Lexical Analysis

C Language

Reyner Marxell Arias Muñoz, Kenneth Ibarra Vargas, David Benavides Naranjo

Project 1, Compilers and Interpreters course, I 2022 Semester

April 26, 2022

Scanning

Since the source file has include and define directives, the preprocessor previously applied each of them correctly so that the scanner receives only a temporary input file.
Afterwards, the scanner goes through the temporary file returning the tokens one by one and in order, according to the parsed source program.
In c there are tokens of different types such as operators, identifiers, literals, reserved words and separator characters.
Any character not belonging to the c lexicon that can be parsed is returned as a lexical error.

FLEX

The scanner is a lex file, which are comprised of three sections:

- The Definition Section: This sections is made up of several regular expressions that act as global declarations that may be used in the next section.
- The Rules Section: This section uses the global declarations of the previous section to define what actions must be taken when a specific regular expression is found.
- The Code Section: This section is attached at the end of the lex output file and may contain any code written and executed by the C code, due to lex usually being paired with yacc.

Tokens

- *Operators*
- Intliterals
- **Floatliterals**
- ***Doubleliteral***
- *Charliteral*
- Stringliteral
- **Reserved Words**
- **Separator characters**
- Identifiers
- Errors

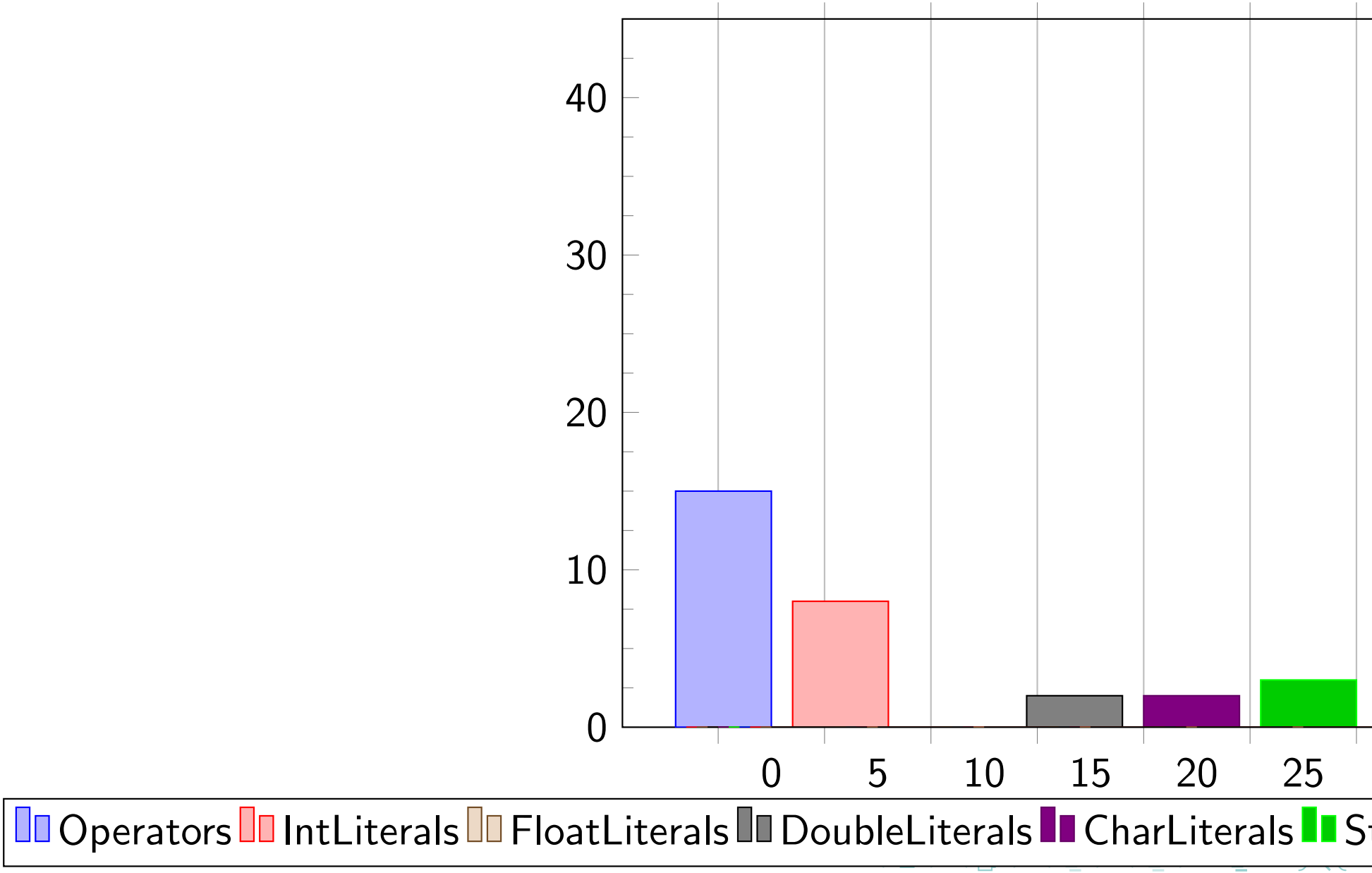
Font Lines

```
int test ( ) {
printf ( "%d" , 1 ) ;
}
int cinco = 5 ;
int main ( ) {
int t = 100301 ;
int r = hola ( ) ;
char f [ 10 ] = " ;
int v = 1 ;
double r = 2.435e2 ;
char r = 'y' ;
printf ( "%s" , f ) ;
}
int main ( ) {
int t = 100301 ;
int r = hola ( ) ;
```

Font Lines

```
char f [ 10 ] = '' ;
int v = 1 ;
double r = 2.435e2 ;
char r = 'y' ;
printf ( "%s" , f ) ;
}
```

Histogram


40
30
20
10
0
0
5
10
15
20
25
Operators
IntLiterals
FloatLiterals
DoubleLiterals
CharLiterals