Lab 1 - Lexical Analyzer

Course: Principles of Programming Languages (Code: IT092IU)

HCMIU-CSE, Summer 2024

Instructor: Le Thi Ngoc Hanh, PhD

1 ANTLR Guidelines

- Install Python 3.

- Download ANTLR at this link

- Add directory to PATH with the location of your downloaded ANTLR jar file (antlr-4.9.2-complete.jar) (optional)

- Install the ANTLR runtime for Python: pip install antlr4-python3-runtime

- Write your ANTLR grammar file (*.g4)

- Generate the lexer using ANTLR with command:

java -jar *path_to_*.jar -Dlanguage=Python3 *your-designed-language*.g4

- Write Python script to use the generated lexer.

2 Example

Use ANTLR to write regular expressions describing a language that must begin with a lowercase letter (“a” to “z”).

3 Exercises

Use ANTLR to write regular expressions describing:

Exercise 1:

A language that must begin with a lowercase letter (“a” to “z”), but may continue with many characters which are lowercase letter or digit (“0” to “9”).

Exercise 2:

A language accepts input consisting of the keyword “hello” followed by an identifier made up of lowercase letters, with optional whitespace in between.

Exercise 3:

A language that accepts input consisting integers (“0” to “9”), lowercase letters (“a” to “z”), question marks, optional whitespace in between.

Exercise 4:

A language accepts all integer numbers, integer numbers do not start with 0; Float numbers do not start with 0, example: 1.0, 1., 12e2, 9e-2?

Exercise 5:

A language is defined by digit (“0” to “9”), 4 arithmetic operators (+, -, *, /), round brackets for parentheses. The production rules are:

expr : term ((PLUS | MINUS) term)* ;

term : factor ((MULT | DIV) factor)* ;

factor : NUMBER | LPAREN expr RPAREN ;