**编译原理实践第5次课**

**（PLY入门）**

**一、实验名称**

利用PLY识别词法单元

**二、实验目的**

识别词法单元

**三、实验内容**

编写程序，能够把如下程序（prog.txt）中的词法单元都识别出来

**四、实验代码**

# ------------------------------------------------------------

# main.py

#

# tokenizer for a simple expression evaluator for

# numbers and +,-,\*,/,

# ------------------------------------------------------------

import ply.lex as lex

reserved = {

'main' : 'MAIN',

'int' : 'INT',

'if' : 'IF',

'while' : 'WHILE',

'printf' : 'PRINTF',

}

# List of token names. This is always required

tokens = [

'ID',

'NUMBER',

'PLUS',

'MINUS',

'TIMES',

'DIVIDE',

'LPAREN',

'RPAREN',

'OPENCURLY',

'CLOSECURLY',

'SEMICOLON',

'QUOTATIONMARK',

'LT',

'EQUAL',

'POINT',

'SENTENCE',

] + list(reserved.values())

# Regular expression rules for simple tokens

t\_PLUS = r'\+'

t\_MINUS = r'-'

t\_TIMES = r'\\*'

t\_DIVIDE = r'/'

t\_LPAREN = r'\('

t\_RPAREN = r'\)'

t\_OPENCURLY = r'\{'

t\_CLOSECURLY = r'\}'

t\_SEMICOLON = r'\;'

t\_QUOTATIONMARK = r'\"'

t\_LT = r'\<'

t\_EQUAL = r'\='

t\_POINT = r'\.'

t\_SENTENCE = r'".+?"' #\"(\\.|[^\\"])\*\"

# A regular expression rule with some action code

def t\_NUMBER(t):

r'\d+'

t.value = int(t.value)

return t

# Define a rule so we can track line numbers

def t\_newline(t):

r'\n+'

t.lexer.lineno += len(t.value)

def t\_ID(t):

r'[a-zA-Z\_][a-zA-Z\_0-9]\*'

t.type = reserved.get(t.value, 'ID') # Check for reserved words

return t

# A string containing ignored characters (spaces and tabs)

t\_ignore = ' \t'

# Error handling rule

def t\_error(t):

print "Illegal character '%s'" % t.value[0]

t.lexer.skip(1)

def find\_column(input, token):

last\_cr = input.rfind('\n', 0, token.lexpos)

if last\_cr < 0:

last\_cr = 0

column = (token.lexpos - last\_cr) + 1

return column

#literals = ['+', '-', '\*', '/', '{', '}', ';', '"', '.', '=']

# Build the lexer

lexer = lex.lex()

# Test it out

data = '''

asd 3 + 4 \* 10

+ -20 \*2

'''

f = open('prog.txt', 'r')

data = ''

for lines in f.readlines():

data += str(lines)

f.close()

# Give the lexer some input

lexer.input(data)

# Tokenize

while True:

tok = lexer.token()

if not tok:

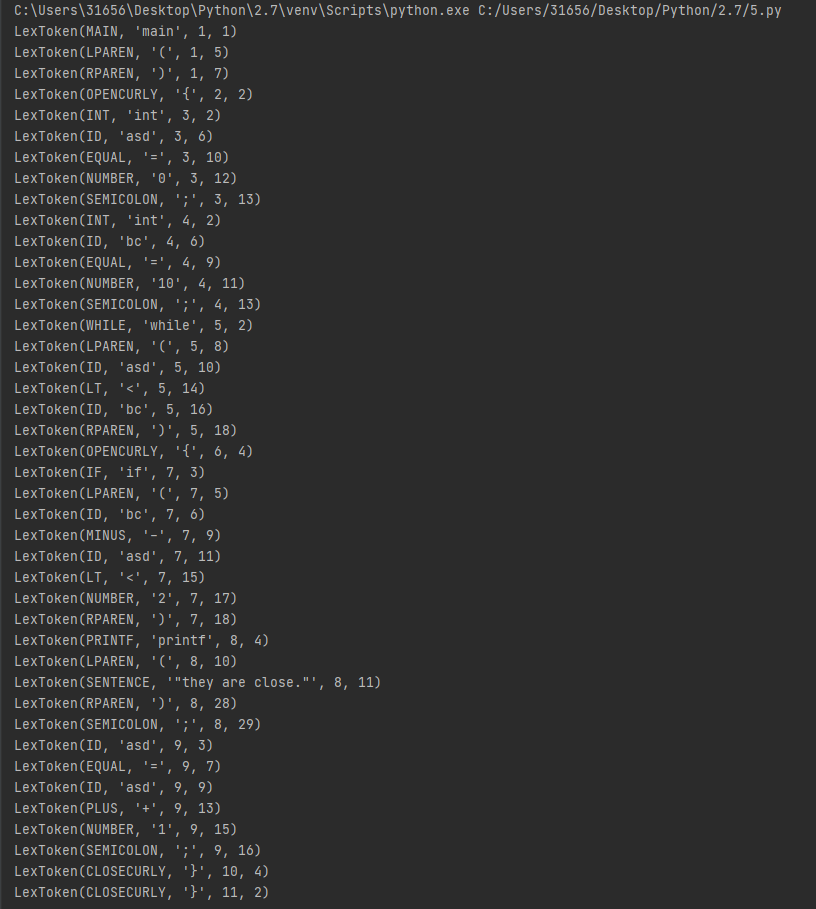
break # No more input

#print tok

print 'LexToken(' + str(tok.type) + ', \'' + str(tok.value) + '\', ' + str(tok.lineno) + ', ' + str(find\_column(data, tok)) + ')'

#print tok.type, tok.value, tok.lineno, tok.lexpos - pos

**五、运行结果**



**六、实验总结**

实验中，成功利用PLY包识别出了’prog.txt’中的词法单元，但是并不能对变量进行分类，只能识别类别。