Лексичний аналіз методом діаграми станів Варіант № 12

Арифметика: цілі та дійсні числа, основні чотири арифметичні операції (додавання, віднімання, ділення та множення), піднесення до степеня(правоасоціативна операція), дужки

Особливості: унарний мінус

Інструкція повторення: for <iд>=<вираз> by <вираз> while<відношення> do <оператор>

Інструкція розгалуження: if <відношення> {<список операторів>}

Повна граматика мови КК

```
Program = start ProgName ';' StatementList
ProgName = Ident
Алфавіт
Letter = 'a' | 'b' | 'c' | 'd' | 'e' | 'f' | 'g' | 'h' | 'i' | 'j' | 'k' | 'l' | 'm' | 'n' | 'o' | 'p' | 'q' | 'r' | 's' | 't' | 'u' |
'v'|'w'|'x'|'y'|'z'
Digit = '0' | '1' | '2' | '3' | '4' | '5' | '6' | '7' | '8' | '9'
SpecSsign = '.' | ',' | ':' | ';' | '(' | ')' | '=' | '+' | '-' | '*' | '/' | '^' | '<' | '>' | WhiteSpace | EndOfLine
WhiteSpace = '' | '\t'
EndOfLine = '\n' | '\r' | '\r\n' | '\n\r'
Символи
SpecSymbols = ArithOp | RelOp | BracketsOp | AssignOp | Punct
ArithOp = AddOp | MultOp | NeltOp
AddOp = '-'
MultOp = '*' | '/'
NeltOp = '^'
RelOp = '==' | '<= ' | '<' | '>' | '>=' | '<>' | '!='
BracketsOp = '(' | ')'
AssignOp = '='
Punct = '.' | ',' | ':' | ';'
Індифікатор
Ident = Letter { Letter | Digit }
Конастанти
Const = IntNumb | RealNumb | BoolConst
```

```
IntNumb = [Sign] UnsignedInt
RealNumb = [Sign] UnsignedReal
Sign = '-'
UnsignedInt = Digit {Digit}
UnsignedReal = '.' UnsignedInt | UnsignedInt '.' | UnsignedInt '.' UnsignedInt
BoolConst = true | false
Ключові слова
KeyWords = start | read | print | for | by | while | do | if | end
Синтаксис
Expression = ArithmExpression | BoolExpr
BoolExpr = Expression RelOp Expression | true | false
ArithmExpression = [Sign] Term | ArithmExpression '+' Term | ArithmExpression '-' Term
Term = Factor | Term '*' Factor | Term '/' Factor | '(' Tern '^' Factor ')'
Factor = Ident | Const | '('ArithmExpression ')'
Арифметичні операції
AddOp = '+'| '-'
MultOp = '*' | '/'
NeltOp = '^'
Оголошення
IdenttList = Ident {',' Ident}
Інструкції
DoSection = StatementList
StatementList = Statement {';' Statement }
Statement = Assign | Inp | Out | ForStatement | IfStatment
Присвоєення
Assign = Ident '=' Expression
Введення
Inp = read '('IdenttList ')'.
Виведення
Out = print '(' IdenttList ')'.
Інструкція повторення
ForStatement = for '(' IndExpr1 ')' by '(' IndExpr2 ')' while '(' BoolExpr ')' do '{' DoBlock '}'
IndExpr1 = Ident '=' ArithmExpression1
IndExpr2 = ArithmExpression2
DoBlock = '{' Statement '}' | '{' StatementList '}'
```

Умовний onepamop IfStatement = if '(' Relation ')' '{' DoBlock '}' Relation = BoolExpr

Таблиця лексем

Код	Приклади лексем	Токен	Неформальний опис
1	a, xyz1, data3	ident	ідентифікатор
2	123, 0, 521	intnum	ціле без знака
3	34.76	realnum	дійсне без знака
4	-6.7	realnum	дійсне зі знаком
5	true	boolval	логічне значення
6	start	keyword	символ start
7	read	keyword	символ read
8	print	keyword	символ print
9	for	keyword	символ for
10	by	keyword	символ by
14	do	keyword	символ do
16	while	keyword	символ while
17	if	keyword	символ if
18	end	keyword	символ end
19	false	boolval	символ false
20	=	assign_op	символ =
21	+	add_op	символ +
22	-	add_op	символ -
23	*	mult_op	символ *
24	/	mult_op	символ /
25	^	nelt_op	символ ^
26	<	rel_op	символ <
27	<=	rel_op	символ <=

28	==	rel_op	символ ===
29	>	rel_op	символ >
30	>=	rel_op	символ >=
31	!=	rel_op	символ !==
32	(par_op	символ (
33)	par_op	символ)
34	{	par_op	символ {
35	}	par_op	символ }
36		punct	символ .
37	,	punct	символ,
38	:	punct	символ:
39	÷,	punct	символ;
40	\32,\t	ws	пробільні символи
41	\n, \r, \r\n, \n\r	ls	кінець рядка

Лексичний аналізатор

Лексичний аналізатор перевіряє коректність символів програми і формує таблицю символів.

Формат таблиці символів

Таблиця розбору реалізована як словник tableOfSymb у форматі:

{ n_rec : (num_line, lexeme, token, idxIdConst) } де:

n rec – номер запису в таблиці символів програми;

num line – номер рядка вхідної програми;

lexeme - лексема;

token – токен лексеми;

idxIdConst – індекс ідентифікатора або константи у таблиці ідентифікаторів та констант відповідно; для інших лексем – порожній рядок.

Базовий приклад програми

```
start ProgramName
  c = 1.0 * 9.0 + 3.0;
  c = -c;
  d = 3.0 ^ (5.0 / 10.0);
  for ( I = 1 ) by ( I = I + 1) while (I < 10) do {
     print(i);
  }
  I = 1;</pre>
```

```
I = -I;

if ( I != 10 ) {
    print (i);
}

if ( I == 10 ) {
    print (i);
}

if ( I > 10 ) {
    print (i);
}

if ( I < 10 ) {
    print (i);
}

if ( I <= 10 ) {
    print (i);
}

end</pre>
```

Програмна реалізація

Почнемо з правила для кореневого нетермінала граматики Program = start StatementList

Для цього визначимо функцію parseProgram() через parseToken(lexeme,token) та parseStatementList():

```
def parseProgram():
    print('\n' + '=' * 60 + '\n')
    try:
        parseToken('start', 'keyword', '')
        parseToken('', 'ident', '')
        parseStatementList()
        parseToken('end', 'keyword', '')

        print('\nParser: Синтаксичний аналіз і трансляція завершені успішно')
        return True
    except SystemExit as e:
        print('\nParser: Аварійне завершення програми з кодом {0}'.format(e))
        exit()
```

перевіряємо наявність ключового слова start і назви програми, а потім списку інструкцій. При успішному аналізі виведемо відповідне повідомлення. У разі неуспішного аналізу відловлюємо помилку і завершуємо виконання.

parseToken(): Функція перевіряє, чи у поточному рядку таблиці розбору зустрілась вказана лексема lexeme з токеном token

Прочитати з таблиці розбору поточний запис. Повертає номер рядка програми, лексему та її токен

```
def getSymb():
    if numRow > len_tableOfSymb:
        return False
    numLine, lexeme, token, _ = tableOfSymb[numRow]
    return numLine, lexeme, token
```

Функція parseStatementList() для розбору за правилом StatementList = Statement { Statement } викликає функцію parseStatement() доти, доки parseStatement() повертає True

```
def parseStatementList(specInstr=''):
    print('\t parseStatementList():')
    while parseStatement(specInstr):
        pass
    return True

def parseStatement(specInstr=''):
    global numRow
    if getSymb():
        numLine, lex, tok = getSymb()
    else:
        return False
    if (lex, tok) == ('end', 'keyword'):
```

```
parsePrint()
```

До функції parseAssign() переходимо, уже знаючи, що поточна лексема – ідентифікатор. У відповідності до правила граматики Assign = Ident '=' Expression беремо цей ідентифікатор і, якщо далі зустрічається лексема =, то викликаємо функцію parseExpression() для розбору арифметичного виразу:

```
def parseAssign():
    global numRow
    if viewSyntax:
        print('\t' * 4 + 'parseAssign():')
    numLine, lex, tok = getSymb()
    postfixCode.append((lex, tok)) # Трансляція
    # if viewTranslation:
    # configToPrint(lex, numRow)
    numRow += 1
    if viewSyntax:
        print('\t' * 5 + '[{0}]: {1}'.format(numLine, (lex, tok)))
    if parseToken('=', 'assign_op', '\t\t\t\t\t\t'):
        numRowCopy = numRow - 1
```

```
parseExpression()
postfixCode.append(('=', 'assign_op')) # Трансляція
# Бінарний оператор '=' додається після своїх операндів
# if viewTranslation:
# configToPrint('=', numRowCopy)
return True
else:
return False
```

Правило для арифметичного виразу:

```
def parseExpression():
    global numRow, postfixCode

if viewSyntax:
        print('\t' * 5 + 'parseExpression():')
    parseTerm()
    F = True
    while F:
        if getSymb():
            numLine, lex, tok = getSymb()
        else:
            return True
        if tok in 'add_op':
            numRowCopy = numRow
            numRow += 1
            if viewSyntax:
                print('\t' * 6 + '[{0}]: {1}'.format(numLine, (lex, tok)))
            parseTerm()
            postfixCode.append((lex, tok)) # трансляція
            # if viewTranslation:
            # # configToPrint(lex, numRowCopy)
            else:
            F = False
            return True
```

```
def parseFactor():
  if viewSyntax:
      postfixCode.append((lex, tok))
```

обробка оператора читання:

```
def parseRead():
    global numRow
    if viewSyntax:
        print('\t' * 4 + 'parseRead():')
    _, lex, tok = getSymb()
    if lex == 'read' and tok == 'keyword':
        numRow += 1
        parseToken('(', 'brackets_op', '\t' * 5)
        parseIdentList()
        parseToken(')', 'brackets_op', '\t' * 5)
        return True
    else:
        return False
```

обробка оператора виводу:

```
def parsePrint():
    global numRow
    if viewSyntax:
        print('\t' * 4 + 'parsePrint():')
    _, lex, tok = getSymb()
    if lex == 'print' and tok == 'keyword':
        numRow += 1
        parseToken('(', 'brackets_op', '\t' * 5)
        parseExpressionList()
        parseToken(')', 'brackets_op', '\t' * 5)
        return True
    else:
        return False
```

Обробка списку виразів:

```
def parseExpressionList():
    if viewSyntax:
        print('\t' * 5 + 'parseExpressionList():')
    while parseExpression():
        if getSymb():
            numLine, lex, tok = getSymb()
        else:
            return True
        if lex == ')':
            break
```

```
parseToken(',', 'punct', '\t\t\t\t\t')
  postfixCode.append(('OUT', 'out'))
  return True
```

розбір інструкції розгалудження за правилом IfStatement = if '(' Relation ')' '{ StatementList '}'

```
def parseIf():
    global numRow
    if viewSyntax:
        print('\t' * 4 + 'parseIf():')
    _, lex, tok = getSymb()
    if lex == 'if' and tok == 'keyword':
        numRow += 1
        parseToken('(', 'brackets_op', '\t' * 5)
        parseBoolExpr()
        parseToken(')', 'brackets_op', '\t' * 5)
        parseToken('{', 'brackets_op', '\t' * 5)
        parseStatementList('IF')
        parseToken('}', 'brackets_op', '\t' * 5)
        return True
else:
    return False
```

розбір інструкції повторювання за правилом ForStatement = for '(' IndExpr1 ')' by '(' IndExpr2 ')' while '(' BoolExpr ')' do '{' DoBlock '}'

```
def parseFor():
    global numRow
    if viewSyntax:
        print('\t' * 4 + 'parseFor():')
    _, lex, tok = getSymb()
    if lex == 'for' and tok == 'keyword':
        numRow += 1
        numLine, lex, tok = getSymb()

        parseToken('(', 'brackets_op', '\t' * 5)
        parseExpression()
        parseToken('=', 'assign_op', '\t' * 5)
        parseToken(', 'brackets_op', '\t' * 5)
        parseToken('by', 'keyword', '\t' * 5)
        parseToken('(', 'brackets_op', '\t' * 5)
        parseToken('(', 'brackets_op', '\t' * 5)
        parseToken('', 'assign_op', '\t' * 5)
        parseToken('', 'brackets_op', '\t' * 5)
        parseToken('', 'brackets_op', '\t' * 5)
        parseToken('', 'brackets_op', '\t' * 5)
        parseToken('(', 'brackets_op', '\t' * 5)
        parseToken(''), 'brackets_op', '\t' * 5)
        parseToken('', 'brackets_op', '\t' * 5)
        return True
else:
        return False
```

розбір логічного виразу за правилом BoolExpr = Expression ('='|'<='|'>='|'<'|'>'') Expression

```
def parseBoolExpr():
    global numRow
    if viewSyntax:
        print('\t' * 5 + 'parseBoolExpression():')
    parseExpression()
    if getSymb():
        numLine, lex, tok = getSymb()
    else:
        return True
    numRow += 1
    parseExpression()
    if tok in ('rel_op'):
        postfixCode.append((lex, tok)) # Трансляція
        if viewSyntax:
            print('\t' * 5 + '[{0}]: {1}'.format(numLine, (lex, tok)))
    else:
        failParse('невідповідність у ВооlExpr', (numLine, lex, tok, '== != <= >= < >'))
    return True
```

Тестування

- 1. Пропуск термінала
 - а. Пропуск назви програми (Error 111)

Код:

```
start end
  d = 3.0 ^ (5.0 / 10.0);
  if ( d != 10 ) {
    print (i);
  }
end
```

Результат:

```
parseToken():
[1]: ('start', 'keyword')
Parser ERROR:
[1]: Неочікуваний елемент ('for', keyword).
Очікувався ідентифікатор
Parser: Аварійне завершення програми з кодом 111
```

b. Пропуск назви змінної (Error 108)

Код:

```
parseToken():
[1]: ('start', 'keyword')
parseToken():
[1]: ('ProgramName', 'ident')
    parseStatementList():
    parseStatement():

Parser ERROR:
    [2]: Неочікуваний елемент ('=', assign_op).
    Очікувалася інструкція.

Parser: Аварійне завершення програми з кодом 108
```

с. Пропуск оператора присвоювання (Error 107)

Код:

```
start ProgramName
c 1.5
end
```

Результат:

d. Пропуск закриваючої дужки (Error 107)

Код:

```
start ProgramName
    k = 1;
    if (k != 10) {
        print (k);
end
```

```
parseExpression():
    parseExpression():
    parseTerm():
        parseFactor():
        [3]: ('k', 'ident')
    parseExpression():
        parseTerm():
        parseTerm():
        parseFactor():
        [3]: ('10', 'integer')
        [3]: ('10', 'integer')
        [3]: ('10', 'integer')
        [3]: ('l', 'brackets op')
        parseToken():
        [3]: ('l', 'brackets op')
        parseStatementList():
        parseStatement():
        parseFrint():
        parseToken():
        [4]: ('(', 'brackets op')
        parseExpressionList():
        parseExpression():
        parseExpression():
        parseToken():
        [4]: ('k', 'ident')
        parseToken():
        [4]: ('l', 'brackets_op')
        parseToken():
        [4]: ('l', 'punct')

Parser ERROR:
    [6]: Неочікуваний елемент ('end', keyword).
        Очікувався - (')', brackets_op).
```

е. Пропуск знака пунктуації ; (Error 107)

Код:

```
start ProgramName
   k = 1
   print (k);
end
```

2. Зайвий термінал

а. Зайвий ідентифікатор (Error 107)

Код:

```
start ProgramName bla
  print(5);
end
```

Результат:

b. Зайвий знак (Error 109)

Код:

```
  k = 1 + / 3;  end
```

```
parseFactor():
        [2]: ('1', 'integer')
        [2]: ('+', 'add_op')
        parseTerm():
            parseFactor():
        [2]: ('/', 'mult_op')

Parser ERROR:
    [2]: Heovikyваний елемент ('/', mult_op).
    Очікувався - 'integer, real, ident aбо '(' Expression ')''.

Parser: Аварійне завершення програми з кодом 109
```

с. Помилка у наборі ключового слова

Код:

```
start ProgramName
  red(k);
end
```

Результат:

3. Порожній файл

Parser ERROR:

4. Shorted basic example without errors

Код:

```
start ProgramName
    c = 1.0 * 9.0;
    for (i = 1) by (i = i + 1) while (i < 10) do {
        print(i);
    }
    if (c != 10) {
        print (i);
    }
end</pre>
```

```
parseToken():
[1]: ('start', 'keyword')
parseToken():
              parseFor():
                  [3]: ('i', 'ident')
parseToken():
[3]: ('=', 'assign_op')
                  [3]: ('(', 'brackets op')
```

```
[3]: ('i', 'ident')
[3]: ('+', 'add_op')
parseTerm():
```

```
parseFactor():
        [6]: ('10', 'integer')
        [6]: ('10', 'integer')
        [6]: ('10', 'integer')
        parseToken():
        [6]: (')', 'brackets_op')
        parseToken():
        [6]: ('[', 'brackets_op')
        parseStatementList():
        parseStatement():
        parsePrint():
        parseToken():
        [7]: ('(', 'brackets_op'))
        parseExpressionList():
        parseExpression():
        parseTerm():
            parseToken():
        [7]: ('i', 'ident')
        parseToken():
        [7]: (')', 'brackets_op')
        parseStatement():
        [7]: (';', 'punct')
        parseToken():
        [8]: (']', 'brackets_op')

parseToken():
        [9]: ('end', 'keyword')

Parser: Синтаксичний аналіз і трансляція завершені успішно
```

Висновок: в ході даної роботи я освоїла основи розробки синтаксичного аналізатору та задіяла їх на практиці. Також мною було розроблено синтаксичний аналізатор для мови програмування КК.