HCMC University of Technology
Faculty of Computer Science & Engineering

Assignment 1

Lexer & Recognizer

Author   Dr. Nguyen Hua Phung

August 23, 2018

Contents

Assignment 1

version 1.1

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 prepare for this assignment, you need to:

- Download initial.zip and unzip it.

- Download antlr-4.7.1-complete.jar from antlr.org, set the environment variable ANTLR_LIB to this file and follow the íntructions in initial/README.txt to test the initial code.

- Remove all files in folders initial/src/main/mp/utils, initial/src/main/mp/astgen, initial/src/main/mp/checker, initial/src/main/mp/codegen.

- Delete files initial/src/test/ASTGenSuite.py, initial/src/test/CheckerSuite.py, and initial/src/test/CodeGenSuite.py

- Comment out five lines 11-15 and from line 103-end of file initial/src/test/TestUtils.py and test the initial code again with just three following íntructions:
  python run.py gen
  python run.py test LexerSuite
  python run.py test ParserSuite

- Change folder **initial** into **assignment1**

To complete this assignment, you need to:

- read carefully the specification of MP language

- Modify MP.g4. in the initial code to describe formally MP language.**Please fill in your id in the header of this file**.

- Add more test in LexerSuite and ParserSuite in the initial code.

This assignment is divided two phases: lexer phase and recognizer phase. **These phases are assessed independently.**

1.1 Phase 1: Lexer

In this phase, you are required to write a lexer for a program written in ANTLR. To complete this phase, you need to:

- Modify MP.g4 to detect tokens in MP language.

- Make 100 testcases for LexerSuite to test your code.

- For lexical errors, please print out as follows:

  khi lexer phát hiện một ký tự không được công nhận
  - "ErrorToken "+ <char>: when the lexer detects an unrecognized character

  - "Unclosed string: "+<unclosed string>: when the lexer detects an unterminated string. khi lexer phát hiện chuỗi chưa hoàn thành

  - "Illegal escape in string: "+<wrong string>: when the lexer detects an illegal escape in string. The wrong string is from the beginning of the string to the illegal escape. khi lexer phát hiện một lối thoát bất hợp pháp trong chuỗi. Chuỗi sai là từ đầu chuỗi đến lối thoát bất hợp pháp.

- You can assume that there is only one error in each test case.
  Bạn có thể giả định rằng chỉ có một lỗi trong mỗi trường hợp thử nghiệm.

1.2 Phase 2: Recognizer

In this phase, you are required to write a recognizer for a program written in MP. To complete this phase, you need to:

- Modify MP.g4.

- Make 100 testcases for ParserSuite to test your code.

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

2 Requirements

Note that you must NOT compress your files when submit them. You may test your code at the website:
http://www.cse.hcmut.edu.vn/onlinejudge, but you MUST submit three files MP.g4, LexerSuite.py and ParserSuite.py in BKeL.

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

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.

3 Change Log

From 1.0,

- Add the preparation instructions for the assignment 1.