**编译原理课程实验报告**

**实验1：词法分析**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 姓名 | 郭茁宁 | | 院系 | | 计算机学院 | | | 学号 | | | 1183710109 | | |
| 任课教师 | | 陈鄞 | | | | 指导教师 | 陈鄞 | | | | | | |
| 实验地点 | | 格物213 | | | | 实验时间 | 2021.4.17 | | | | | | |
| 实验课表现 | | 出勤、表现得分 | |  | | 实验报告  得分 |  | | 实验总分 | | |  | |
| 操作结果得分 | |  | |
| **一、需求分析** | | | | | | | | | | 得分 | | |  |
| 输入源程序，输出单词符号(token)。即：把构成源程序的字符串转换成“等价的”单词(记号)序列，需要实现的功能有：  1.根据词法规则识别及组合单词  2.对数字常数完成识别  3.查填符号表  4.删去空格字符和注释  5.错误检查  具体而言，词法分析器识别了：  标识符（由大小写字母、数字以及下划线组成，但必须以字母或者下划线开头）  关键字（①类型关键字：整型、浮点型、布尔型、结构体；②分支结构中的if和else；③循环结构中的do和while；）  运算符（①算术运算符；②关系运算符**；**③逻辑运算；）  界符（①用于赋值语句的界符，如“=”；②用于句子结尾的界符，如“；”；  常数（无符号整数和浮点数等）  注释（/\*……\*/形式） | | | | | | | | | | | | | |
| **二、文法设计** | | | | | | | | | | 得分 | | |  |
| **1.标识符id**  ①digit → 0|1|2|…|9  ②letter\_ → A|B|…|Z |a | b|…|z|  ③id → [letter\_](letter|\_|digit) \*    **2.无符号数**  ①digit → 0|1|2|…|9  ②digits → digit digit\*  ③ number → digits (.digits)? (E [+-]? digits)?    **3.运算符**  Operator → +|-|\*|/|++|--|=|==|&&| || | ! | & | | | ~ | > | < | >= | <= | > | <  **4.界符**  Delimiters → (|)|{|}|;|”  **5.注释**  Annotation→/\* | \*/ | | | | | | | | | | | | | |
| **三、系统设计** | | | | | | | | | | 得分 | | |  |
| 要求：分为系统概要设计和系统详细设计。   1. **系统概要设计：**     **读入源程序，经过词法分析器，输入单词序列表。**    **种别码定义如下：**   |  |  | | --- | --- | | \_ID(标识符) | 0 | | \_constString(字符串常量) | 1 | | \_constFloat(浮点数常量) | 2 | | \_constInteger(整数常量) | 3 | | for | 4 | | do | 5 | | if | 6 | | else | 7 | | int | 8 | | char | 9 | | float | 10 | | String | 11 | | while | 12 | | struct | 13 | | boolean | 14 | | true | 15 | | false | 16 | | “ | 17 | | + | 18 | | - | 19 | | \* | 20 | | / | 21 | | = | 22 | | ++ | 23 | | -- | 24 | | & | 25 | | | | 26 | | ~ | 27 | | != | 28 | | == | 29 | | && | 30 | | || | 31 | | ! | 32 | |  | 33 | | < | 34 | | > | 35 | | <= | 36 | | >= | 37 | | ( | 38 | | ) | 39 | | { | 40 | | } | 41 | | ; | 42 |   **（2）系统详细设计：**   * **核心数据结构的设计**   1.Tool类：主要用于辅助程序实现，包括判断一个字符是数字还是字母等等信息。  2.DFA类：主要是用于构造识别单词和无符号常数的DFA。   * **主要功能函数说明**   1.DFA类  ①public static String getDigit(String input)  识别数字的DFA  ②public static String getWord(String input)  识别单词的DFA  **2.Tool类**  ①public static boolean isunderline(char x)  判断字符是否是下划线  ②public static boolean isletter(char x)  判断字符是否是字母  ③public static boolean isdigit(char x)  判断字符是否是数字  **3.Analysis包**  ①private static void lineAnalysis(String line,int LineNumber)  按行进行词法分析  ②private static void classifyType(char x, String line, int lineNumber)  判断单词的类别  ③ public static void handleAnnotation(String line, int lineNumber)  处理注释的函数   * **程序核心部分的程序流程图** | | | | | | | | | | | | | |
| **四、系统实现及结果分析** | | | | | | | | | | 得分 | | |  |
| 要求：对如下内容展开描述。   1. **系统实现过程中遇到的问题；**   **①DFA的构建问题：**  首先就是如果读到了首字符是字母或者是下划线，那么需要调用DFA去识别做一个最长的前缀匹配，来得到完整的字符串来做标识符或者是关键字。  如果首字符是数字，那么调用DFA去将其数字部分全部识别，那么就需要构建一个识别整数和浮点数的DFA。这一部分不只是需要写程序的能力，还需要先构建自动机再在程序中实现。具体在程序中，我使用的是循环加上switch\_case语句来做的，如果读入下一个字符，利用判断语句进到下一个状态，循环执行，直到遇到自动机不能识别的字符，自动机结束。  **②如何根据输入的程序，识别各类单词？**  为了能够识别各类单词，使其进入正确的DFA转换表进行识别，我采用逐个字符读入的方式，根据读入的起始字符进入相应的处理程序。如果起始字符是数字或者字母或者下划线，那么进入相应的DFA进行处理。  值得一提的是算符，由于我定义的算符有“+”和“++”，如果只看到“+”，还需要继续往下查找看是否有“++”符号，因此看到一个字符需要再向前搜索下一个字符才能判断。   1. **针对某测试程序输出其词法分析结果；**   **测试程序如下：**    **输出的结果是：**      进行分析，  **（a）**如最后一个符号“;”，能够成功识别。并且输出正确的种别码是37，与定义的种别码表相符合。  **（b）**对于标识符a，在符号表的0号位置。对于标识符b，在符号表的1号位置。因此输出的是<0,0>和<0,1>。其中，由于是标识符，所以种别码是0,属性值是符号表的位置，分别是0和1。  **（c）**发现，注释的部分全部被删除掉了。与处理相符合。  **（d）**如果注释不封闭，那么程序会有提示，并且注释以后的部分全部去除掉。   1. **输出针对此测试程序对应的词法错误报告；**     其中，在第六行有定义 String s=”1dvfg;，这里是有错误的，因为字符串引号不封闭。  并且，在第九行有定义 float b=1.23.5，这里由于有两个“.”，那么这个是错误的，不符合浮点数的定义。   1. **对实验结果进行分析。**   本词法分析器读入源程序后，对于各类正确单词能够成功地进行识别，并进行输出Token序列、单词类别、相应的DFA转换表。  对于错误的输入，本词法分析器能够进行成功的识别，并在错误显示区进行输出，同时输出对应的行号，便与使用者清楚地知道错误的位置和原因。  其中，关于注释的识别，成功的进行了注释的识别，而且成功的识别了多行的注释，对于注释未封闭的情况也能成功识别并且输出错误信息。 | | | | | | | | | | | | | |
| 指导教师评语：  日期： | | | | | | | | | | | | | |