

Compiler HW1

Simple java scanner

Flex version :

- 2.6.4

Running Platform :

- Ubuntu 20.04.2 LTS

How to run :

```
make clean all
./main < testfile.java
```

How the program is done :

- Lexical definitions :
 - Symbols, Reserved word, Identifiers, Integer, Float, String
 - Whitespace, Comments
 - `commentnewline()` will count the newline char '\n' in the comment
- Symbol table :

A global variable `table[100]` is defined as symbol table, it contains 100 struct object

```
struct sym{
    char *data;
};
```

- **`create()`**: init the table
- **`lookup(s)`**: return the index of string `s` in table. If not found, return -1
- **`insert(s,len)`**: insert string `s` with length `len` to the table, return the index
- **`dump()`**: output value in table

Problems I met :

- complex c-style comments are hard to identify


eg :


```
/*comment */ int mycode /*more comment*/
```

Testfile outputs :

```
18:50:53 ...NSYSU_CSE_2021/Compiler/HW1 v10.19.0 master x
$ ./main < 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 an "identifier" .
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 an "identifier" .
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 an "identifier" .
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 an "identifier" .
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 an "identifier" .
Line: 4, 1st char: 18, "+" is an "operator" .
Line: 4, 1st char: 20, "b" is an "identifier" .
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 an "identifier" .
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 an "identifier" .
Line: 10, 1st char: 15, "=" is an "operator" .
Line: 10, 1st char: 17, "5" is an "integer" .
Line: 10, 1st char: 18, ";" is a "symbol" .
Line: 11, 1st char: 9, "c" is an "identifier" .
Line: 11, 1st char: 11, "=" is an "operator" .
Line: 11, 1st char: 13, "add" is an "identifier" .
Line: 11, 1st char: 16, "(" is a "symbol" .
Line: 11, 1st char: 17, "a" is an "identifier" .
Line: 11, 1st char: 18, "," is a "symbol" .
Line: 11, 1st char: 20, "10" is an "integer" .
Line: 11, 1st char: 22, ")" is a "symbol" .
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 an "identifier" .
```

```
Line: 12, 1st char: 15, ">" is an "operator" .
Line: 12, 1st char: 17, "10" is an "integer" .
Line: 12, 1st char: 19, ")" is a "symbol" .
Line: 13, 1st char: 13, "print" is a "reserved word" .
Line: 13, 1st char: 18, "(" is a "symbol" .
Line: 13, 1st char: 19, "c = " is a "string" .
Line: 13, 1st char: 26, "+" is an "operator" .
Line: 13, 1st char: 28, "-" is an "operator" .
Line: 13, 1st char: 29, "c" is an "identifier" .
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 an "identifier" .
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, "Hello World" is a "string" .
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:
Test1
add
a
b
c
```

 Captured with Xnip
19:06:48 ...NSYSU_CSE_2021/Compiler/HW1 v10.19.0 master x ● ★
\$./main < 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 an "identifier" .
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 an "identifier" .
Line: 4, 1st char: 11, "=" is an "operator" .
Line: 4, 1st char: 13, "-100" is an "integer" .
Line: 4, 1st char: 17, ";" is a "symbol" .
Line: 5, 1st char: 5, "double" is a "reserved word" .
Line: 5, 1st char: 12, "d" is an "identifier" .
Line: 5, 1st char: 14, "=" is an "operator" .
Line: 5, 1st char: 16, "12.25e+6" is a "float" .
Line: 5, 1st char: 24, ";" 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: 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:
Test2

 Captured with Xnip
19:07:01 ...NSYSU_CSE_2021/Compiler/HW1 v10.19.0 master x ● ★
\$./main < 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 an "identifier" .
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 an "identifier" .
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 an "identifier" .
Line: 5, 1st char: 5, "double" is a "reserved word" .
Line: 5, 1st char: 12, "b" is an "identifier" .
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 an "identifier" .
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 an "identifier" .
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 an "identifier" .
Line: 9, 1st char: 11, "=" is an "operator" .
Line: 9, 1st char: 13, "1" is an "integer" .
Line: 9, 1st char: 14, ";" is a "symbol" .
Line: 10, 1st char: 9, "A" is an "identifier" .
Line: 10, 1st char: 11, "=" is an "operator" .
Line: 10, 1st char: 13, "2" is an "integer" .
Line: 10, 1st char: 14, ";" is a "symbol" .
Line: 11, 1st char: 9, "b" is an "identifier" .
Line: 11, 1st char: 11, "=" is an "operator" .
Line: 11, 1st char: 13, "-1.2" is a "float" .
Line: 11, 1st char: 17, ";" is a "symbol" .
Line: 12, 1st char: 5, "}" is a "symbol" .
Line: 13, 1st char: 1, "}" is a "symbol" .
The symbol table contains:
Test3
A
a
b