

report

学号	姓名
20337091	马佳欣

实验过程

1. 在 `lexer/lexer.l` 模板的基础上完成更多的 `token`：

注：实验模板欠缺一下 `token`：

```
1 token      name
2
3 float      float
4 ...        ellipsis
```

2. 在 `lexer/lexer.l` 中增加 `genflag()` 函数，处理无关字符 `[StartOfLine]` `[LeadingSpace]`：

思路：增加全局变量 `yynewline`、`yynewspace` 用于标记新行、新空格，分情况设置 `flag`：

```
1  if(yynewline && yynewspace){
2      flag = " [StartOfLine] [LeadingSpace]";
3  }
4  else if(yynewline){
5      flag = " [StartOfLine]";
6  }
7  else if(yynewspace){
8      flag = " [LeadingSpace]";
9  }
10 else{
11     flag = "";
12 }
```

3. 挑战方向：扩展更多 C 语言的 `token`

- 踩过的坑：
 - 不支持数据类型 `bool`，`Clang` 将其识别为 `identifier`；
 - 不需要写 `long long` 类型对应的规则，`Clang` 将其识别为两个 `long`。
- 更多 `token`：

```
1 token      name
2
3 :          colon
4 auto       auto
5 double     double
6 long       long
7 enum       enum
8 register   register
9 short      short
10 signed     signed
```

11	unsigned	unsigned
12	static	static
13	struct	struct
14	union	union
15	typedef	typedef
16	do	do
17	switch	switch
18	case	case
19	default	default
20	for	for
21	goto	goto
22	sizeof	sizeof
23	extern	extern
24	++	plusplus
25	+=	plusequal
26	--	minusminus
27	-=	minusequal
28	*=	starequal
29	/=	slashequal
30	%=	percentequal

4. 根据新扩展的 `token`, 编写新的测例 `tester/mytest.c` 和评测脚本 `lexer/mytest.sh`。

难点

设计字符串的正则表达式

1. 初始版本 `"[^"]"`:

- 思路: 不允许字符串中单独出现 `"`
- 缺点: 出现 `\` 不被允许

2. 最终版本 `"([^\\""]|\\.)*"`

- 思路: | 左边规定: 不允许单独出现 `"` 和 `\`; 右边规定: `\` 的后面需要带任意字符 (`\n` 除外)

设计小数的正则表达式

总共有三种表示:

- 十进制小数点计数法
- e-计数法
- p-计数法

自动评测结果

0

The screenshot displays the Visual Studio Code interface. The Explorer sidebar on the left shows the project structure for 'wukun0621/sysu-lang'. The 'lexertest.sh' script is selected, showing its contents in the main editor. The script sets environment variables for PATH, CPATH, LIBRARY_PATH, and LD_LIBRARY_PATH, and then runs a series of 'sysu-compiler' commands with various flags, followed by 'echo pass all!'. The TERMINAL panel at the bottom shows the output of the script, which consists of a long list of test results, each starting with 'tester/third_party/SYSU-lang-tester-performance/performance_test2022-public/instruction-combining-1.sysu.c' and ending with a specific test identifier (e.g., 's12.sysu.c'). The status bar at the bottom indicates the current file is 'Container wukun0621/sysu-lang...' and the terminal is running 'Ln, Col 17'.

新的测例评测结果

```
$ mytest.sh X
SYsU-lang > lexer > $ mytest.sh
1  ( export PATH=$HOME/sysu/bin:$PATH \
2    CPATH=$HOME/sysu/include:$CPATH \
3    LIBRARY_PATH=$HOME/sysu/lib:$LIBRARY_PATH \
4    LD_LIBRARY_PATH=$HOME/sysu/lib:$LD_LIBRARY_PATH &&
5    sysu-compiler --unittest=lexer-0 "../tester/mytest.c" &&
6    sysu-compiler --unittest=lexer-1 "../tester/mytest.c" &&
7    sysu-compiler --unittest=lexer-2 "../tester/mytest.c" &&
8    sysu-compiler --unittest=lexer-3 "../tester/mytest.c" &&
9    echo mytest pass!)
```

PROBLEMS OUTPUT DEBUG CONSOLE TERMINAL PORTS

```
● root@21a56bf57171:/workspace/SYsU-lang/lexer# ./mytest.sh
[0/1] ../tester/mytest.c
[0/1] ../tester/mytest.c
[0/1] ../tester/mytest.c
[0/1] ../tester/mytest.c
mytest pass!
```