Compiler Hw1 Readme

Cheng-Han Hsieh, 謝承翰

April 11, 2024

1 Environment

kernel name	Linux
kernel version	5.15.0-101-generic
processor type	x86_64
hardware platform	x86_64
architecture	x86_64
system version	#111 20.04.1-Ubuntu
flex version	2.6.4
GNU gcc version	13.1.0
GNU g++ version	13.1.0
GNU Make version	4.2.1
GNU ar version	2.34

Table 1: Environment

2 How to execute

Type `make` and then type `./lex.elf < your_file`.

3 Method

1. Since Pascal is case-insensitive, the flag `-i` for flex is set.
2. The regular expression for identity is `[A-Za-z_][_[:alnum:]]0,14`. And the regular expression for invalid identity is `[A-Za-z_][_[:alnum:]]{15,}` and `[0-9^#%$][_[:alnum:]^#%$]*`.
3. The regular expression for symbols is `;|:|(|)|:=|>|<|=|==|>=|<=|[|]|+|-|*|/`.
4. The regular expression for real number is `{int}(unsigned)?([eE]int)?` where `{int}` is `[+-]?[0-9]+` and `{unsigned}` is `[0-9]+`.
5. The regular expression for string is `'([^'\r\n]|''){0,30}'`.

6. I use start state to deal with comments. When match the string `(*`, flex enters into COMMENT status and process the content of comment until hit the string `*)`.

7. After matching the invalid token, the program will print out the invalid token and the reasons.

4 Problems that I confronted

1. How to use `std::unordered_set` to implement symbol table in C? (namely, how to link C++ std lib in C)

2. How to consume the comment?

3. How to tokenize the expression that has no space? (e.g. 1+2 is integer, symbol, integer, instead of integer, integer)

5 Result

1.pas

```
(base)  ch0923@231  ~/compiler/hw1  master  ./lex.elf < testfile_lab1_2022/test\ data/1.pas
Line: 1, 1st char: 1, "program" is a "reserved word".
Line: 1, 1st char: 9, "test" is a "identity".
Line: 1, 1st char: 13, ";" is a "symbol".
Line: 2, 1st char: 1, "var" is a "reserved word".
Line: 3, 1st char: 3, "i" is a "identity".
Line: 3, 1st char: 5, ":" is a "symbol".
Line: 3, 1st char: 7, "integer" is a "reserved word".
Line: 3, 1st char: 14, ";" is a "symbol".
Line: 4, 1st char: 1, "begin" is a "reserved word".
Line: 5, 1st char: 3, "read" is a "reserved word".
Line: 5, 1st char: 7, "(" is a "symbol".
Line: 5, 1st char: 8, "i" is a "identity".
Line: 5, 1st char: 9, ")" is a "symbol".
Line: 5, 1st char: 10, ";" is a "symbol".
Line: 6, 1st char: 1, "end" is a "reserved word".
Line: 6, 1st char: 4, ";" is a "symbol".
symbol table:
i
test
```

Figure 1: The result of 1.pas

2.pas

```
(base)  ch0923@231  ~/compiler/hw1  master  ./lex.elf < testfile_lab1_2022/test\ data/2.pas
Line: 1, 1st char: 1, "program" is a "reserved word".
Line: 1, 1st char: 9, "test" is a "identity".
Line: 1, 1st char: 13, ";" is a "symbol".
Line: 2, 1st char: 1, "var" is a "reserved word".
Line: 3, 1st char: 3, "3i" is a "invalid identity".
    3 | ERROR: Starts with the invalid character: 3.
Line: 3, 1st char: 6, ":" is a "symbol".
Line: 3, 1st char: 8, "string" is a "reserved word".
Line: 3, 1st char: 14, ";" is a "symbol".
Line: 4, 1st char: 1, "begin" is a "reserved word".
Line: 5, 1st char: 3, "3i" is a "invalid identity".
    5 | ERROR: Starts with the invalid character: 3.
Line: 5, 1st char: 6, ":=" is a "symbol".
Line: 5, 1st char: 9, "'ab;" is a "invalid string".
    5 | ERROR: missing terminating ' character.
Line: 6, 1st char: 1, "end" is a "reserved word".
Line: 6, 1st char: 4, ";" is a "symbol".
symbol table:
test
```

Figure 2: The result of 2.pas

3.pas

```
(base)  ch0923@231  ~/compiler/hw1  master  ./lex.elf < testfile_lab1_2022/test\ data/3.pas
Line: 2, 1st char: 1, "(* comment 1
   comment 2 *)" is a "comment".
Line: 3, 1st char: 1, "program" is a "reserved word".
Line: 3, 1st char: 9, "test" is a "identity".
Line: 3, 1st char: 13, ";" is a "symbol".
Line: 4, 1st char: 1, "var" is a "reserved word".
Line: 5, 1st char: 3, "i" is a "identity".
Line: 5, 1st char: 5, ":" is a "symbol".
Line: 5, 1st char: 7, "integer" is a "reserved word".
Line: 5, 1st char: 14, ";" is a "symbol".
Line: 6, 1st char: 1, "begin" is a "reserved word".
Line: 7, 1st char: 3, "read" is a "reserved word".
Line: 7, 1st char: 7, "(" is a "symbol".
Line: 7, 1st char: 8, "i" is a "identity".
Line: 7, 1st char: 9, ")" is a "symbol".
Line: 7, 1st char: 10, ";" is a "symbol".
Line: 8, 1st char: 1, "end" is a "reserved word".
Line: 8, 1st char: 4, ";" is a "symbol".
symbol table:
i
test
```

Figure 3: The result of 3.pas

4.pas

```
(base)  ch0923@231  ~/compiler/hw1  master  ./lex.elf < testfile_lab1_2022/test\ data/4.pas
Line: 1, 1st char: 1, "program" is a "reserved word".
Line: 1, 1st char: 9, "test" is a "identity".
Line: 1, 1st char: 13, ";" is a "symbol".
Line: 2, 1st char: 1, "var" is a "reserved word".
Line: 3, 1st char: 3, "f" is a "identity".
Line: 3, 1st char: 5, ":" is a "symbol".
Line: 3, 1st char: 7, "float" is a "reserved word".
Line: 3, 1st char: 12, ";" is a "symbol".
Line: 4, 1st char: 1, "begin" is a "reserved word".
Line: 5, 1st char: 3, "f" is a "identity".
Line: 5, 1st char: 5, ":=" is a "symbol".
Line: 5, 1st char: 8, "12.25e+6" is a "float".
Line: 5, 1st char: 16, ";" is a "symbol".
Line: 6, 1st char: 1, "end" is a "reserved word".
Line: 6, 1st char: 4, ";" is a "symbol".
symbol table:
f
test
```

Figure 4: The result of 4.pas

5.pas

```
Line: 1, 1st char: 1, "(* a**b) *)" is a "comment".
Line: 2, 1st char: 1, "program" is a "reserved word".
Line: 2, 1st char: 9, "test" is a "identity".
Line: 2, 1st char: 13, ";" is a "symbol".
Line: 3, 1st char: 1, "var" is a "reserved word".
Line: 4, 1st char: 3, "i" is a "identity".
Line: 4, 1st char: 5, ":" is a "symbol".
Line: 4, 1st char: 7, "integer" is a "reserved word".
Line: 4, 1st char: 14, ";" is a "symbol".
Line: 5, 1st char: 3, "_s" is a "identity".
Line: 5, 1st char: 5, "," is a "symbol".
Line: 5, 1st char: 7, "_s2" is a "identity".
Line: 5, 1st char: 10, "," is a "symbol".
Line: 5, 1st char: 12, "_s3" is a "identity".
Line: 5, 1st char: 15, "," is a "symbol".
Line: 5, 1st char: 17, "_s4" is a "identity".
Line: 5, 1st char: 20, "," is a "symbol".
Line: 5, 1st char: 22, "_s5" is a "identity".
Line: 5, 1st char: 26, ":" is a "symbol".
Line: 5, 1st char: 28, "string" is a "reserved word".
Line: 5, 1st char: 34, ";" is a "symbol".
Line: 6, 1st char: 1, "begin" is a "reserved word".
Line: 7, 1st char: 3, "i" is a "identity".
Line: 7, 1st char: 5, ":=" is a "symbol".
Line: 7, 1st char: 8, "-100" is a "integer".
Line: 7, 1st char: 12, ";" is a "symbol".
Line: 8, 1st char: 3, "_s" is a "identity".
Line: 8, 1st char: 6, ":=" is a "symbol".
Line: 8, 1st char: 9, "'db lab'" is a "string".
Line: 8, 1st char: 17, ";" is a "symbol".
Line: 9, 1st char: 3, "_s2" is a "identity".
Line: 9, 1st char: 7, ":=" is a "symbol".
Line: 9, 1st char: 10, "'You''ll see'" is a "string".
Line: 9, 1st char: 23, ";" is a "symbol".
Line: 10, 1st char: 3, "_s3" is a "identity".
Line: 10, 1st char: 7, ":=" is a "symbol".
Line: 10, 1st char: 10, "''" is a "string".
Line: 10, 1st char: 12, ";" is a "symbol".
Line: 11, 1st char: 3, "_s4" is a "identity".
Line: 11, 1st char: 7, ":=" is a "symbol".
```

Figure 5: The result of 5.pas

Figure 6: The result of 5.pas (cont.)

6.pas

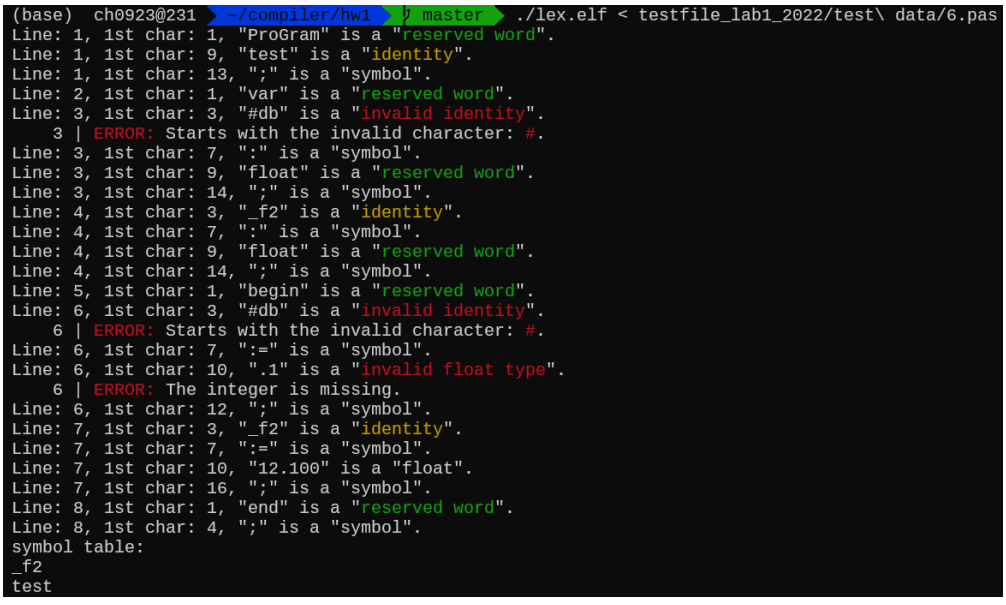

Figure 7: The result of 6.pas

7.pas

```
(base)  ch0923@231  ~/compiler/hw1  master  ./lex.elf < testfile_lab1_2022/test\ data/7.pas
Line: 1, 1st char: 1, "(* This line is a comment. *)" is a "comment".
Line: 2, 1st char: 1, "program" is a "reserved word".
Line: 2, 1st char: 9, "test" is a "identity".
Line: 2, 1st char: 13, ";" is a "symbol".
Line: 3, 1st char: 1, "var" is a "reserved word".
Line: 4, 1st char: 3, "i" is a "identity".
Line: 4, 1st char: 5, ":" is a "symbol".
Line: 4, 1st char: 7, "integer" is a "reserved word".
Line: 4, 1st char: 14, ";" is a "symbol".
Line: 5, 1st char: 1, "begin" is a "reserved word".
Line: 6, 1st char: 3, "i" is a "identity".
Line: 6, 1st char: 5, ":=" is a "symbol".
Line: 6, 1st char: 8, "1" is a "integer".
Line: 6, 1st char: 9, "+" is a "symbol".
Line: 6, 1st char: 10, "2" is a "integer".
Line: 6, 1st char: 11, ";" is a "symbol".
Line: 7, 1st char: 1, "end" is a "reserved word".
Line: 7, 1st char: 4, ";" is a "symbol".
symbol table:
i
test
```

Figure 8: The result of 7.pas