HCMC University of Technology
Faculty of Computer Science & Engineering

Assignment 1

Lexer & Recognizer

Author Dr. Nguyen Hua Phung

August 25, 2019

Contents

Assignment 1

version 1.0

After completing this assignment, you will be able to

- define formally lexicon of a programming language.
- use ANTLR to implement a lexer for a programming language.
- define formally grammar of a programming language.
- use ANTLR to implement a recognizer for a programming language.

1 Specification

In this assignment, you are required to write a lexer and a recognizer for a program written in MC. To complete this assignment, you need to:

- Set up the environment
  - Make sure that **java** can run on your machine otherwise JAVA JDK must be installed (*https://www.oracle.com/technetwork/java/javase/downloads/jdk8-downloads-2133151.html*)
  - Download and install python 3 from *https://www.python.org/*
  - Download file antlr-4.7.2-complete.jar from *https://www.antlr.org/download.html* (the link below "Complete ANTLR-4.7.2 Java binaries jar").
  - Set an environment variable named $ANTLR\_JAR$ keep the path to the file antlr-4.7.2-complete.jar.
  - Follow the instructions in Section **Python Targets** of *https://www.antlr.org/download.html* to install/have *antlr4-python3-runtime*.
  - Download assignment1.zip, unzip it and follow the instructions in initial/README.txt to test your environment.
- read carefully the specification of MC language
- Modify initial/src/main/mc/parser/MC.g4. in the initial code to describe formally MC language.**Please fill in your id in the comment in the header of this file**.
- Add more test in initial/src/test/LexerSuite.py and initial/sr/test/ParserSuite.py in the initial code.

2 Submission

In this assignment, you are required to submit three files MC.g4, LexerSuite.py and ParserSuite.py. **Note that you must submit 3 files, NOT compress them.**

- Modify MC.g4 to detect tokens and check grammar of MC programs.
- Make 100 testcases in LexerSuite.py to test your lexer rule.
- For lexical errors, please throw the exception as follows:
  - ErrorToken(<char>): when the lexer detects an unrecognized character
  - UnclosedString(<unclosed string>): when the lexer detects an unterminated string. The unclosed string is from the beginning of the string (without the quote) to the newline or end of file, exclusively.
  - IllegalEscapeInString(<wrong string): when the lexer detects an illegal escape in string. The wrong string is from the beginning of the string (without the quote) to the illegal escape, inclusively.
- Make 100 testcases in ParserSuite.py to test your grammar rules.

You can assume that there is at most one error in each test case.

The deadline of both phases of assignment 1 is announced in the class website.

3 Plagiarism

You must complete the assignment by yourself and do not let your work seen by someone else, otherwise, you will be punished by the university rule for plagiarism.