



华南师范大学

本科学生实验（实践）报告

院系：计算机学院

实验课程：编译原理

实验项目：C++ 源代码单词扫描程序（词法分析）

指导老师：黄煜廉

开课时间：2020 ~ 2021 年度第 1 学期

专 业：计算机科学与技术

班 级：18 级 4 班

姓 名：陈伟卓

学 号：20182131018

华南师范大学教务处

华南师范大学实验报告

学生姓名 陈伟卓 学 号 20182131018
专 业 计算机科学 年级、班级 2018级软工4班
课程名称 编译原理 实验项目 陈伟卓
实验类型 ☐验证 ☐设计 ☒综合 实验时间 2020 年 9 月 20 日
实验指导老师 黄煜廉 实验评分 _____

一、实验内容

1.1 C++源代码扫描程序识别C++记号。C++语言包含了几种类型的记号：标识符，关键字，数（包括整数、浮点数），字符串、注释、特殊符号（分界符）和运算符等。

1.2 打开一个C++源文件，打印出所有以上的记号。

1.3 要求应用程序应为Windows界面。

1.4 选作部分：为了提高C++源程序的可读性，C++程序在书写过程中加入了空行、空格、缩进、注释等。假设你想牺牲可读性，以节省磁盘空间，那么你可以存贮一个删除了所有不必要空格和注释的C++源程序的压缩文本。因此，程序中还应该有这样的压缩功能。

1.5 选作部分：进一步思考或实现——如何进一步实现减小源文件大小的压缩功能。

二、实验目的

熟悉C++编译分词方法，对文本编译预处理有基本的了解。

三、实验文档

实现项目分析：

本次实验需要对C++源代码进行扫描，需要使用windows界面，核心技术设计文件读写和界面展示。本次实验选用Java/Kotlin作为编码语言，使用JavaFX作为界面展示框架。对于CPP文件的读写需要掌握一定的C++语法，在关键点，如数字的识别，符号的识别，特殊符号的识别上可能存在一定的困难。

实验设计：

软件界面方面，主界面左侧提供编辑框和文件拖拽选择两种方式供用户导入待分析的CPP文件。软件右侧展示分析结果。如下图所示：

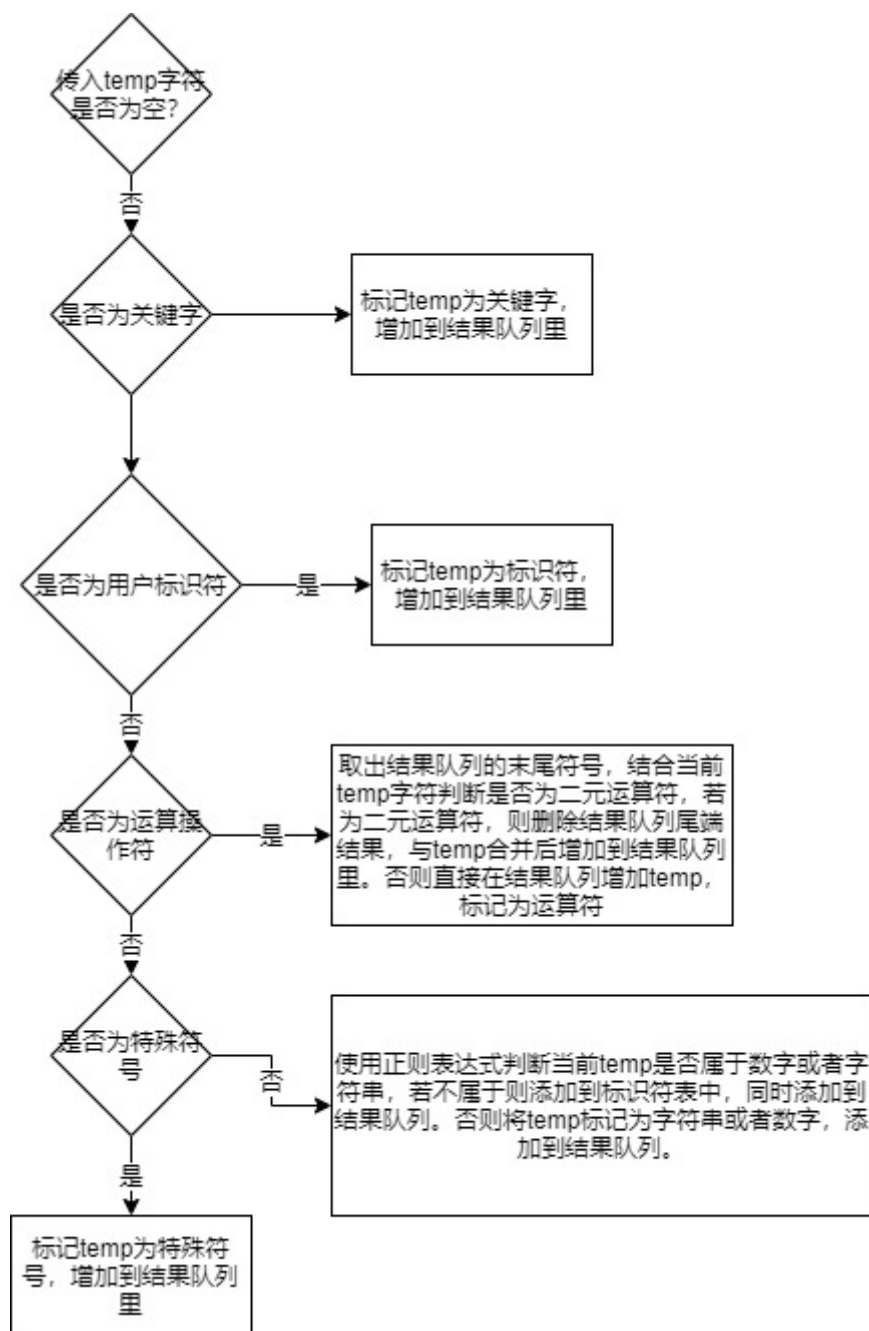


实验实现：

1. 根据要求提供普通CPP文件的分析和识别功能，包括标识符，关键字，数（包括整数、浮点数），字符串、注释、特殊符号（分界符）和运算符等。
2. 用户可牺牲可读性，以节省磁盘空间，提供删除了所有不必要空格和注释的C++源程序的压缩文本的功能。

核心功能实现：

1. 读入文件进行逐个字符分析。
2. 构建四个哈希表，分别储存关键词表(如include, main, const等)，运算符表(如+, ++, >=, >>等)，特殊字符表（如# { } 括号等）和用户定义字符表
3. 结果的储存，构建一个列表，储存分析结果，并进行后续的合并。
4. 总体思路，根据特殊符号和运算符等对单词进行截断，分析截断的结果，存入最终结果列表中。



6. 其中在printType函数中用于判断是否为数字的正则表达式为：

```
[ - ] ? 0 x [ A - F a - f 0 - 9 ] * | [ - ] ? [ 0 - 9 . e ] *
```

用于判断是否为注释的正则表达式为：

```
".*?"
```

7. 压缩文件函数onCompressButtonPress，对目标代码先进行分析，对得到的结果列表进行逐一排查，仅在关键字后追加空格。同时对注释内容不作处理，拼接其余字符串得到压缩后的代码结果。输出到界面上。
8. 采用FXML作为前端界面展示，用Controller类对界面事件进行捕获和处理。采用Maven作为项目管理工作。最后通过JLink生成最小化JRE环境，通过exe4j打包成可执行文件。

软件测试：

具体测试文本于test目录可见

测试内容：

cpp文本内容测试(文本测试.txt)、cpp文件拖拽功能测试(文件测试.cpp)

测试结果：

均能正确识别文件内容。对于像-2，0x8，8e6等的数字能正确识别。对于如“hello”的串能正确识别。对于“==”等符号能正确识别，对于using等关键字能正确识别。具体测试结果于文件（测试结果.txt）中可见。

文件分析测试结果及截图：



代码压缩截图：



四、实验总结（心得体会）

在基本功能方面看似不难，但是对代码字符的状态进行判断是一件相当繁琐的工作。需要提前安排好状态图，程序编写起来才会更快。

五、参考文献

Kenneth C.Louden：《编译原理及实践》.2000.机械工业出版社

