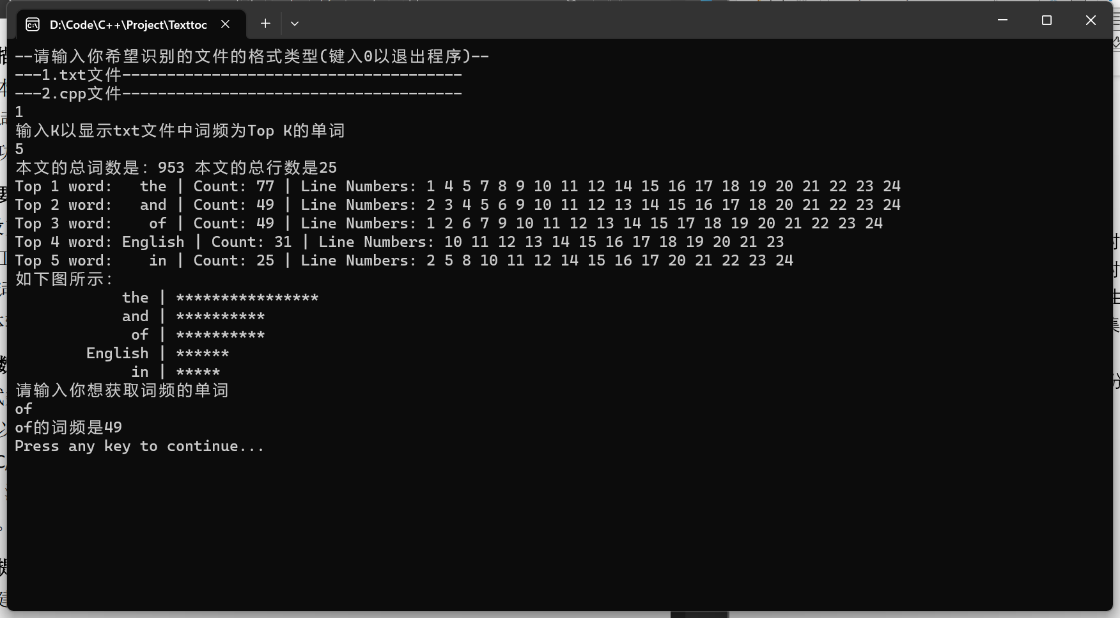
**Project 3 实验报告**

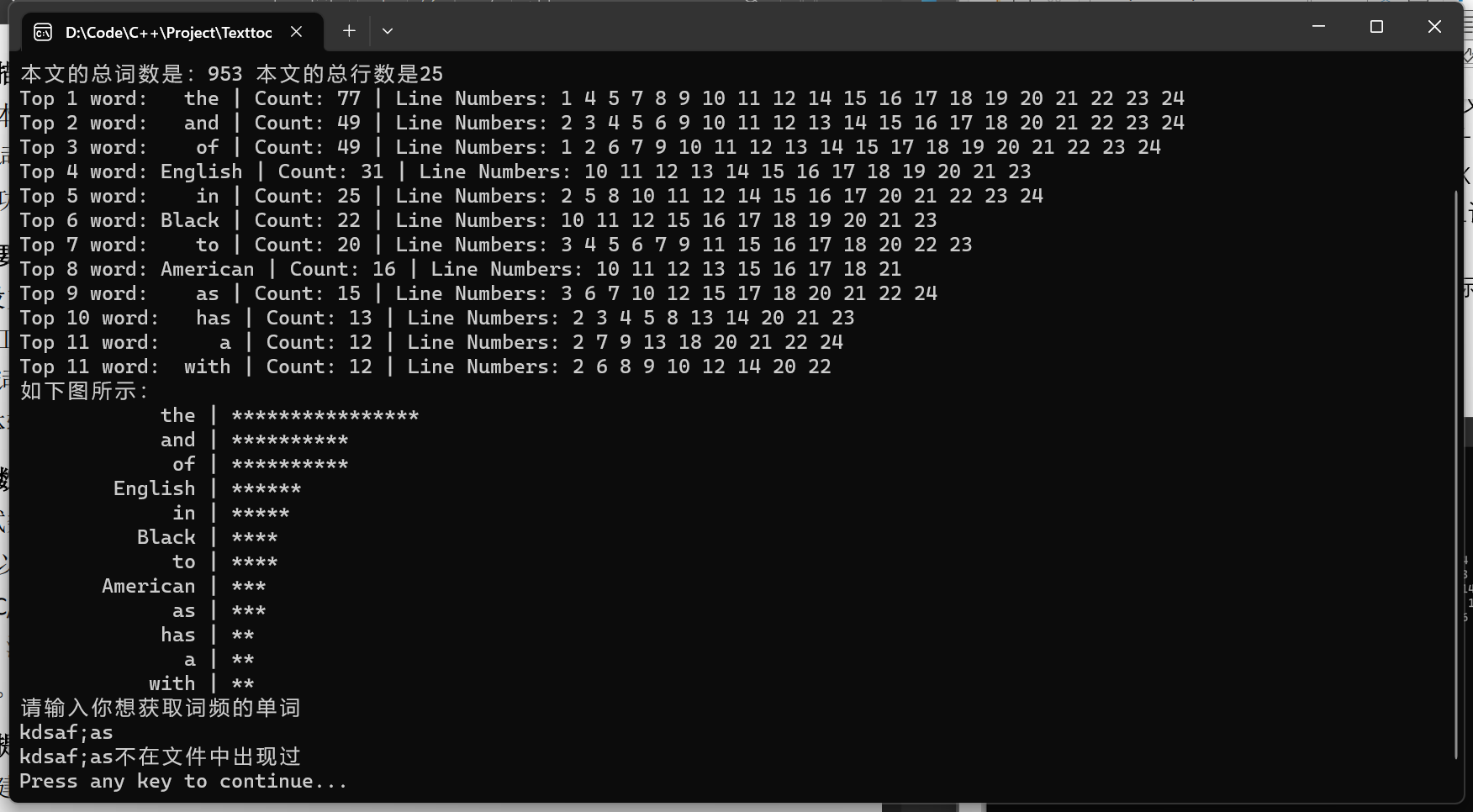
**学号**22336313 **姓名** 郑鸿鑫

1. 程序功能简要说明

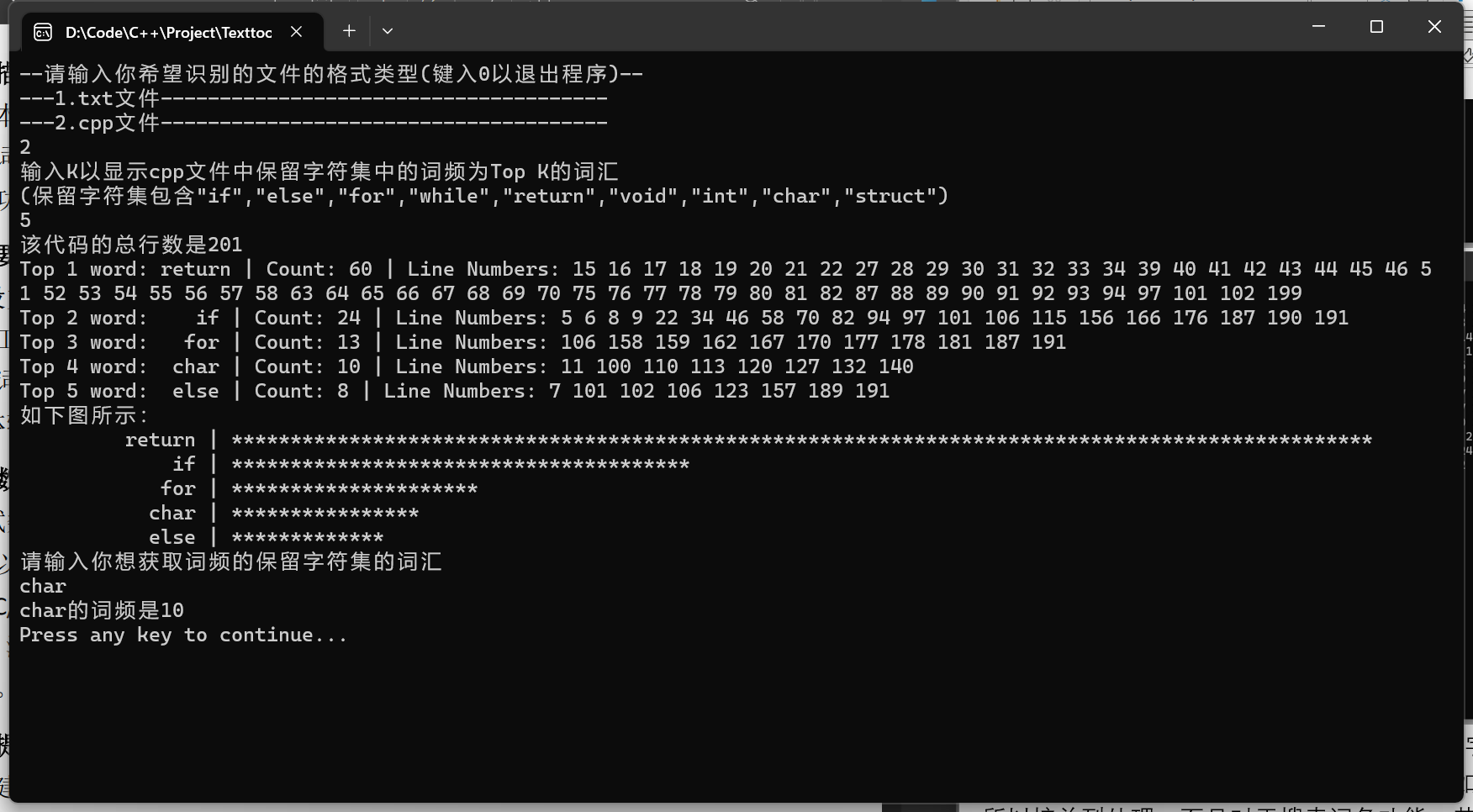
这是一个文本分析工具，可以实现对txt文件或cpp文件的文本分析，具体功能有，统计词频，统计总词数和行数，并以统计图的形式输出。针对txt文件，程序可以方便的统计总词数，并且输出词频为Top-K的单词（其中k可由用户设置）。针对cpp文件，程序可以统计代码的总行数，并且计算其中保留字符集中的词汇出现的次数。程序还允许输入单词以查找词频。

1. 程序运行截图，包括计算功能演示、部分实际运行结果展示、命令行或交互 式界面效果等

程序运行截图如下：  
对于txt文件：  




针对cpp文件：



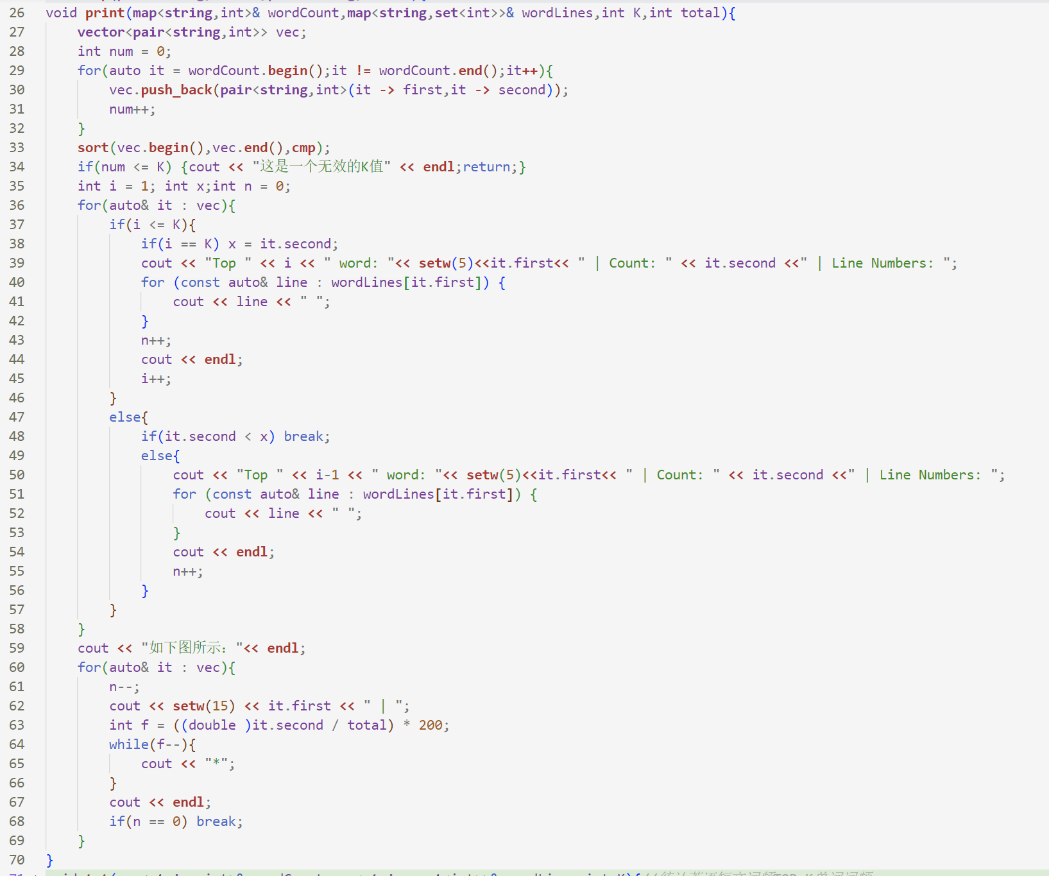
说明：（如果词频出现并列的情况，则会按照字典序排名，但如果是到了排行的最后，则会将并列的情况按照同一排名输出，如图2中的Top11，a和with都为12次，所以按并列处理，而且对于搜索词条功能，若单词没出现，则打印提示）

1. 部分关键代码及其说明





上图为处理txt文件和cpp文件的子函数，通过两个map来储存单词的词频及出现的行数等信息，在函数末尾调用了print函数，来打印输出，其中cpp函数中用到了KMP算法，以实现对保留字符集中字符串的匹配。

以下是print函数的源码

其中第一个循环是将map中的数据导入到一个vector中进行排序。第二个循环实现将词频为Top-k的单词逐行打印输出，并且最后一个循环中会让他们以条形统计图的方式进行输出，其中需要注意的是，条形统计图中以“\*”来体现词频的多少，但并不代表绝对数值，而是，用该单词的词频占总词数的比例后换算得到的，故只能体现相对的多少。

1. 程序运行方式简要说明

程序设置了两个输入文件，分别为input.txt和input.cpp，用户可以将自己想要分析的文本导入到这两个文件中，然后再运行text.cpp进行相应的输入以得到输出。需要注意的是程序会一直重复进行，直到用户键入0以退出。