

Translacja z Pythona – dokumentacja wstępna

Autor: Jakub Ficek

Temat projektu

Celem projektu jest stworzenie translatora skróśnego dla wybranego podzbioru języka Python do języka C++. Program będzie napisany w języku Python3.

Założenia

- Obsługiwanymi typami są int, float, bool
- Wszystkie zmienne są statycznie typowane (zmienna nie może być najpierw typu int a potem float)
- Obsługa własnych funkcji z adnotowanymi typami argumentów i zwracanej wartości
- Obsługa instrukcji warunkowych if elif else
- Obsługa pętli while
- Obsługa wyrażeń matematycznych i logicznych
- Obsługa wyjścia przy pomocy funkcji print

Przykłady

```
x = 5
while x > 0:
    if x % 3 == 0:
        print(0)
    elif x % 3 == 1:
        print(1)
    else:
        print(2)
    x = x - 1
```

```
#include <iostream>

using namespace std;

int main()
{
    int x;
    x = 5;
    while(x > 0)
    {
        if(x % 3 == 0)
        {
            cout << 0 << endl;
        }
        else if(x % 3 == 1)
        {
            cout << 1 << endl;
        }
        else
        {
            cout << 2 << endl;
        }
        x = x - 1;
    }
    return 0;
}
```

<pre>def sum(a: int, b: int) -> int: return a + b x = 2 while x > 0 and 1 == 1: x = x - 1 print(x) y = 3 + 2 * 2 print(sum(x, y))</pre>	<pre>#include <iostream> using namespace std; int sum(int a, int b) { return a + b; } int main() { int x; int y; x = 2; while(x > 0 && 1 == 1) { x = x - 1; cout << x << endl; } y = 3 + 2 * 2; cout << sum(x, y) << endl; return 0; }</pre>
---	--

<pre>x=2 x++</pre>	<pre>error: line 2</pre>
--------------------	--------------------------

Lista zdefiniowanych tokenów

'def', ':', 'if', 'elif', 'else', 'while', '(', ')', 'int', 'float', 'bool', 'None', 'return', '<', '>', '<=', '>=', '==', '!=', '=', '->', 'and', 'or', 'not', '+', '-', '*', '/', '%', 'print'

Gramatyka

program = statements

statements = statement | statement statements

statement_block = indent statements dedent

statement = assignment_statement | function_statement | return_statement | while_statement | if_statement | print_statement | expression_statement | EOL

assignment_statement = identifier '=' expression_statement

function_statement = 'def' identifier '(' [identifier ':' type] { ',' identifier ':' type } ')' '->' (type | 'None') ':' EOL statement_block

return_statement = 'return' expression_statement

while_statement = 'while' expression_statement ':' EOL statement_block

if_statement = 'if' expression_statement ':' EOL statement_block [el_statement]

el_statement = else_statement | elif_statement

elif_statement = 'elif' expression_statement ':' EOL statement_block [el_statement]

else_statement = 'else' ':' EOL statement_block

print_statement = 'print' '(' {expression_statement} ')'

expression_statement = func_call | operation

operation = { unary_op | unary_logic_op } expression_statement { binary_op | binary_logic_op
| comparison_op expression_statement }

func_call = identifier '(' [identifier | value] { ',' (identifier | value) } ')'

unary_op = '+' | '-'

unary_logic_op = 'not'

binary_op = '+' | '-' | '*' | '/' | '%'

binary_logic_op = 'and' | 'or'

comparison_op = '==' | '!=' | '<' | '<=' | '>=' | '>'

type = 'int' | 'float' | 'bool'

value = int | float | bool

identifier = [a-zA-Z_][a-zA-Z0-9_]*

intend – wzrost liczba początkowych tabulacji o 1 względem poprzedniej linii

dedend – spadek liczby początkowych tabulacji o 1 względem poprzedniej linii

Sposób uruchomienia

Program na wejściu dostaje ścieżkę do pliku z kodem napisanym w języku Python, jako wyjście program wypisuje lub tworzy plik z analogicznym kodem napisanym w C++ albo wypisuje komunikat o odpowiednim błędzie (błąd może wystąpić na etapie analizy leksykalnej lub składniowej)

Moduły

- **Moduł obsługi plików** – odpowiedzialny za wczytywanie tekstu z pliku i tworzenie pliku wyjściowego
- **Lekser** – odpowiedzialny za rozbicie tekstu wczytanego z pliku na tokeny, które trafiają do analizatora składniowego.
- **Parser** – odpowiedzialny za sprawdzenie czy wczytany ciąg tokenów jest zgodny z gramatyką zdefiniowanego podzbioru oraz zbudowanie drzewa rozbioru.
- **Generator kodu** – odpowiedzialny za konstrukcję kolejnych instrukcji w C++, przechodząc drzewo rozbioru oraz korzystając z tablicy symboli, aby zainicjalizować wszystkie zmienne
- **Moduł obsługi błędów** – odbiera błędy z leksera, parsera oraz prezentuje błędy w czytelny sposób (numer linii)

Dodatkowe struktury danych:

- **Tablica symboli** – zawierająca używane w kodzie wejściowym identyfikatory oraz informacje o ich typie i zasięgu, uzupełniana przez lekser oraz parser
- **Tablica akceptowalnych tokenów** – predefiniowana tablica wszystkich zdefiniowanych symboli, używana przez lekser

Reguły translacji

- Stała struktura kodu w C++, pojawiająca się zawsze:

```
#include <iostream>
```

```
using namespace std;
```

```
//definicje funkcji
```

```
int main()
```

```
{
```

```
    // inicjalizacja zmiennych o zasięgu globalnym
```

```
    // translacja pythonowych instrukcji
```

```
    return 0;
```

```
}
```

- Na koniec każdej instrukcji (równoznaczną niepustej linii w pythonie), która w pythonie nie kończy się dwukropkiem (aby wyeliminować pętle oraz definicje funkcji), dodawany jest średnik
- Na początku funkcji inicjowane są wszystkie zmienne lokalne
- Zamiana 'int', 'float', 'bool', 'None' na kolejno 'int', 'float', 'bool', 'void'
- Zamiana funkcji postaci 'def func(arg1: type1, arg2:type2) -> type3:' na 'type3 func(type1 arg1, type2 arg2)'
- Zwiększenie liczby wcięć oznaczane przez '{', a zmniejszenie przez '}'
- Zamiana 'if expression:' na 'if(expression)'
- Zamiana 'elif expression:' na 'else if(expression)'
- Zamiana 'else:' na 'else'
- Zamiana 'while expression' na 'while(expression)'
- Operatory porównania, przypisania oraz arytmetyczne pozostają bez zmian
- Zamiana operatorów logicznych 'not', 'and', 'or' na kolejno '!', '&&', '||'
- Zamiana 'print(expression1, expression2)' na 'cout << expression1 << expression2 << endl'

Testowanie

Testowanie będzie odbywać się przy użyciu krótkich przykładów testujących, porównując uzyskany przez program kod z oczekiwanym.