**西安电子科技大学**

**Java程序设计 课程实验报告**

**实验名称 文本文件字符统计程序**

计算机科学与技术 学院 2103051 班

成 绩

姓名 张平 学号 21030540006

同作者

实验日期 2022 年 05 月 22 日

|  |
| --- |
| 指导教师评语：  指导教师：  年 月 日 |
| **实验报告内容基本要求及参考格式**  一、实验目的  二、实验内容  三、实验过程  四、实验结果分析  五、实验小结（实验过程感受和建议） |

# 实验目的

1. 熟悉File对象的操作，文件信息的获取与测试；
2. 掌握典型的流式输入输出（文件流、缓存流、数据流、标准输入输出流），典型的流接口的使用。
3. 了解java.util.Scanner类以及输入输出的重定向方法。

# 实验内容

1．编写一个程序，程序实现对用户指定的文本文件中的英文字符和字符串的个数进行统计的功能，并将结果根据用户选择输出至结果文件或屏幕。

1) 构建统计类，该类实现对I/O的操纵；实现对文本文件中英文字符、字符串的统计；实现对统计结果的输出。

2) 构建测试类，该类实现与用户的交互，向用户提示操作信息，并接收用户的操作请求。

程序应具有良好的人机交互性能，即：程序应向用户提示功能说明，并可根据用户的功能选择，执行对应的功能，并给出带详细描述信息的最终执行结果。

# 实验过程

1. **实验环境**

操作系统：Windows 11

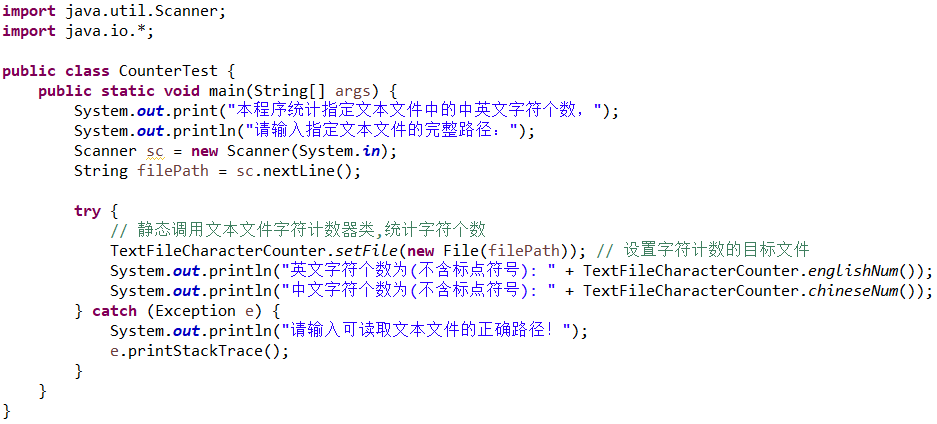
集成开发环境：Eclipse IDE for Enterprise Java and Web Developers (includes Incubating components) 2022-03 (4.23.0)

1. **题目分析**

本实验编写一个文本文件字符统计程序，还要求输出统计结果。为此，这里特别限定：**统计英文字母**（不包含英文标点符号、空格、数字等）**和中文汉字**（不包含中文标点符号、空格、数字等）的个数。至于统计字符串什么的，由于字符串的判断标准太不清晰（空格还是.或。隔开？），这里我不决定实现。

1. **代码说明**

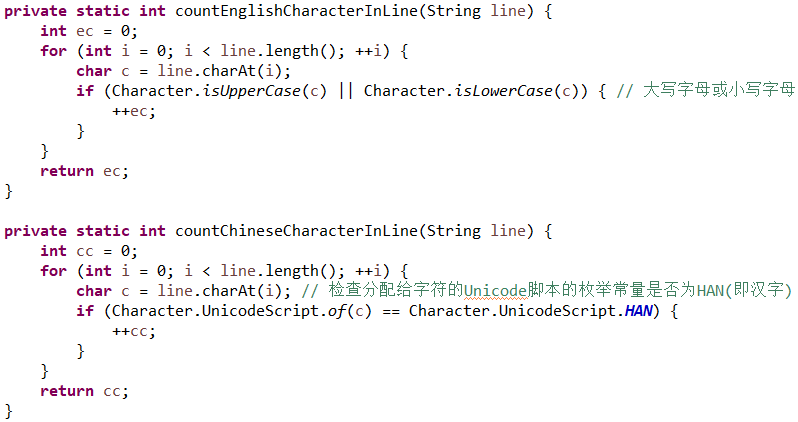
这里简要说明所写的代码。测试类CounterTest如下所示，在main方法中提示用户输入文本文件的完整路径，调用计数器类时还使用try-catch语句处理可能的异常。使用计数器类时，通过调用setFile方法来传入目标文件，调用englishNum方法返回英文字符个数，调用chineseNum方法返回中文字符个数：



由于输入是文本文件，文本文件字符计数器类中使用字符输入流FileReader，为了使读取字符效率更高，代码中还将字符输入流包装为带缓冲的字符输入流BufferedReader。为了自动关闭输入流，代码中还使用了try with resource语法。

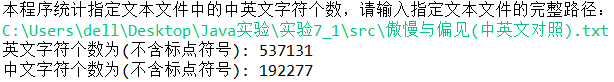


将字符串按行从输入流中读出后，对每行的字符串进行字符统计。此处采用Character类的isUpperCase方法和isLowerCase方法，统计英文字母个数；采用Character.UnicodeScript.of方法，判断是否为汉字。此处不使用正则表达式”[\u4e00-\u9fa5]”，是因为现在这一表达式无法匹配到所有汉字；不使用正则表达式”[\u4e00-\u9fff]”，是因为这实际上是Unicode的中日韩认同表意文字区，且硬编码不符合个人习惯。



# 实验结果分析

完成程序编写后，在Eclipse中点击Run，执行结果如下。傲慢与偏见这一文本文件也放在代码目录中：



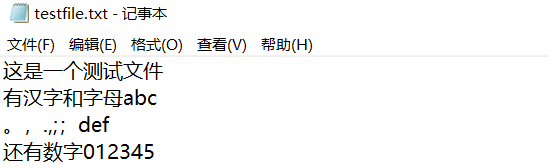
为了验证结果是否正确，在网上搜索了一个统计方式较为接近的程序，全文复制进去后，统计结果如下。程序网址见https://www.a-site.cn/tool/zi/。可见汉字个数是正确的。



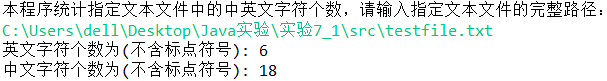
后来换了一个网站https://www.hao353.com/zishutongji，发现字母和汉字个数都是正确的：



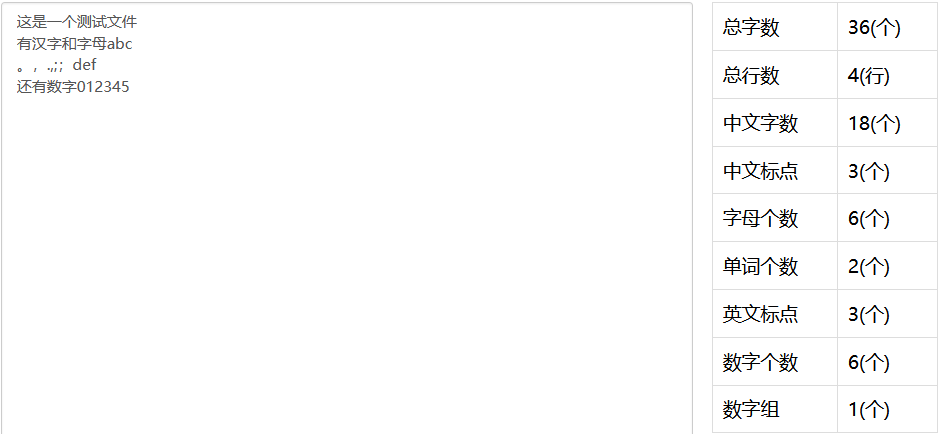
小规模的测试也不可少：



程序运行结果如下：



测试网站的结果如下：



由上可见，实验结果正确。

# 实验小结

本次实验实现了对文本文件的读写、对字符信息的统计，使我在一定程度上熟悉了Java的IO体系和相关使用规范。

在代码中的一个关键点是使用了try with resource语法来自动关闭输入输出流，虽然在本程序中没什么用处（因为GC会自动回收垃圾），但是对某些使用缓冲区的输出流来说，如果不正确关闭流，可能导致文件内容无法及时写入。当然，这一语法适用的资源对象不止输入输出流，可以说是极大方便了代码的编写。

最后在测试中，偶然发现了一个问题：如果我在Eclipse中创建文本文件、输入文本内容，字符统计时结果是正确的；但如果使用记事本来创建文本文件、输入文本内容，字符统计时结果却是错误的。Debug发现是编码问题，记事本创建的文本文件在Eclipse中打开是乱码（它使用的是UFT-8编码），而我的Eclipse使用的是GBK编码。虽然可以在代码中使用语句new InputStreamReader(new FileInputStream(targetFile), “gbk”)来指定数据流的编码方式，但这不能一劳永逸解决所有问题。暂时我还没有找到解决办法。