**编译规格文档**

**撰写人：罗溥晗**

**1.数据结构**

1.1概述

编译部分将会按行读取源代码，逐行解析代码并且最后返回给执行部分一个语句列表。这个语句数组里的每个元素代表了一行代码执行时所必需的信息以及执行完的跳转信息。

1.2语句列表

该列表的定义是ArrayList<Statement\_AST>，列表中元素的类型是Statement\_AST，将在下方定义。列表的第一项置为空，这样可以使列表的下标和代码的行数直接对应。

1.3语句AST

语句AST是一个类，类名为Statement\_AST。属性和方法如下方代码所示。Pythy中有四种语句，分别是赋值语句（y=x+1），分支语句（if），循环语句（while）和空语句。并不是所有语句都用到了所有属性。其中type、line和next是共有属性。type表明该语句的类型，line指示了当前语句位于代码的哪一行，next指示了下一行的行号。空语句只有共有属性。赋值语句有left和right两个属性，分别代表赋值语句的左值和右值，其中语句的左边一定是一个变量名，右边则是一个表达式（将在1.4定义）。分支和循环语句有right、in和out三个属性，这是因为这两条语句会打乱程序的顺序执行，开启一段特殊的程序段。in属性表示该程序段的开始行号（通常就是line+1，也就是next），out属性表示如果分支或循环条件不成立，程序应当跳转到的行号（无法在解析这条语句时直接确定out的值，将在1.5节讨论），right属性用来存储分支或循环条件。语句AST还有一个copy方法，表示将自己复制一遍并返回。

class Statement\_AST

{

String type;//= if while null......

String left;

Expression\_AST right;

int line;

int next;

int in;

int out;

Statement\_AST copy()

{

Statement\_AST y = new Statement\_AST();

y.type = this.type;

y.left = this.left;

y.right = this.right;

y.line = this.line;

y.next = this.next;

y.in = this.in;

y.out = this.out;

return y;

}

}

1.4表达式AST

表达式AST是一个类，类名为Expression\_AST。属性和方法如下方代码所示。Pythy中有八种表达式，用公共属性type区分。四种运算表达式——加减乘除，和四种常变量表达式（我们认为一个常变量也是一种特殊的表达式）——变量（name），整型（int），浮点型（float）和布尔型（boolean）。和语句AST一样，并不是所有表达式都用到了所有属性。除了公共属性之外，运算表达式只用到了left和right两个属性，分别代表运算的左值和右值（采用最左推导？还是最右？，即始终将最左单项展开为常变量）。常变量表达式用到了type属性以及一个和type对应的value属性。表达式AST的copy方法同1.3。

class Expression\_AST

{

String type;//name int float boolean + - \* / > <......

Expression\_AST left;

Expression\_AST right;

int value\_int;

float value\_float;

boolean value\_boolean;

String value\_name;

Expression\_AST copy()

{

Expression\_AST y = new Expression\_AST()

y.type = this.type;

y.left = this.left;

y.right = this.right;

y.value\_int = this.value\_int;

y.value\_float = this.value\_float;

y.value\_boolean = this.value\_boolean;

y.value\_name = this.value\_name;

return y;

}

}

1.5处理缩进

不同于C、JAVA等语言使用大括号区分层级，Pythy采用的是tab缩进来区分层级。我们认为该行首有几个tab缩进，该行就处于第几级（设为level方便后续讨论）。在处理缩进时，我们首先要引入一些全局变量。current表示当前行的level，last表示上一行的level。在1.3节时我们提到，语句的out值无法在解析当前语句时确定，这是因为我们不知道if或while结束在什么地方。因此我们还需要引入一个栈暂时存放尚未填入out值的语句。如果发现了level上升（或是直接根据if/while来压栈？似乎都可以），则压栈。如果发现了level下降，则出栈并将出栈语句的out设置为当前行号。需要注意的是，level可能不止一次减1，也可能下降很多，因此我们规定下降几级就退栈几次（退栈次数=last-current）。同时，如果读到文件结束符，则认为该处level=1，全部出栈。具体算法将在后文给出。

**2.编译动作**

2.1概述

Pythy语言中有4种类型的语句，分别是分支语句、循环语句、赋值语句和空语句。我们依次读取每行代码，根据语句中的一些特征来判断具体是哪一种语句从而执行不同的编译动作。

2.2分支语句

当词法分析器扫描到当前行除缩进外的首个token为“if”时，判定该行为分支语句。将type字段赋值为“if”，line字段为当前行号，next字段为下一行行号，in字段为下一行行号。由于out字段暂时无法确定，此时需要将该行行号压栈表示尚未处理完成，后续将细述out字段的计算算法。同时将分支条件表达式存储进right字段。

2.3循环语句

当词法分析器扫描到当前行除缩进外的首个token为“while”时，判定该行为循环语句。将type字段赋值为“while”，line字段为当前行号，next字段为下一行行号，in字段为下一行行号。由于out字段暂时无法确定，此时需要将该行行号压栈表示尚未处理完成。同时将循环条件表达式存储进right字段。

2.4赋值语句

当词法分析器扫描到当前行除缩进外的首个token不为“if”也不为“while”时，判定该行为赋值语句。将type字段赋值为“assignment”，line字段为当前行号，next字段为下一行行号，left字段为词法分析器扫描到的第一个token，right字段为扫描到“=”后的表达式。

2.5处理缩进及计算out字段的回填算法

**描述：**在进行编译时，称遇到一个tab为级别上升，反之回退一个tab为级别下降。在级别上升时，采用将if、while语句压栈。而该算法用于求解在遇到

n个级别下降时，各个语句AST的out属性的求法。

**输入：**last为级别下降前一条语句，n为下降的级数,比如下面的程序中，last为语句x = 1, n为2

while True:

if True:

x = 1

**算法描述**

令集合set <- {last}

while n >= 0, do

从语句栈中弹出一个语句，记作x

n <- n - 1

*// n*相当于还有几个语句

if x 是"if"类型, then

将 x 加入到集合 set 中

*//*因为*x*的*out*和集合*set*中的*out*或*next*一样

if n == 0, then

fill(set, 当前的行号)

*//*相当于后面有一个和*if*同级别的语句，递归出口。

end if

else if x 是 "while"类型, then

fill(set, x语句的行号)

*//while*中最后一句的*out*或*next*是确定的，都是*while*的行号

if n == 0 then

令x的 out 为当前的行号

else

另 set 为新的空集合，并将 x 加入 set

end if

end if

过程：fill(set, m)：

for set中每一个语句x，do

如果x是if或while，则令x的out为m

否则令x的next为m

end for