HCMC University of Technology
Faculty of Computer Science & Engineering

Assignment 1

Recognizer

Author Dr. Nguyen Hua Phung

April 22, 2020

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 BKIT. To complete this assignment, you need to:

- Install Python 3 if you have not installed it yet.
- Download initial.zip and unzip it.
- Download antlr-4.7.2-complete.jar from https://www.antlr.org/download.html, set the environment variable ANTLR_JAR to this file; install antlr4-python3-runtime (see instructions in section Python Targets of the above webpage) and follow the instructions in initial/README.txt to test the initial code.
- Remove all files in folders initial/src/main/bkit/utils, initial/src/main/bkit/astgen, initial/src/main/bkit/checker if any.
- ~~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 tructions:
  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 language

- Modify BKIT.g4. in the initial code to describe formally BKIT 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 BKIT.g4 to detect tokens in BKIT language.
- Make 100 testcases for LexerSuite to test your code.
- For lexical errors, please print out as follows:
  - "Error Token "+ <char>: when the lexer detects an unrecognized character
  - "Unclosed String: "+<unclosed string>: when the lexer detects an unterminated string.
  - "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.
- You can assume that there is only one error in each test case.

1.2 Phase 2: Recognizer

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

- Modify BKIT.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 BKIT.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

- Changes from version 1.0
  - Replace ANTLR_LIB with ANTLR_JAR
  - Remove line "Comment out ..."
  - Fix error messages in Section 1.1