**《编译原理》课程**

**实验报告**

**题目 词法分析**

**专业 软件工程**

**班级 10065502**

**学号 1006840639**

**姓名 杨达一**

**指导教师 朱保平**

**南京理工大学计算机学院**

**2012年4月**

**实验1：词法分析**

|  |
| --- |
| **一、实验目的** |
| 1.巩固对词法分析的基本功能和原理的认识。  2.能够应用自动机的知识进行词法分析。  3.理解并处理词法分析中的异常和错误。 |
| **二、实验内容** |
| 要求：对如下工作进行展开描述  **(1)** 给出语言的词法规则描述  •关键字、标识符  关键字：常用的c++关键字   |  |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | --- | | 1 | bool | 2 | true | 3 | false | 4 | char | | 5 | int | 6 | long | 7 | short | 8 | float | | 9 | double | 10 | signed | 11 | unsigned | 12 | void | | 13 | const | 14 | sizeof | 15 | if | 16 | else | | 17 | switch | 18 | case | 19 | default | 20 | while | | 21 | for | 22 | break | 23 | continue | 24 | return | | 25 | goto | 26 | static | 27 | extern | 28 | struct | | 29 | union | 30 | enum | 31 | typedef | 32 | new | | 33 | delete | 34 | class | 35 | private | 36 | public | | 37 | protected | 38 | this | 39 | operator | 40 | friend | | 41 | namespace | 42 | using |  |  |  |  |   预编译指令：   |  |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | --- | | 1 | #include | 2 | #define | 3 | #undef | 4 | #pragma | | 5 | #import | 6 | #error | 7 | #line | 8 | #ifdef | | 9 | #ifndef | 10 | #if | 11 | #else | 12 | #elif | | 13 | #endif |   标识符：以字母或下划线开始，由字母、数字及下划线组成的字符序列，关键字不能作为关键字，严格区分大小写。  • 整型常量，浮点常量，字符常量，字符串常量  整型常量：十进制，以1~9开头，由0~9组成的数；八进制，以0开头，由0~7组成的数，十六进制，以0x或0X开头，由0~9、A~F或a~f组成的数。  浮点常量：小数表示法，由整数和小数两部分组成，中间用小数点隔开，在小数点左边或右边有数字；指数表示法，由尾数和指数两部分组成，中间用e或E隔开，指数只能是整数，可以有+，-表示正负。只允许十进制.  字符常量：用单引号括起的单个字符，允许如下转义：**\n \t \\ \’ \”**  字符串常量：用双引号括起的0个或多个字符，只能写在同一行。  • 运算符  **. -> ++ -- ! ~ \* / % + - << >> < <= > >= == != & ^ | = += -= \*= /= %= <<= >>= &= ^= |= ,**  • 括号  **( ) [ ] { }**  • 界符  **; :**  • 注释  //单行注释  /\*多行注释  多行注释\*/  **(2)** 针对这种单词的状态转换图和程序框图  图A，图B，图C为单词转化的DFA，分别主要描述了数字，字符（串），标点符号的转化规则。(绿色的为结束状态)              **(3)** 核心数据结构的设计  •查询到的单词类型：  NUMBER,IDENTIFIER,KEYWORD,OPERATOR,UNKNOWN,STRING,PUNCTUATION,ANNOTATION,PARENTHESES  **(4)** 附加功能  整理格式，补全括号 |
| **三、实验结果** |
| 要求：将实验获得的结果进行描述，基本内容包括：   1. 针对某测试程序输出其词法分析结果； 2. 输出针对此测试程序对应的词法错误报告； 3. 输出针对此测试程序对应的符号表。   注：其中的测试样例自行产生。 |
| **四、实验中遇到的问题总结** |
| 要求：主要阐述两方面的问题   1. 实验过程中遇到的问题如何解决的？   算法上面没有遇到什么太大的问题，整个算法难度不大，但是状态转换上比较烦琐，在代码上会经常出现一些细节问题，经过测试可以检查出来并且改正。  （二）思考题的思考与分析  创建类DFA，维护一个整数Status，表示当前状态，用函数dfa.run(intStatus,char c)来处理状态转换，当接受到一个字符时，将Status做出相应的改变或是抛出异常。状态的表示和状态之间的转换以及判断是否为终态，完全按照之前给出的状态转换图来实现。这样做的好处在于方便实现，根据状态图也很好理解。  关于单词查找，可以想到以下改进：1、用正则表达式去匹配单词，代码量小，不易出错，可维护，效率高等诸多优点，但这完全抛弃了编译课上学到的东西（其实这是我最先想到的实现方法，但是不敢用）；3、将DFA最小化再进行编程 |
| **五、实验体会** |
| 包括收获、心得体会、存在的问题及解决问题的方法、建议等  存在问题：   1. 数字读入时无法区分正负数，后来明白过来，在不考虑语义的前提下，无法区分”-“到底是符号还是减号。 2. 一个标识符在不同的作用域中出现时，应该在符号表中拥有不同的标号，但是我的程序无法判断。   心得体会：  通过这段时间的课程设计，我理解了词法分析在编译过程中的作用，以及它是如何工作的，感谢老师对我的指导。 |