

Homework 1

Readme of

Simple Java – Scanner

Chih-Hsuan Yang

National Sun-Yet-San University, Taiwan

Instructor: Ye-In Chang

March 28, 2021

Contents

1	Lex version	3
2	Operating System	3
3	Execution manual	3
3.1	Build	3
3.2	Run	3
4	How to implement	3
5	Problems	4
6	Outcome	5

1 Lex version

```
$ lex -version  
flex 2.6.4
```

2 Operating System

```
$ uname -voms  
  
Linux #1 SMP Debian 4.19.171-2 (2021-01-30) x86_64 GNU/Linux
```

3 Execution manual

3.1 Build

Symbol table (Built, but if you want to compile by yourself)

```
$ cd lib && mkdir build && cd build  
$ cmake ..  
$ make  
$ mv libsymbol_table.a ..
```

Lexer

```
$ flex -o lex.c demo.l  
$ gcc -c lex.c  
$ g++ lex.o lib/libsymbol_table.a -lfl -o my_lex.elf
```

L^AT_EX

```
$ cd Readme  
$ make
```

3.2 Run

```
$ # Build it first  
$ ./my_lex.elf < input.java
```

4 How to implement

Regex in Lex and `std::unordered_set` in C++

5 Problems

1. About the symbol table: the requirement has said:

create() 建立一個symbol table。
lookup(s) 傳回字串s的index；假如s沒找到的話，就傳回-1。
insert(s) 新增s到symbol table中，並傳回存放位置的index。
dump() 將symbol table中所有的資料印出。

- (a) About the lookup(s) function:

Why you indexing? And how you indexing?

The symbol table is a set mathematically. A set does not have any argument is called index.

However, the requirement asked:

lookup(s) 傳回字串s的index；假如s沒找到的話，就傳回-1。

Why?

How can I return a index of a set?

What is the index of an element in a set?

- (b) About the create() function:

This is an implementation issue.

According to the RAII pattern: In RAII, holding a resource is a class invariant, and is tied to object lifetime: resource allocation (or acquisition) is done during object creation (specifically initialization), by the constructor, while resource deallocation (release) is done during object destruction (specifically finalization), by the destructor.

Therefore, request a function to construct an object, but no destruct is unreasonable. It would cause the memory leak issue.

- (c) About the dump() function:

Well, this is a naming issue.

We use the word "dump" in programming, that means some error(s) happened, and we would get some information about the "crash" event. However, there is no "crash" happened while we were printing the symbol table, right?

Hence, we shouldn't take the "dump" to be the function name.

For clean code, we should named it as "export" or "view". Which give users a decision to handle the exported table, no matter they want to "print", "store" etc.

Again, we shouldn't use "dump" here.

2. The GNU Flex does not fully support POSIX:

I tried `[[:alpha:]]` for `[A-Za-z]` in regex. However, the flex does not support.

And I tried the dollar sign `'$'` for the "End-of-line". However the flex does not support too.

3. The test file is DOS format:

Well, this is not a mistake. I just not want to precessing these formatting issue. Every UNIX(-like) system are using `'\n'` as newline. I don't want to add a special for DOS format.

6 Outcome

```
rm -rf lib/build lex.o my_lex.elf lex.c
flex -o lex.c demo.l
gcc -c lex.c
g++ lex.o lib/libsymbol_table.a -lfl -o my_lex.elf
./my_lex.elf < TestFile_Lab1_2021/Test1.java
Line: 2, 1st char: 1, "public" is a "reserved word".
Line: 2, 1st char: 8, "class" is a "reserved word".
Line: 2, 1st char: 14, "Test1" is a "ID".
Line: 2, 1st char: 20, "{" is a "symbol".
Line: 3, 1st char: 5, "public" is a "reserved word".
Line: 3, 1st char: 12, "static" is a "reserved word".
Line: 3, 1st char: 19, "int" is a "reserved word".
Line: 3, 1st char: 23, "add" is a "ID".
Line: 3, 1st char: 26, "(" is a "symbol".
Line: 3, 1st char: 27, "int" is a "reserved word".
Line: 3, 1st char: 31, "a" is a "ID".
Line: 3, 1st char: 32, "," is a "symbol".
Line: 3, 1st char: 34, "int" is a "reserved word".
Line: 3, 1st char: 38, "b" is a "ID".
Line: 3, 1st char: 39, ")" is a "symbol".
Line: 3, 1st char: 41, "{" is a "symbol".
Line: 4, 1st char: 9, "return" is a "reserved word".
Line: 4, 1st char: 16, "a" is a "ID".
Line: 4, 1st char: 18, "+" is a "operator".
Line: 4, 1st char: 20, "b" is a "ID".
Line: 4, 1st char: 21, ";" is a "symbol".
Line: 5, 1st char: 5, "}" is a "symbol".
Line: 7, 1st char: 5, "public" is a "reserved word".
Line: 7, 1st char: 12, "static" is a "reserved word".
Line: 7, 1st char: 19, "void" is a "reserved word".
Line: 7, 1st char: 24, "main" is a "reserved word".
Line: 7, 1st char: 28, "(" is a "symbol".
Line: 7, 1st char: 29, ")" is a "symbol".
Line: 7, 1st char: 31, "{" is a "symbol".
Line: 9, 1st char: 9, "int" is a "reserved word".
Line: 9, 1st char: 13, "c" is a "ID".
Line: 9, 1st char: 14, ";" is a "symbol".
Line: 10, 1st char: 9, "int" is a "reserved word".
Line: 10, 1st char: 13, "a" is a "ID".
Line: 10, 1st char: 15, "=" is a "operator".
Line: 10, 1st char: 17, "5;" is a "float".
Line: 11, 1st char: 9, "c" is a "ID".
Line: 11, 1st char: 11, "=" is a "operator".
```

```

Line: 11, 1st char: 13, "add" is a "ID".
Line: 11, 1st char: 16, "(" is a "symbol".
Line: 11, 1st char: 17, "a" is a "ID".
Line: 11, 1st char: 18, "," is a "symbol".
Line: 11, 1st char: 20, "10)" is a "float".
Line: 11, 1st char: 23, ";" is a "symbol".
Line: 12, 1st char: 9, "if" is a "reserved word".
Line: 12, 1st char: 12, "(" is a "symbol".
Line: 12, 1st char: 13, "c" is a "ID".
Line: 12, 1st char: 15, ">" is a "operator".
Line: 12, 1st char: 17, "10)" is a "float".
Line: 13, 1st char: 13, "print" is a "reserved word".
Line: 13, 1st char: 18, "(" is a "symbol".
Line: 13, 1st char: 19, "" is a "error".
Line: 13, 1st char: 20, "c" is a "ID".
Line: 13, 1st char: 22, "=" is a "operator".
Line: 13, 1st char: 24, "" is a "error".
Line: 13, 1st char: 26, "+" is a "operator".
Line: 13, 1st char: 28, "-" is a "operator".
Line: 13, 1st char: 29, "c" is a "ID".
Line: 13, 1st char: 30, ")" is a "symbol".
Line: 13, 1st char: 31, ";" is a "symbol".
Line: 14, 1st char: 9, "else" is a "reserved word".
Line: 15, 1st char: 13, "print" is a "reserved word".
Line: 15, 1st char: 18, "(" is a "symbol".
Line: 15, 1st char: 19, "c" is a "ID".
Line: 15, 1st char: 20, ")" is a "symbol".
Line: 15, 1st char: 21, ";" is a "symbol".
Line: 16, 1st char: 9, "print" is a "reserved word".
Line: 16, 1st char: 14, "(" is a "symbol".
Line: 16, 1st char: 15, "" is a "error".
Line: 16, 1st char: 16, "Hello" is a "ID".
Line: 16, 1st char: 22, "World" is a "ID".
Line: 16, 1st char: 27, "" is a "error".
Line: 16, 1st char: 28, ")" is a "symbol".
Line: 16, 1st char: 29, ";" is a "symbol".
Line: 18, 1st char: 5, "}" is a "symbol".
Line: 20, 1st char: 1, "}" is a "symbol".
The symbol table contains:
add
World
Test1
Hello
b
a
c
./my_lex.elf < TestFile_Lab1-2021/Test2.java
Line: 1, 1st char: 1, "// this is a comment // line */ /* with /* delimiters */ before the end
" is a "comment".
Line: 3, 1st char: 1, "public" is a "reserved word".
Line: 3, 1st char: 8, "class" is a "reserved word".
Line: 3, 1st char: 14, "Test2" is a "ID".
Line: 3, 1st char: 20, "{" is a "symbol".
Line: 4, 1st char: 5, "int" is a "reserved word".
Line: 4, 1st char: 9, "i" is a "ID".
Line: 4, 1st char: 11, "=" is a "operator".
Line: 4, 1st char: 13, "-100;" is a "float".
Line: 5, 1st char: 5, "double" is a "reserved word".
Line: 5, 1st char: 12, "d" is a "ID".
Line: 5, 1st char: 14, "=" is a "operator".
Line: 5, 1st char: 16, "12.25e+6;" is a "float".
Line: 7, 1st char: 5, "public" is a "reserved word".

```

```

Line: 7, 1st char: 12, "static" is a "reserved word".
Line: 7, 1st char: 19, "void" is a "reserved word".
Line: 7, 1st char: 24, "main" is a "reserved word".
Line: 7, 1st char: 28, "(" is a "symbol".
Line: 7, 1st char: 29, ")" is a "symbol".
Line: 7, 1st char: 31, "{" is a "symbol".
Line: 8, 1st char: 1, "/* this is a comment // line with some /* and
// delimiters */" is a "comment".
Line: 10, 1st char: 5, "}" is a "symbol".
Line: 11, 1st char: 1, "}" is a "symbol".
The symbol table contains:
d
Test2
i
./my_lex.elf < TestFile_Lab1_2021/Test3.java
Line: 2, 1st char: 1, "public" is a "reserved word".
Line: 2, 1st char: 8, "class" is a "reserved word".
Line: 2, 1st char: 14, "Test3" is a "ID".
Line: 2, 1st char: 20, "{" is a "symbol".
Line: 3, 1st char: 5, "int" is a "reserved word".
Line: 3, 1st char: 9, "A" is a "ID".
Line: 3, 1st char: 10, ";" is a "symbol".
Line: 4, 1st char: 5, "int" is a "reserved word".
Line: 4, 1st char: 9, "a" is a "ID".
Line: 5, 1st char: 5, "double" is a "reserved word".
Line: 5, 1st char: 12, "b" is a "ID".
Line: 5, 1st char: 13, ";" is a "symbol".
Line: 6, 1st char: 5, "double" is a "reserved word".
Line: 6, 1st char: 12, "A" is a "ID".
Line: 6, 1st char: 13, ";" is a "symbol".
Line: 8, 1st char: 5, "public" is a "reserved word".
Line: 8, 1st char: 12, "Test3" is a "ID".
Line: 8, 1st char: 17, "(" is a "symbol".
Line: 8, 1st char: 18, ")" is a "symbol".
Line: 8, 1st char: 20, "{" is a "symbol".
Line: 9, 1st char: 9, "a" is a "ID".
Line: 9, 1st char: 11, "=" is a "operator".
Line: 9, 1st char: 13, "1;" is a "float".
Line: 10, 1st char: 9, "A" is a "ID".
Line: 10, 1st char: 11, "=" is a "operator".
Line: 10, 1st char: 13, "2;" is a "float".
Line: 11, 1st char: 9, "b" is a "ID".
Line: 11, 1st char: 11, "=" is a "operator".
Line: 11, 1st char: 13, "-1.2;" is a "float".
Line: 12, 1st char: 5, "}" is a "symbol".
Line: 13, 1st char: 1, "}" is a "symbol".
The symbol table contains:
b
Test3
a
A

```