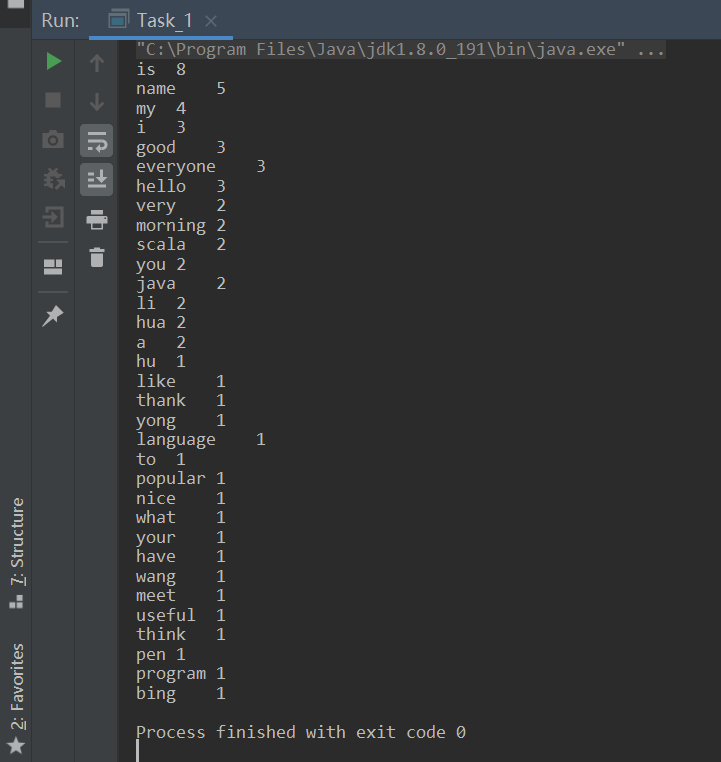
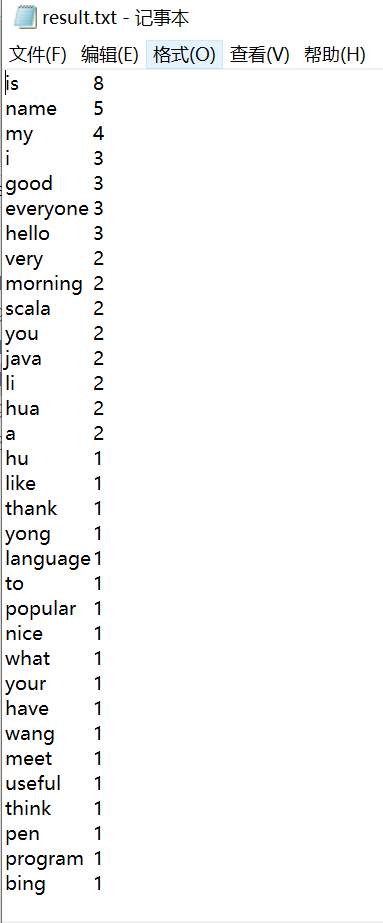
**第一次作业：使用scala 实现wordcount**

源程序：



运行截图：





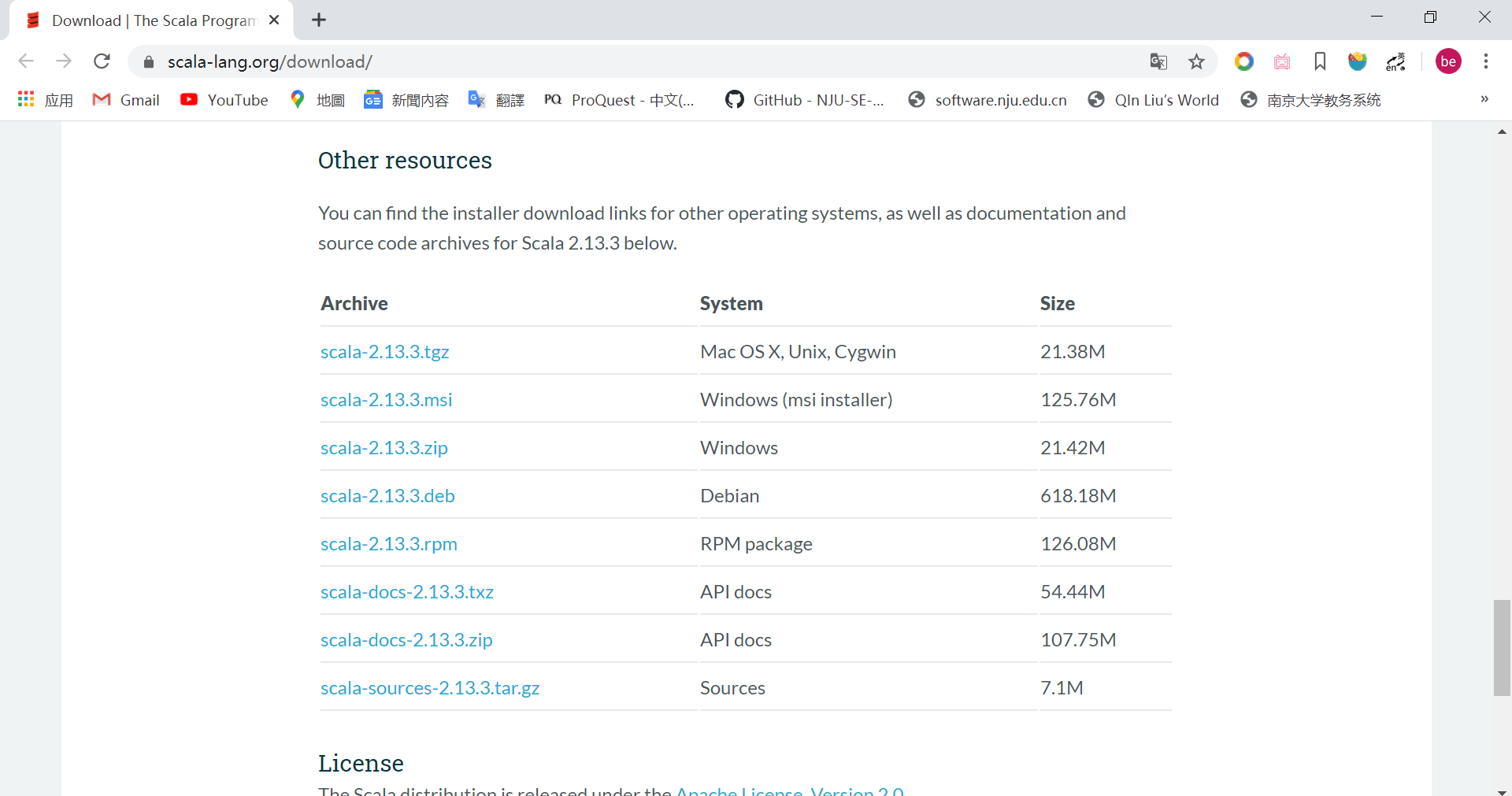
实验思路：

1. 使用fromfile读取文件
2. 使用flatMap将读取的List按照空格切分然后压平
3. 将单词都转化为小写
4. 将每个单词映射为元组，然后根据单词使用groupBy分组
5. 在每个组内部根据出现次数求和并转为list
6. 根据出现次数倒序排序
7. 在控制台打印并且写入文件

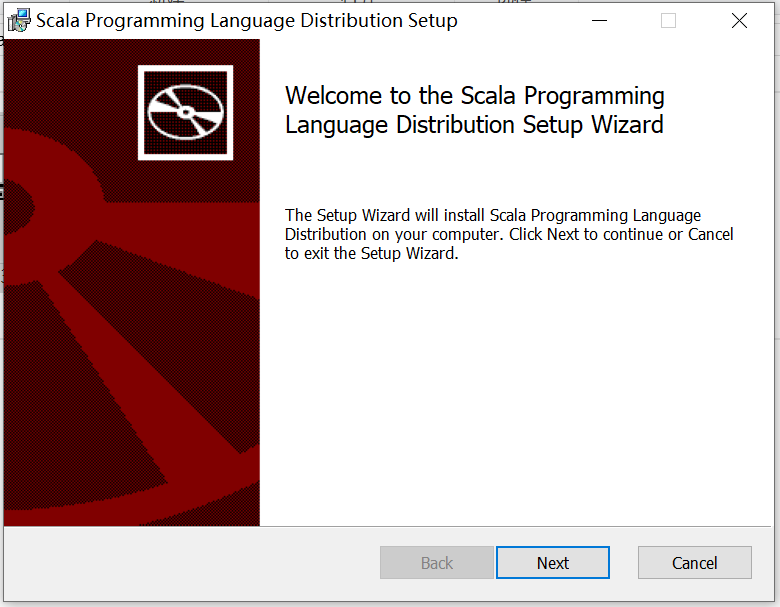
操作思路：

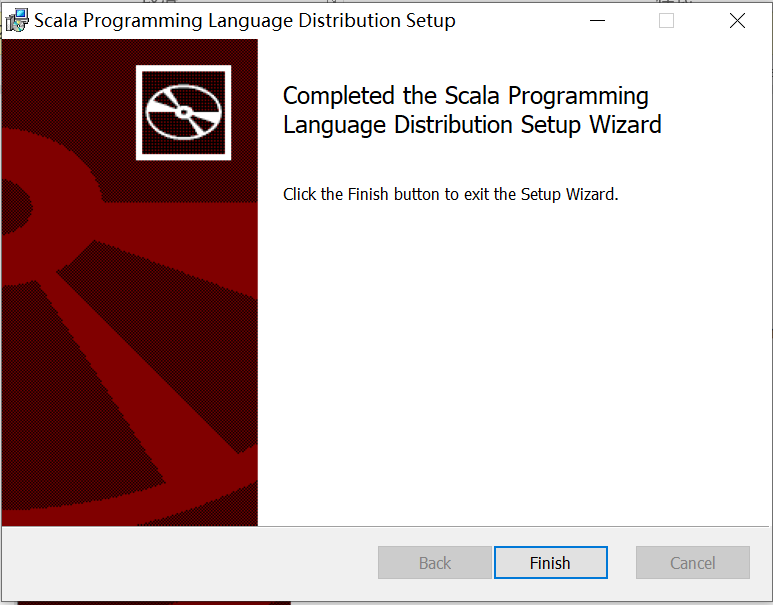
一、安装scala

1. 前往官网下载scala

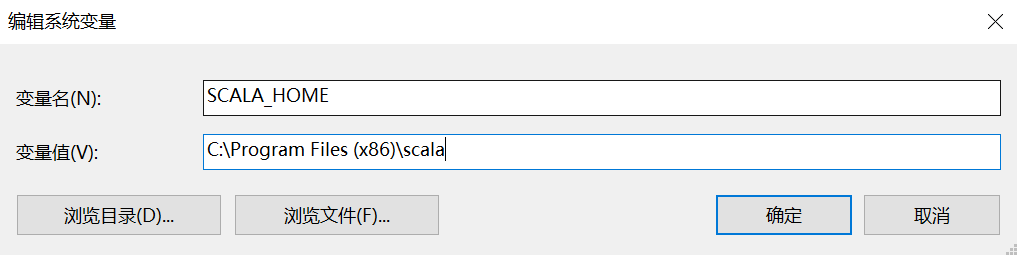


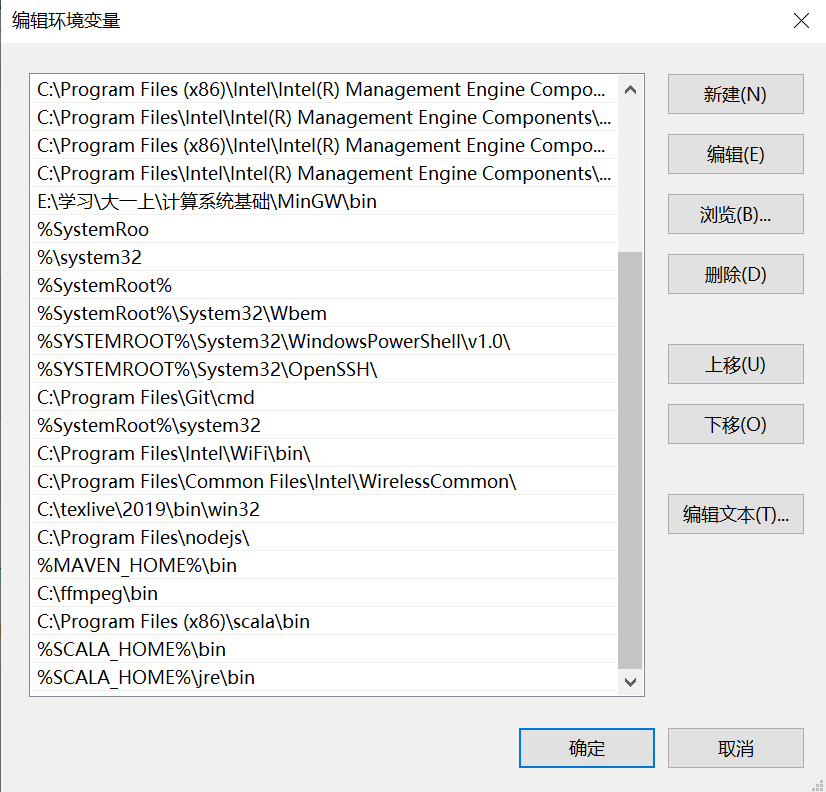
1. 安装scala

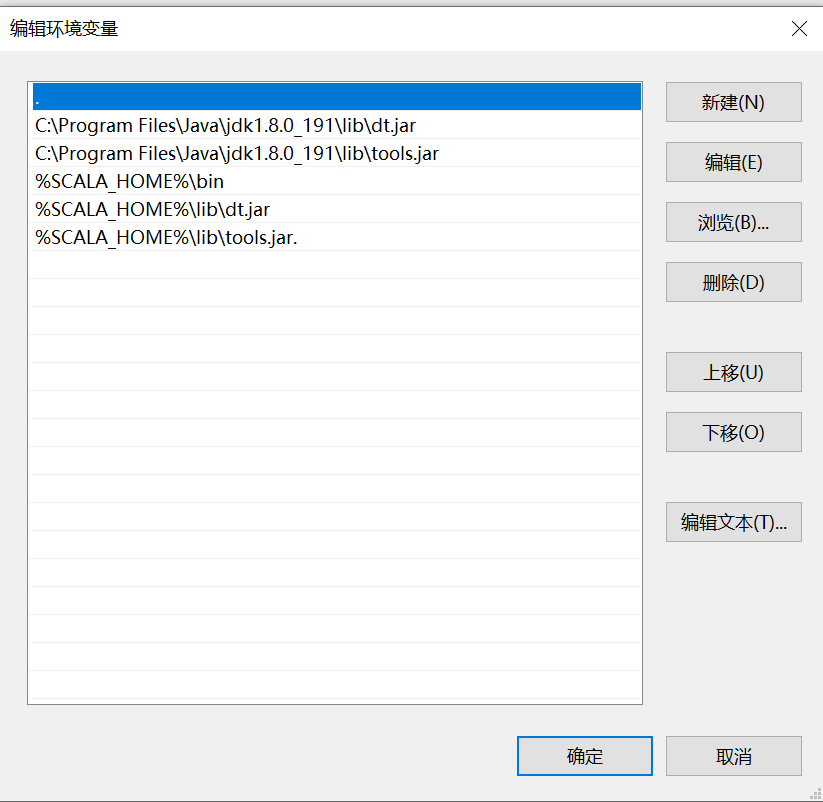




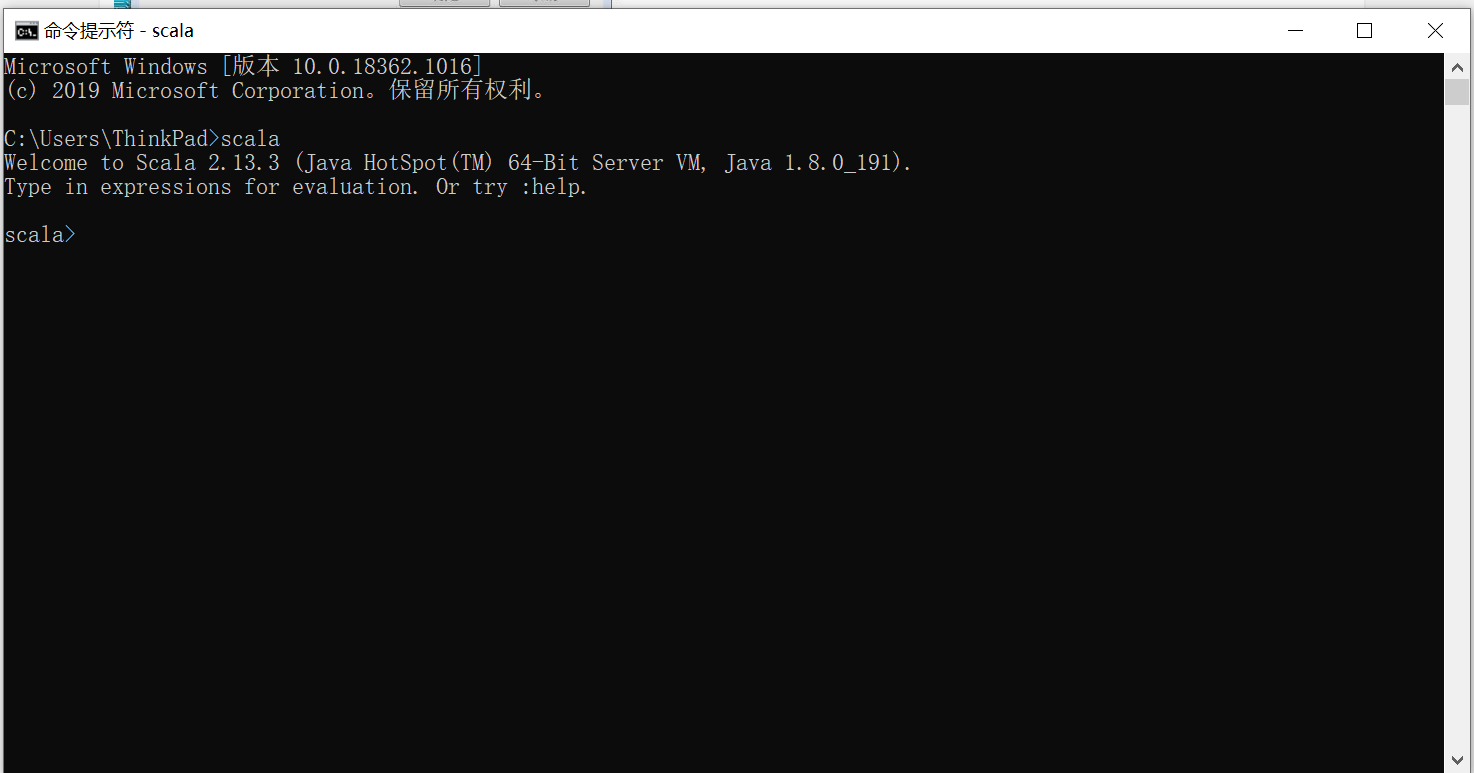
1. 配置环境变量





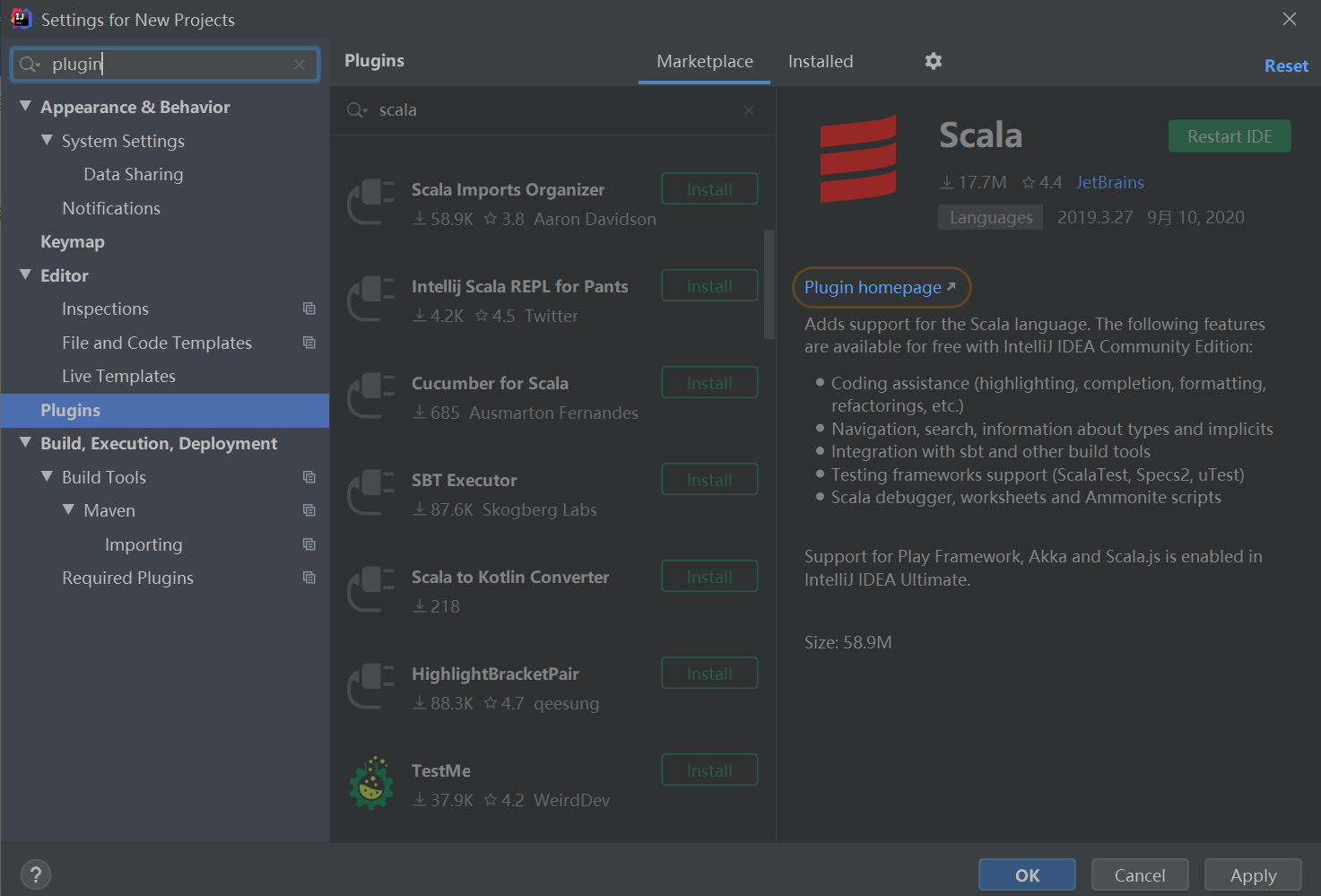


1. 在cmd中启动，安装成功

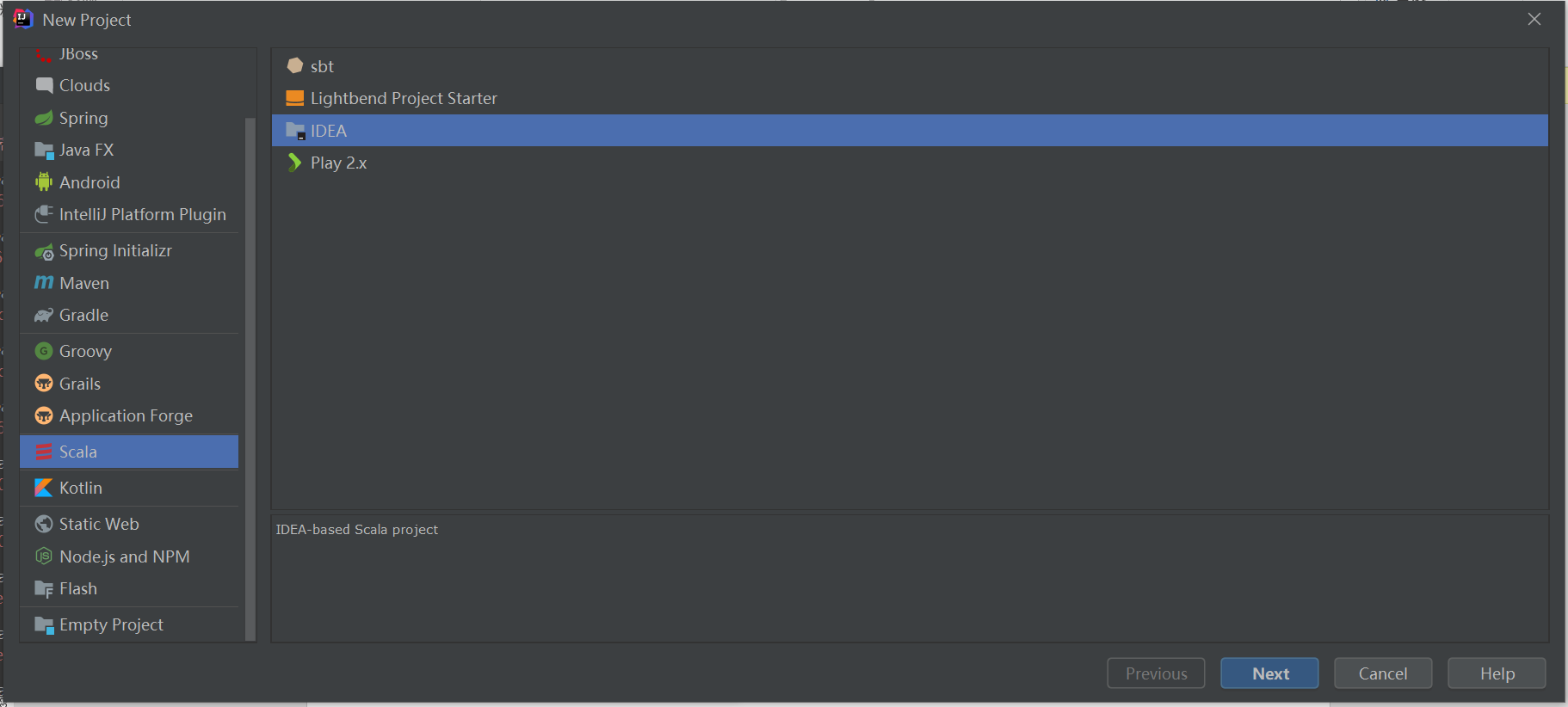


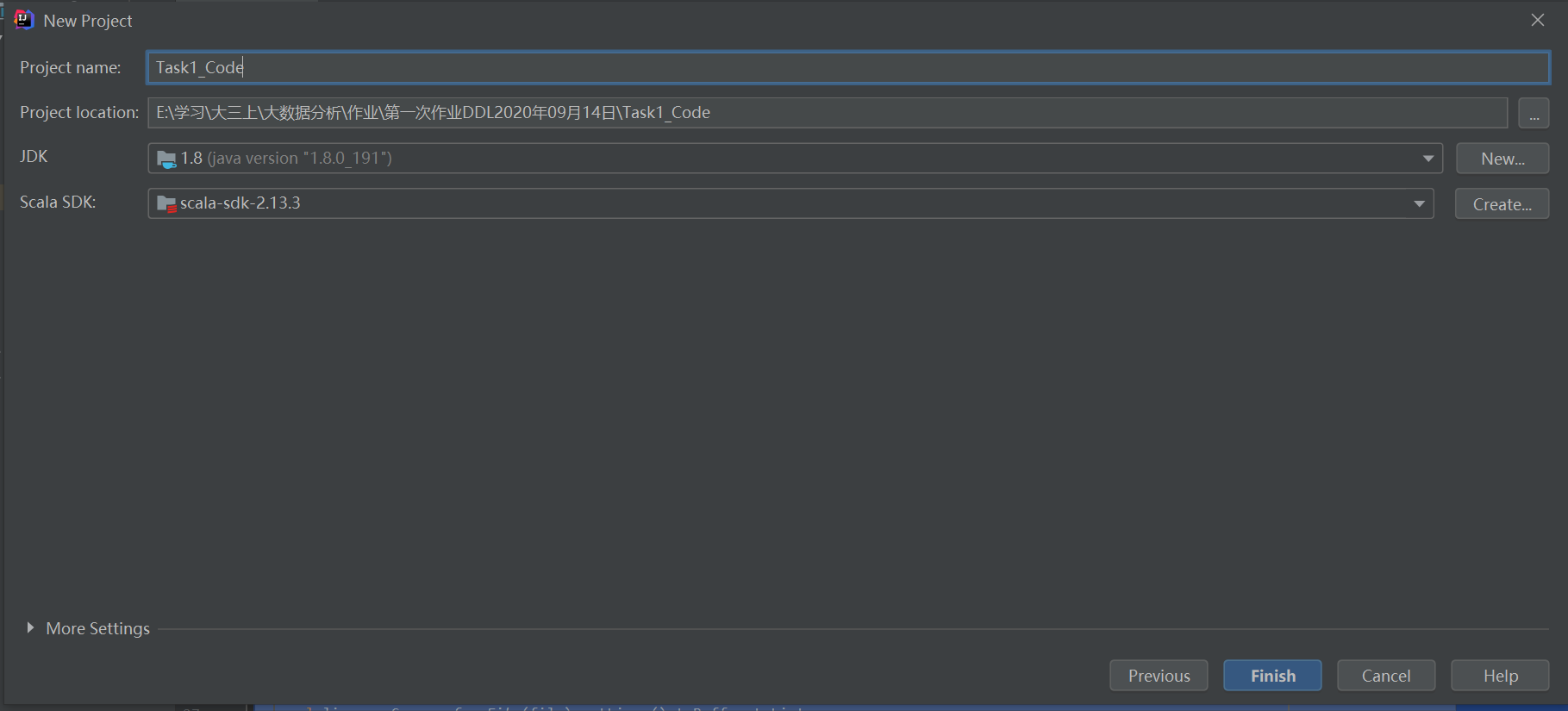
二、idea配置

1. idea安装scala插件



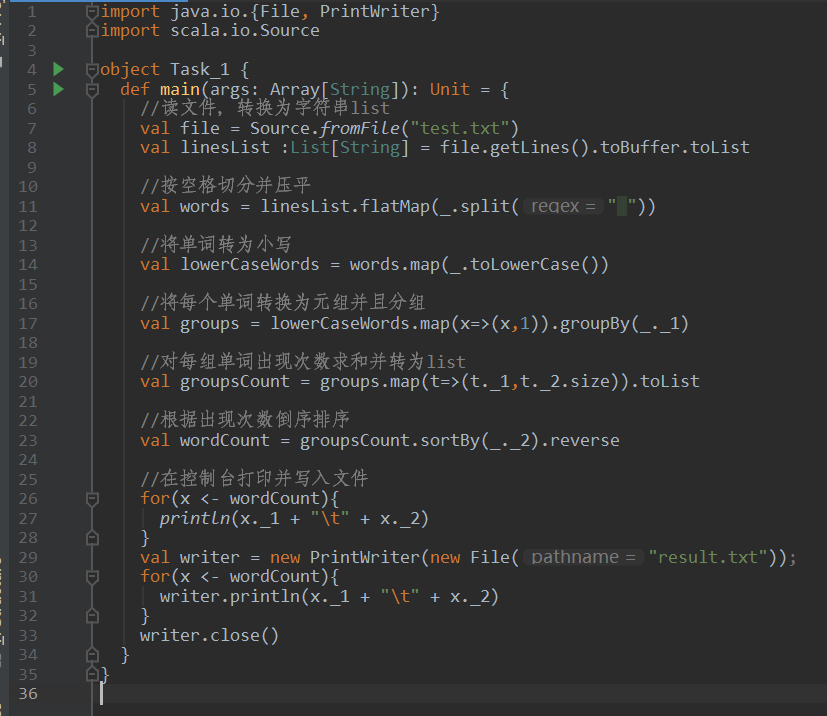
2. 创建一个scala项目



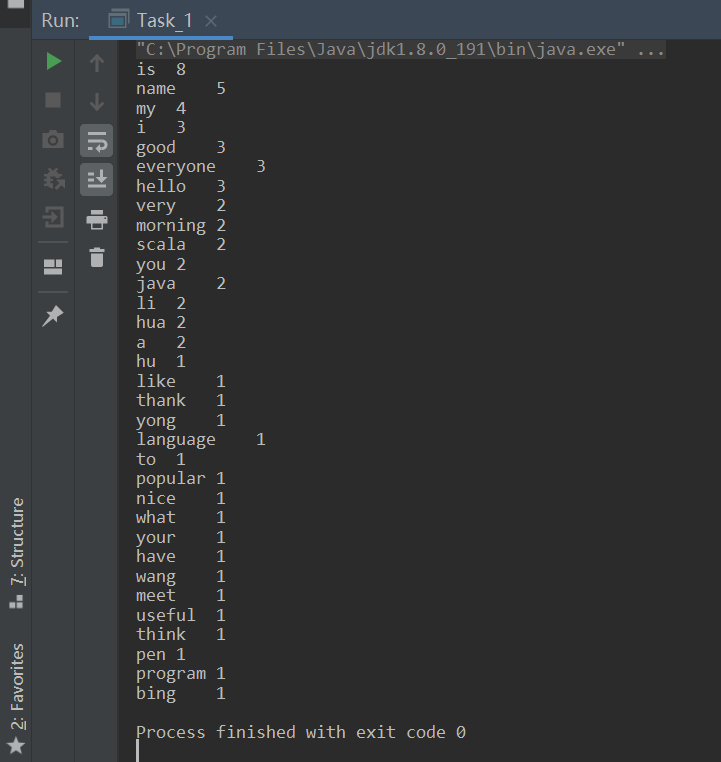


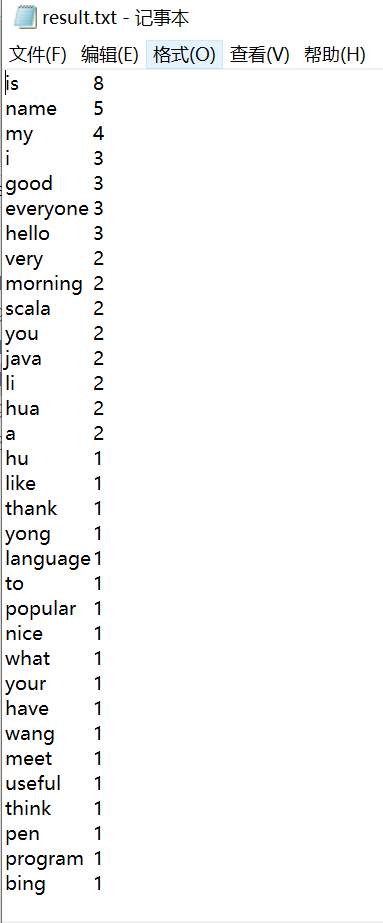
三、代码编写运行

1. 编写代码并且运行



2.运行结果





实验心得：

此次实验，第一次在电脑上安装并且配置了了Scala的运行环境，第一次在idea上面编写并运行了简单的RDD相关操作的程序，实现了对一个文本的词数统计并写入文件。接触到了Scala的语法，如map、flatMap、groupBy、sortBy等方法的初步使用。接触到了map、reduce功能，主要思想是分治。

Hadoop能提供分布式数据存储功能HDFS和数据处理功能MapReduce。Hadoop可用于搜索引擎、视频图片分析、推荐系统等。Spark对数据分析处理的速度比Hadoop的MapReduce更快。Spark可以用于较为复杂的如机器学习的数据处理。