Homework 1: scanner

Compiler ConstructionInstructor: Zhiyao Liang
Macau University of Science and Technology
2024 FAll

1. Introduction

We will construct a *lexical analyzer*, also called a *scanner,* for the source language **PyC**, a new language designed to combine the features of C and Python.

Automata techniques can be used to build a scanner. Automata is fascinating, like the image shown above [1].

2 Helpful files

Some helpful files can be found at the Moodle site of this course at our university, under the section "Slides and Handouts".

- **Guidance with Helpful C code**: in the folder "PyC_scanner_ideas," some helpful C code fragments are listed, together with the needed guidance for this project.
- **Files of PyC**: Some source code programs of PyC are uploaded on Moodle.

- **Document of C-minus**: This file describes some techniques for building a scanner for the language C-minus, which can be similarly used for creating a scanner for the PyC.
- **Scanner algorithm**: A folder called "A simple DFA-based scanner algorithm" is on the Moodle site, which describes a DFA-based algorithm used to build a scanner which can help implement a scanner with high quality.

Implementing the scanner

Expected behavior of the scanner

Test your scanner with each of the provided sample language Pyr files (.arr), ensuring the correct list of tokens can be built and printed.

How to submit

- Files to be submitted on Moodle:
    - All source code files (.c and .h if using C to write the scanner). Using C++ is also allowed.
    - A "hmk1-report" text file. The format can be like .txt, .docx, .md, .pdf, or .html. This file should describe the following:
        - The full names of each group member.
        - What features of the scanner are successfully implemented?
        - If there are some remaining problems, what are they?
        - Any technical detail that is worthy of documentation. For example, if you have drawn some DFA graph, like what we see in the C-minus document, you are welcome to put the pictures in the document, although doing so is not required.
        - If there is more than one group member, describe each member's workload and contribution in detail.
            - Note: We do not want one member to do all the work or one member to do nothing.
- The due time is about three weeks after releasing this coment. The exact due time should refer to the information at the setting of Assignment 1 on Moodle.
- At most, three students can form a group to do the homework together and submit the homework together with an email. You are encouraged to do the homework alone (a one-person group).
- No plagarism, do not share your work online, no sharing allowed between groups.

References

1. This Programmable 6,000-Part Drawing Boy Automata is Arguably the First Computer and It Was Built 240 Years Ago
https://www.thisiscolossal.com/2013/11/the-writer-automata/