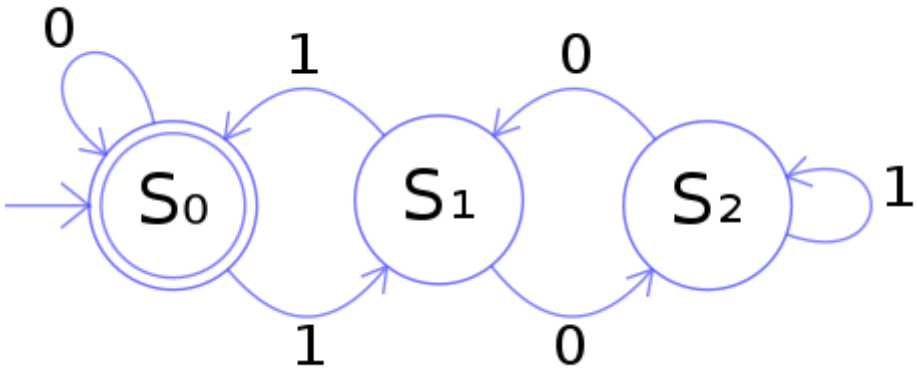

Homework 1: scanner

Instructor: Zhiyao Liang
Macau University of Science and Technology
2020 FAll

Introduction

We will construct a *lexical analyzer*, also called a *scanner* for the specified source language.

The source language

The source language is documented in the file `Pym_language.pdf` . Some sample programs of this language are in the folder `Sample_Pym_programs` .

Implementing the scanner

- Use the algorithm of scanner discussed in class, which is also documented in the file `simple_scanner_algorithm.pdf` .
- Do not use some automatically generated scanner code.
- You can use C or C++ to implement the scanner, but C is recommended, because the helping code given by the teacher will be C code.
- Design your program with reasonable modules. A possible choice is having the following files
  - `tools.h` : inclcude the declarations of tools for general purpose.
  - `tools.c` : the definitions of general-purpose tools.
  - `token.h` : declaration of token types, and related function prototypes.

- `token.c` : token related function definitions.
- `list.h` : declarations related to the list data structure.
- `list.c` : function definitions for the list data structure.
- `automata.h` : declaration of some DFA.
- `automata.c` : definitions for the DFA. The DFA can be defined as a function according to the *simple scanner algorithm*.
- `scanner.c` : the `main` function, the driver of the program.

Expected behavior of the scanner

- Test your scanner with the provided some Pym programs as the source. You can use the programs in the folder `Sample_Pym_programs` . Or you can use some Pym program that you create.
- When a source program is correct,
  - some token list should be generated, which will be used by the parser in the future.
  - The generated token list should be printed.
    - For each token, its token type should be printed. If the string cannot be decided by by its type, like ID or Number, or String the string should also be printed.
- When some error occurs, the corresponding error message should be printed. At least, one error message is printed for the first lexical error found by the compiler.

For example, suppose a Pym source file `sample.pym` contains the following code

```
temperature = 115
while temperature > 3\7: # first while loop code
    printnum(temperature)
    printstr(" is too hot.\n")
    temperature = temperature - 1
printstr("The tea is cool enough.")
```

When the scanner runs, the screen printing could be like the sample run listed below, where the lines starting with a smile face `:)` is a sentence output by the program in the conversation, and line starting with `>>>` is input from the user; each token is printed in a line, where the starting line number is optional.

```
:)  Hello, what is the name of the Pym source file?
>>> sample.pym.

:) The listed of tokens are printed as follows:

1 ID temperature
1 ASSN
1 NUM 115
1 NEWLINE
2 WHILE
2 ID temperature
2 GT
2 NUM 3\7
2 COLON
2 NEWLINE
3 INDENT
3 ID printnum
3 LPR
3 ID temperature
3 RPR
3 NEWLINE
4 ID printstr
4 LPR
4 STR " is too hot.\n"
4 RPR
4 NEWLINE
5 ID temperature
5 ASSN
5 ID temperature
5 SUB
5 NUM 1
5 NEWLINE
6 DEDENT
6 ID printstr
6 LPR
6 STR "The tea is cool enough."
6 RPR
6 NEWLINE
6 EOF
```

Note that you can choose to let your program to accept the source file name as a command-line argument. For example, if the executable file of your scanner is `myscanner`, you can do:

```
myscanner sample.pym
```

How to submit

- Upload your files at the webpage address of this homework on Moodle.
- The uploaded files can include only the following:
  - The source files (.c, .cpp or .h files),
  - An optional document file (.txt, .md, .docx) describing your work, like: what features of the homework are achieved; what are the remaining problems; how did you solve the difficulties that you met, some screen record of compiling and running the program; the bonus features that you have implemented, ...
- About homework submission in a group:
  - At most 3 students can form a group to submit the homework together. You can surely do the homework alone.
  - One group only need to submit the homework by once by one student. The other members do not need to (better not) submit the homework again; or, just submit one .txt file saying who are the members of the group and who submitted the homework.
  - In each file that you submit, record the registered names in Chinese (if you are not an international student) and English letters (PinYin), classes of each student, last 5 digits of student ID, as comments. For example:

```
/*   homework 1. Group members:
 李白      Li, Bai
 杜甫      Du, Fu
 李清照    Li, QingZhao
*/
```

- Deadline: Saturday Nov 14 2020
- Plagiarism is not allowed

References

1. "Among the Automata" https://www.theparisreview.org/blog/2012/05/22/among-the-automata/