CPS2000 - Compiler Theory and Practice
Course Assignment 2018/2019

Department of Computer Science, University of Malta
Sandro Spina

March 22, 2019

Instructions

- This is **an individual** assignment. This assignment carries **50**% of the final **CPS2000** grade.
- It is strongly recommended that you start working on the tasks as soon as the related material is covered in class.
- The submission deadline is Friday **24**th **May 2019**. A soft-copy of the report and all related files (including code) must be uploaded to the VLE by 5pm of the indicated deadline. Hard copies **are not** required to be handed in. Source and executable files must be archived into a single .zip file before uploading to the VLE. It is the student's responsibility to ensure that the uploaded .zip file is valid. The PDF report must be submitted separately through the Turnitin submission system on the VLE.
- A report describing how you designed and implemented the different tasks of the assignment **is required**. Tasks (1-6) for which no such information is provided in the report will **not be** assessed.
- You are welcome to share ideas and suggestions. However, under no circumstances should code be shared among students. Please remember that plagiarism will not be tolerated; the final submission must be entirely your work.
- You are to allocate approximately **40** hours for this assignment.

Description

In this assignment you are to develop a lexer, parser, interpreter and an ILOC instructions generator for a small programming language - *MiniLang*. The specification of this language can be found in the next sections. The assignment is composed of three major parts: i) a FSA-based lexer and hand-crafted top-down recursive-descent parser, ii) visitor classes to perform XML generation, semantic analysis and interpreter execution on the abstract syntax tree (AST) produced by the parser and iii) a report detailing how you designed and implemented the different tasks. These parts are subdivided in tasks and explained in detail in the Task Breakdown section.

MiniLang is an expression-based strongly-typed programming language. The language has C-style comments, that is, `//...` for line comments and `/*...*/` for block comments. The language is case-sensitive and each function is expected to return a value. MiniLang has 3 types: 'float', 'int' and 'bool'. Binary operators, such as '+', require that the operands have matching types and the language does not perform any implicit/automatic typecast.

The following is a syntactically and semantically correct MiniLang program:

```
fn Square(x:float) : float {
 return x*x;
}

fn XGreaterThanY(x:float, y:float) : bool {
 var ans:bool = true;
 if (y > x) { ans = false; }
 return ans;
}

fn AverageOfThree(x:float, y:float, z:float) : float {
  var total:float = x + y + z;
  return total / 3;
}

var x:float = 2.4;
var y:float = Square(2.5);
print y;                                  //6.25
print XGreaterThanY(x, 2.3);              //true
print XGreaterThanY(Square(1.5), y);      //false
print AverageOfThree(x, y, 1.2);          //3.28
```

MiniLang (in EBNF)

⟨*Letter*⟩ ::= [A-Za-z]

⟨*Digit*⟩ ::= [0-9]

⟨*Printable*⟩ ::= [\x20-\x7E]

⟨*Type*⟩ ::= ‘`float`’ | ‘`int`’ | ‘`bool`’

⟨*BooleanLiteral*⟩ ::= ‘`true`’ | ‘`false`’

⟨*IntegerLiteral*⟩ ::= ⟨*Digit*⟩ { ⟨*Digit*⟩ }

⟨*FloatLiteral*⟩ ::= ⟨*Digit*⟩ { ⟨*Digit*⟩ } ‘`.`’ ⟨*Digit*⟩ { ⟨*Digit*⟩ }

⟨*Literal*⟩ ::= ⟨*BooleanLiteral*⟩
  | ⟨*IntegerLiteral*⟩
  | ⟨*FloatLiteral*⟩

⟨*Identifier*⟩ ::= (‘`_`’ | ⟨*Letter*⟩) { ‘`_`’ | ⟨*Letter*⟩ | ⟨*Digit*⟩ }

⟨*MultiplicativeOp*⟩ ::= ‘`*`’ | ‘`/`’ | ‘`and`’

⟨*AdditiveOp*⟩ ::= ‘`+`’ | ‘`-`’ | ‘`or`’

⟨*RelationalOp*⟩ ::= ‘`<`’ | ‘`>`’ | ‘`==`’ | ‘`!=`’ | ‘`<=`’ | ‘`>=`’

⟨*ActualParams*⟩ ::= ⟨*Expression*⟩ { ‘`,`’ ⟨*Expression*⟩ }

⟨*FunctionCall*⟩ ::= ⟨*Identifier*⟩ ‘`(`’ [⟨*ActualParams*⟩] ‘`)`’

⟨*SubExpression*⟩ ::= ‘`(`’ ⟨*Expression*⟩ ‘`)`’

⟨*Unary*⟩ ::= (‘`-`’ | ‘`not`’) ⟨*Expression*⟩

⟨*Factor*⟩ ::= ⟨*Literal*⟩
  | ⟨*Identifier*⟩
  | ⟨*FunctionCall*⟩
  | ⟨*SubExpression*⟩
  | ⟨*Unary*⟩

⟨*Term*⟩ ::= ⟨*Factor*⟩ { ⟨*MultiplicativeOp*⟩ ⟨*Factor*⟩ }

⟨*SimpleExpression*⟩ ::= ⟨*Term*⟩ { ⟨*AdditiveOp*⟩ ⟨*Term*⟩ }

⟨*Expression*⟩ ::= ⟨*SimpleExpression*⟩ { ⟨*RelationalOp*⟩ ⟨*SimpleExpression*⟩ }

⟨*Assignment*⟩ ::= ⟨*Identifier*⟩ ‘`=`’ ⟨*Expression*⟩

⟨*VariableDecl*⟩ ::= ‘`var`’ ⟨*Identifier*⟩ ‘`:`’ ⟨*Type*⟩ ‘`=`’ ⟨*Expression*⟩

⟨*PrintStatement*⟩ ::= ‘`print`’ ⟨*Expression*⟩

⟨*ReturnStatement*⟩ ::= ‘`return`’ ⟨*Expression*⟩

⟨*IfStatement*⟩ ::= ‘`if`’ ‘`(`’ ⟨*Expression*⟩ ‘`)`’ ⟨*Block*⟩ [‘`else`’ ⟨*Block*⟩]

⟨*ForStatement*⟩ ::= ‘`for`’ ‘`(`’ [⟨*VariableDecl*⟩] ’;’ ⟨*Expression*⟩ ’;’ [⟨*Assignment*⟩] ‘`)`’ ⟨*Block*⟩

⟨*FormalParam*⟩ ::= ⟨*Identifier*⟩ ‘`:`’ ⟨*Type*⟩

⟨*FormalParams*⟩ ::= ⟨*FormalParam*⟩ { ‘`,`’ ⟨*FormalParam*⟩ }

⟨*FunctionDecl*⟩ ::= ‘`fn`’ ⟨*Identifier*⟩ ‘`(`’ [⟨*FormalParams*⟩] ‘`)`’ ‘`:`’ ⟨*Type*⟩ ⟨*Block*⟩

⟨*Statement*⟩ ::= ⟨*VariableDecl*⟩ ‘`;`’
 | ⟨*Assignment*⟩ ‘`;`’
 | ⟨*PrintStatement*⟩ ‘`;`’
 | ⟨*IfStatement*⟩
 | ⟨*ForStatement*⟩
 | ⟨*ReturnStatement*⟩ ‘`;`’
 | ⟨*FunctionDecl*⟩
 | ⟨*Block*⟩

⟨*Block*⟩ ::= ‘`{`’ { ⟨*Statement*⟩ } ‘`}`’

⟨*Program*⟩ ::= { ⟨*Statement*⟩ }

Task Breakdown

Task 1 - A table-driven lexer

In this first task you are to develop the lexer for the *MiniLang* language using the C++ programming language. The lexer is to be implemented using the table-driven approach, which encodes the DFA transition function of the *MiniLang* micro-syntax. The lexer should be able to report any lexical errors in the input program.

[Marks: 20%]

Task 2 - Hand-crafted recursive descent parser

In this task you are to develop a hand-crafted predictive parser for the *MiniLang* language using the C++ programming language. The Lexer and Parser classes interact through the function *GetNextToken*() which the parser uses to get the next valid token from the lexer. The parser should be able to report any syntax errors in the input program. A successful parse of the input should produce an abstract syntax tree (AST) describing the structure of the program.

[Marks: 30%]

Task 3 - AST XML Generation Pass

In OOP programming, the Visitor design pattern is used to describe an operation to be performed on the elements of an object structure without changing the classes on which it operates. In our case this object structure is the AST produced by the parser in Task 2. For this task you are to implement a visitor class to output a properly indented XML representation of the generated AST.

```
var x : float = 8.1 + 2.2;

<Decl>
    <Var Type="float">x</Id>
    <BinExprNode Op="+">
        <FloatConstant>8.1</FloatConstant>
        <FloatConstant>2.2</FloatConstant>
    </BinExprNode>
</Decl>
```

[Marks: 5%]

Task 4 - Semantic Analysis Pass

For this task, you are to implement another visitor class to traverse the AST and perform type-checking (e.g. checking that variables are assigned to appropriately typed expressions, variables are not declared multiple times in the same scope, etc.). In addition to the global program scope, scopes are created whenever a block is entered and destroyed when control leaves the block. Note that blocks may be nested and that to carry out this task, it is essential to have a proper implementation of a symbol table.

[Marks: 20%]

Task 5 - Interpreter Execution Pass

For this task, you are to implement another visitor class to traverse the AST which simulates an interpreter and executes the test program. The *'print'* <Expression> statement can be used in your test programs to output the value of <Expression> to the console and determine whether the computation carried out by the interpreter visitor is correct.

[Marks: 25%]

Report

In addition to the source and class files, you are to deliver a report. Tasks 1 to 5 for which no information is provided in the report will not be assessed. In your report include any deviations from the original EBNF, the salient points on how you developed the lexer / parser / interpreter (and reasons behind any decisions you took) including semantic rules and code execution, and any sample *MiniLang* programs you developed for testing the outcome of your compiler. In your report, state what you are testing for, insert the program AST and the outcome of your test. As an example, the *MiniLang* source script below, computes the answer of a real number raised to an integer power:

```
//Function definition for Power
fn Pow(x:float , n:int) : float
{
    var y : float = 1.0;          //Declare y and set it to 1.0
    if( n>0 )
    {
        for(; n>0 ; n=n-1)
        {
            y = y * x;            //Assignment y = y * x;
        }
    }
    else
    {
        for(; n<0 ; n=n+1)
        {
            y = y / x;            //Assignment y = y / x;
        }
    }
    return y;                     //return y as the result
}
```