

作业二：词法分析与语法分析

唐亚周 519021910804

提交要求：作业文档使用PDF格式提交。问题1的代码命名为p1.java/py/cpp。问题2的变量打印程序代码命名为p2_print.java/py/cpp，示例代码命名为p2_sample.java/py/cpp。将上述一个文档和三个源代码文件打包为RAR/ZIP提交。项目工程文件、第三方库、编译结果等均不要提交。

本次作业涉及语法分析的编程实验，请选择Java/Python/C++任意一种语言进行编程。

一、分别为下列字符串集合**编写正则表达式**，并**阐述构造思路**。同时，对于每个正则表达式，分别构造**三个正例**（可匹配的字符串）和**三个反例**（不可匹配的字符串）。**编写相应的代码**进行验证。

1. 合法的C语言变量名；
2. 大驼峰命名法（UpperCamelCase）所命名的字符串，英文单词之间或字符串最后可能穿插一个或多个数字（如Img2Col, Pcg64）；
3. 使用科学计数法表示的C语言浮点数（如1e-5, -.4e9，不考虑尾数精度或指数范围的限制）；

答：分析如下：

1. 合法的C语言变量名（不考虑保留字）：以大写字母/小写字母/下划线开头，且只包含大写字母/小写字母/下划线/数字。

正则表达式为：`[a-zA-Z_][a-zA-Z0-9_]*`

2. 大驼峰命名法（UpperCamelCase）所命名的字符串：有1个或多个单词；每个单词的第一个字母为大写，后面的字母（0个或多个）为小写，再跟着0个或多个数字。

正则表达式为：`([A-Z][a-z]*[0-9]*)+`

3. 使用科学计数法表示的C语言浮点数：形式为 `AeB`，其中 `A` 等于 `C` 跟着可能有的 `D`；`C` 为可能有的 `-` 号加上0个或多个数字，`D` 为 `.` 号加上1个或多个数字；`B` 为可能有的 `-` 号加上1个或多个数字。

正则表达式为：`-?[0-9]*([.][0-9]+)?e-?[0-9]+`

具体代码与测试样例见 `p1.py`。

二、编写一个变量打印程序，该程序接受一段**源代码文本**作为输入，使用**相应工具**对代码进行语法分析得到语法树。对语法树进行**遍历**，打印所有全局变量、以及每个函数内所有局部变量的名称至标准输出（stdout），每个变量在打印结果中仅出现一次。

然后编写一段**示例代码**，代码中必须包含：

1. 至少一个全局变量（Java语言用类的静态字段代替）；
2. 至少一个函数（Java语言用类的静态方法代替）；
3. 每个函数中必须定义至少一个局部变量。

将上述示例代码输入变量统计程序中，**给出打印结果**。

答：这里对 Python 的 AST 库中的 `NodeVisitor` 类进行继承，并且修改其 `visit_Assign` 函数。这是因为 Python 中每个变量的定义都会有一个赋值语句。遍历赋值语句的 `targets` 列表并将其所有类型为 `ast.Name` 的 target 的 id 加入至存储最终结果的 `variable_list` 即可（注意不要有重复元素）。具体代码见

`p2_print.py`。