**深 圳 大 学 实 验 报 告**

**课程名称： 编译原理**

**实验项目名称：实验二 词法分析程序设计**

**学院： 计算机与软件学院**

**专业： 计算机科学与技术**

**指导教师： 罗成文**

**报告人：谭嘉豪 学号： 2020152087 班级： 数计班**

**实验时间： 2023年4月13日**

**实验报告提交时间： 2023年4月19日**

**教务部制**

|  |
| --- |
| **一、实验目的与要求**   1. 完成 TINY+词法分析程序的编写及测试（使用提供的测试代码或自己编写测 试文件）； 2. 使用实验所提供的模板撰写实验报告，要求内容详实，有具体的设计描 述、关键的代码片段、及实验结果屏幕截图； 3. 在截止日期前将代码、实验报告、测试文件（如有）等所有实验相关文件 压缩到一个压缩包姓名\_学号\_实验二.rar 上传至 Blackboard。   **二、实验内容**   * TINY 语言的词法由 TINY Syntax.ppt 描述； * TINY 语言的词法分析器由 TINY Scanner.rar 的 C 语言代码实现； * TINY+语言的词法由 TINY+ Syntax.doc 描述。   任务：理解 TINY 语言的词法及词法分析器的实现，并基于该词法分析器， 实现拓展语言 TINY+的词法分析器。  要求：  （1） TINY+词法分析器以TINY+源代码为输入，输出为识别出的token序列；  （2） 词法分析器以最长匹配为原则，例如‘:=’应识别为赋值符号而非单 独的‘：’及‘=’；  （3） Token 以（种别码，属性值）表示，包含以下类型的种别码：  a) KEY 为关键字；  b) SYM 为系统特殊字符；  c) ID 为变量；  d) NUM 为数值常量；  e) STR 为字符串常量。  （4） 识别词法错误。词法分析器可以给出词法错误的行号并打印出对应的 出错消息，主要包含以下类型的词法错误：  a) 非法字符。即不属于 TINY+字母表的字符，比如$就是一个非法字符；  b) 字符串匹配错误，比如右部引号丢失，如‘scanner  c) 注释的右部括号丢失或匹配错误，如 {this is an example  **三、实验步骤及说明**   1. **Tiny源代码及修改**   **要在Tiny源代码上实现Tiny+的拓展，必须理解Tiny源代码中一些必要的代码。**   1. **GLOBALS.H文件**   **利用宏定义定义一些标识符FALSE，TRUE，MAXRESERVED（关键字数量，由于Tiny+有23个关键字，古修改为23）。**    **枚举列表TokenType，用来区分识别出的token类型，包括六种类型：关键字、系统特殊符号、错误类型、文件结束符、标识符、数据类型。**    **在原Tiny代码上添加了一些新的关键字。**    **添加新的数据类型字符串。**    **添加新的系统特殊符号。**    **添加具体的错误类型。**     1. **MAIN.C文件**   **pgm为输入的Tiny或Tiny+源代码文件名或路径，且要求格式为2个参数，即：./main 文件名或文件路径，其中main为编译生成的可执行文件名，可以为自定义为其他。**    **接下来拷贝文件名或路径，以只读的方式打开文件流，如果文件不存在，输出提示信息。**    **不断查找并输出token，直到文件末尾，关闭文件流。**     1. **UTIL.C文件**   **在UTIL.C文件中，只需要修改printToken函数，向其中的swicth语句根据token类型添加case输出对应的（种别码，属性值）或错误类型，一些示例如下。**       1. **SCAN.C文件**   **为StateType枚举列表添加以下新的状态元素。**    **在关键词结构体中添加新的关键词，使每个字符串对应一个TokenType类型。**    **reservedLookup函数用于检查一个标识符是否是一个关键字，如果是关键字，将token类型修改为对应的关键字类型，否则不变。**    **当前位于开始状态时：**   * **遇到一个数字，则进入INNUM状态；** * **遇到一个字母，则进入INID状态；** * **遇到冒号，则进入INASSIGN状态。** * **遇到空格、\t、\n、\r，则跳过。** * **遇到单引号，则进入INSTRING状态。** * **遇到大于号，则进入INGET。** * **遇到小于号，则进入INLET。** * **遇到星号，则进入INPOWER。** * **遇到左花括号，则进入INCOMMENT状态，保存标志设为FALSE。** * **遇到右花括号，则进入DONE状态，保存标志设为FALSE，token类型为ERROR\_COMMENT\_LEFT\_MISSING。**       **遇到其他字符，均变为DONE状态，再根据具体字符设置为对应的ToeknType。**    **当前位于INCOMMENT状态：**   * **遇到文件结束符，说明没有右花括号与其匹配，状态变为DONE状态，错误类型为ERROR\_COMMENT\_RIGHT\_MISSING。** * **再次遇到左花括号，说明发生嵌套错误。状态变为DONE状态，错误类型为ERROR\_COMMENT\_NEST。** * **遇到右花括号，与其左花括号匹配。** * **遇到其他字符，均为注释，忽略。**     **当前位于INASSIGN状态：**   * **一定变为DONE状态。** * **遇到‘=’符号，token类型变为ASSIGN。** * **遇到其他字符发生错误，错误类型为ERROR\_ASSGIN\_SYMBOL，save标记变为FALSE，回退一个字符位置，也就是再次从当前字符开始判断。**     **当前位于INNUM状态：**   * **遇到一个字母，说明发生标识符定义错误，错误类型为ERROR\_ID\_DEFINE。** * **遇到非字母也非数字，说明当前得到一个NUM，toeknType为NUM，save变为FALSE，状态变为DONE，回退一个字符位置，也就是再次从当前字符开始判断。** * **遇到数字，save不变，将其打印。**     **当前位于INID状态：**   * **遇到非字母且非数字，得到一个ID，token类型为ID，状态变为DONE，save变为FALSE，回退一个字符位置，也就是再次从当前字符开始判断。** * **遇到字母或数字，将其打印，可能为标识符的一部分。**     **当前位于INSTRING状态：**   * **遇到单引号，说明得到一个字符串，state变为DONE，token类型为STRING。** * **遇到回车或文件结束符，由于字符串不能跨行，因此产生字符串匹配错误，错误类型为ERROR\_STRING\_SINGLE\_MISSING，save变为FALSE，状态变为DONE，回退一个字符位置，也就是再次从当前字符开始判断。** * **遇到其他字符，可能为字符串的一部分。**     **当前位于INGET状态：**   * **状态变为DONE。** * **遇到‘=’，说明是一个‘>=’，toekn类型变为GET。** * **遇到其他字符，save变为FALSE，回退一个字符位置，也就是再次从当前字符开始判断，toekn类型变为GT。**     **当前位于INLET状态：**   * **状态变为DONE。** * **遇到‘=’，说明是一个‘<=’，toekn类型变为LET。** * **遇到其他字符，save变为FALSE，回退一个字符位置，也就是再次从当前字符开始判断，toekn类型变为LT。**     **当前位于INPOWER状态：**   * **状态变为DONE。** * **遇到‘\*’，说明是一个‘\*\*’，toekn类型变为POWER。** * **遇到其他字符，save变为FALSE，回退一个字符位置，也就是再次从当前字符开始判断，toekn类型变为TIMES。**     **当save=true且tokenString长度不超过最大长度时，将其打印。**  **当前状态为DONE时，需要将tokenString已经为一个token类型，提示判断是否为标识符，进而判断是否为关键字。**    **打印行号及（种别码，属性值）或错误类型，返回token，入token不等于EOF，继续查找token。**     1. **测试代码** 2. **测试Tiny.txt**      1. **测试tiny+1.txt**      1. **测试Tiny+2.txt**      1. **测试tiny+3.txt**   **tiny+3.txt主要为错误测试。**    **四、感想及其他**  **通过本次实验，我完成 TINY+词法分析程序的编写及测试，并深切地体会了词法分析的艰难程度，尽管Tiny+语言的词法要求不如现在的C++，Java等高级语言一般严格，但是对于我来说也极难实现了，这次艰难的实验使我对词法分析有了更深的理解。** |

深圳大学学生实验报告用纸

|  |
| --- |
| 指导教师批阅意见：  成绩评定：  指导教师签字：  2023年 月 日 |
| 备注： |

注：1、报告内的项目或内容设置，可根据实际情况加以调整和补充。

2、教师批改学生实验报告时间应在学生提交实验报告时间后10日内。