词法分析器的设计

一、设计思路

词法分析器的输入是待编译程序的源代码。输出是源代码中的所有符号（保持源代码中的顺序）。

输入一般为文件。输出可以有多种形式。譬如：直接输出到文件、输出到控制台，或者暂不输出，等到语法分析要用时，直接给出下一个单词的符号即可。

故设计三个类：

1、词法分析器类，表示一个词法分析器。可以打开源代码文件，读入并分析文件，产生符号序列，并进行输出。

2、符号类，表示一个最小的符号单元。其属性为type和value。表示符号的类型和值。

3、值类，表示符号类中值的类型。这里可以用共用体，也可以用结构体。因为符号类中不同的type对应的值的类型也不同，所以需要不同的解释方式。这里考虑到共用体并不一定能节省空间，故采用结构体，结构体中定义三个变量，分别表示整型、字符型和字符串型的值。

二、符号的类型设计

符号总的来说可以分为以下四类：

1、保留字

2、标识符

3、常数

4、分界符或操作符

对标识符来说，通常将所有用户自定义的标识符归为一类。

对常数来说，通常将不同类型的常数各归一类，例如整型、字符型、实型等

对保留字、分界符和操作符来说，可以分成大类，再用值进行区分，也可以一符一类。由于给定文法中常常由非终结符直接推出"begin"、"if"等保留字和"+"、"-"等操作符和";"等分界符，故这里采用一符一类的方法。

所以符号的类型有：

【保留字】CONST VAR ARRAY RW\_INTEGER RW\_CHAR PROCEDURE FUNCTION IF THEN ELSE CASE OF FOR DOWNTO TO DO BEGIN END READ WRITE

【标识符】IDENTIFIER

【常数】CONST\_INTEGER CONST\_CHAR

【分界符与操作符】PERIOD COMMA SEMICOLON SINGLE\_QUOTATION DOUBLE\_QUOTATION COLON ASSIGN LEFT\_BRACKET RIGHT\_BRACKET LEFT\_PARA RIGHT\_PARA ADD MINUS MUL DIV LT LTE GT GTE EQ NEQ

三、状态图设计

