

# 编译 PJ1: 词法分析

15307130012 罗辉翔    15307130237 冯乐浩    15307130002 冯望

## 1. 实验目的

- (1) 学习使用 flex 这一在文本上做正则匹配的工具;
- (2) 对程序的编译过程有一个基本的认识。

## 2. 实现细节

### (1) 程序的编译过程

参照 demo 文件可知, 原始文件主要有四个:

- 1) .lex 文件中记录了正则表达式的匹配规则;
- 2) .cpp 文件控制了程序在扫面文本之后的输出结果;
- 3) .h 文件记录了程序中定义的常量;
- 4) Makefile 文件中定义了整个文件的编译过程。

我将 demo 的文件略作修改, 删去了 demo.y 文件, 并将编译器设置为 g++ 11.

### (2) 指定匹配规则

#### 1) 注释

注释的定义: 以 (\* 字符开头, 以\*) 字符结尾, 且中间不存在\*) 的字符串。

因为这与 c 语言中的/\*\*/注释规则类似, 我就参照了它的正则匹配规则:

"^\*([\\^\*])(\\^\*)\*(\\^\*)\*/", 将其中的"/"字符改为("("或者")"字符即可。

匹配注释时, 若注释末尾没有\*/, 则需要记录并报错。这时只需要删去上述正则表达式的末尾部分"\*/"即可。

#### 2) 整数

整数的定义: 属于 0 到  $2^{32}-1$  这个范围的所有整数 (闭区间)。

匹配整数时, 如果出现超出范围的 (负整数或者大于 2147483647 的整数), 则要报错。

#### 3) 实数

参照实数的定义, PCAT 语言中, 所有的实数都不能为负。所以如果出现了负实数, 也需要报错。

#### 4) 保留字, 操作符, 终止符

参照 PCAT 语法规则, 将所有保留字匹配出来即可。

#### 5) 助记符

助记符的定义: 长度不超过 255 个字符, 且不为保留字的字符串。

这里我只实现了前半部分，即长度不超过 255 的异常检测；后半部分的检测我没有想出来（我认为这属于语法分析的范畴了…）T\_T。

#### 6) 字符串

字符串是以双引号开头和结尾，中间不包括其他双引号、换行符和制表符的串，所以需要  
对字符串是否定义完整，以及字符串中是否包含了其他字符做检测。

#### 7) ASCII 码中无法打印的字符

这里我们将\x09，即制表符\t 区别对待，则无法打印的字符有\x00-\x08,\x0A-\x1F,\x7F.

### (3) 输出 token 的位置

只需设置两个计数器 row\_index 和 col\_index，分别用来统计当前遍历到的行数和列数即可。需要注意的是，空白字符（空格，换行符和制表符）在程序中虽然出现，但是并无意义，所以在程序末尾无需打印出来，只需要计数即可。

### (4) 程序错误控制

我采取的方式是将各类错误分别检测出来（通过 switch），然后存在一个变量中，最后打印的思路。我认为，如果能将各类错误内容预先定义好，然后根据报错的序号去访问，效率会提升很多。（然而我懒……）

### (5) 样例测试效果

#### 1)test01.pcat

```
luohuixiang@bengio-WorkStation:~/Compile/dem
(行号 列号)      种类      详情
(  4,  1)      保留字      PROGRAM
(  4,  9)      保留字      IS
(  5,  5)      保留字      VAR
(  5,  9)      助记符      i
(  5, 10)      终止符      ,
(  5, 12)      助记符      j
(  5, 14)      终止符      :
(  5, 16)      助记符      INTEGER
(  5, 24)      终止符      :=
(  5, 27)      整数      1
(  5, 28)      终止符      ;
(  6,  5)      保留字      VAR
(  6,  9)      助记符      x
(  6, 11)      终止符      :
(  6, 13)      助记符      REAL
(  6, 18)      终止符      :=
(  6, 21)      实数      2.0
(  6, 24)      终止符      ;
(  7,  5)      保留字      VAR
(  7,  9)      助记符      y
(  7, 11)      终止符      :
(  7, 13)      助记符      REAL
(  7, 18)      终止符      :=
(  7, 21)      实数      3.0
(  7, 24)      终止符      ;
(  8,  1)      保留字      BEGIN
(  9,  5)      保留字      WRITE
```

```
(  9, 11)      终止符      (
(  9, 12)      字符串      "i = "
(  9, 18)      终止符      ,
(  9, 20)      助记符      i
(  9, 21)      终止符      ,
(  9, 23)      字符串      ", j = "
(  9, 31)      终止符      ,
(  9, 33)      助记符      j
(  9, 34)      终止符      )
(  9, 35)      终止符      ;
( 10,  5)      保留字      WRITE
( 10, 11)      终止符      (
( 10, 12)      字符串      "x = "
( 10, 18)      终止符      ,
( 10, 20)      助记符      x
( 10, 21)      终止符      ,
( 10, 23)      字符串      ", y = "
( 10, 31)      终止符      ,
( 10, 33)      助记符      y
( 10, 34)      终止符      )
( 10, 35)      终止符      ;
( 11,  1)      保留字      END
( 11,  4)      终止符      ;
符号数: 50      异常数: 0
```

[illegible]

项目的代码编写部分和报告撰写部分由罗辉翔完成，冯乐浩与冯望全程参与讨论并督促罗辉翔完成了该项目。