Compiler Hw1 Readme

Cheng-Han Hsieh, 謝承翰

April 6, 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)