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

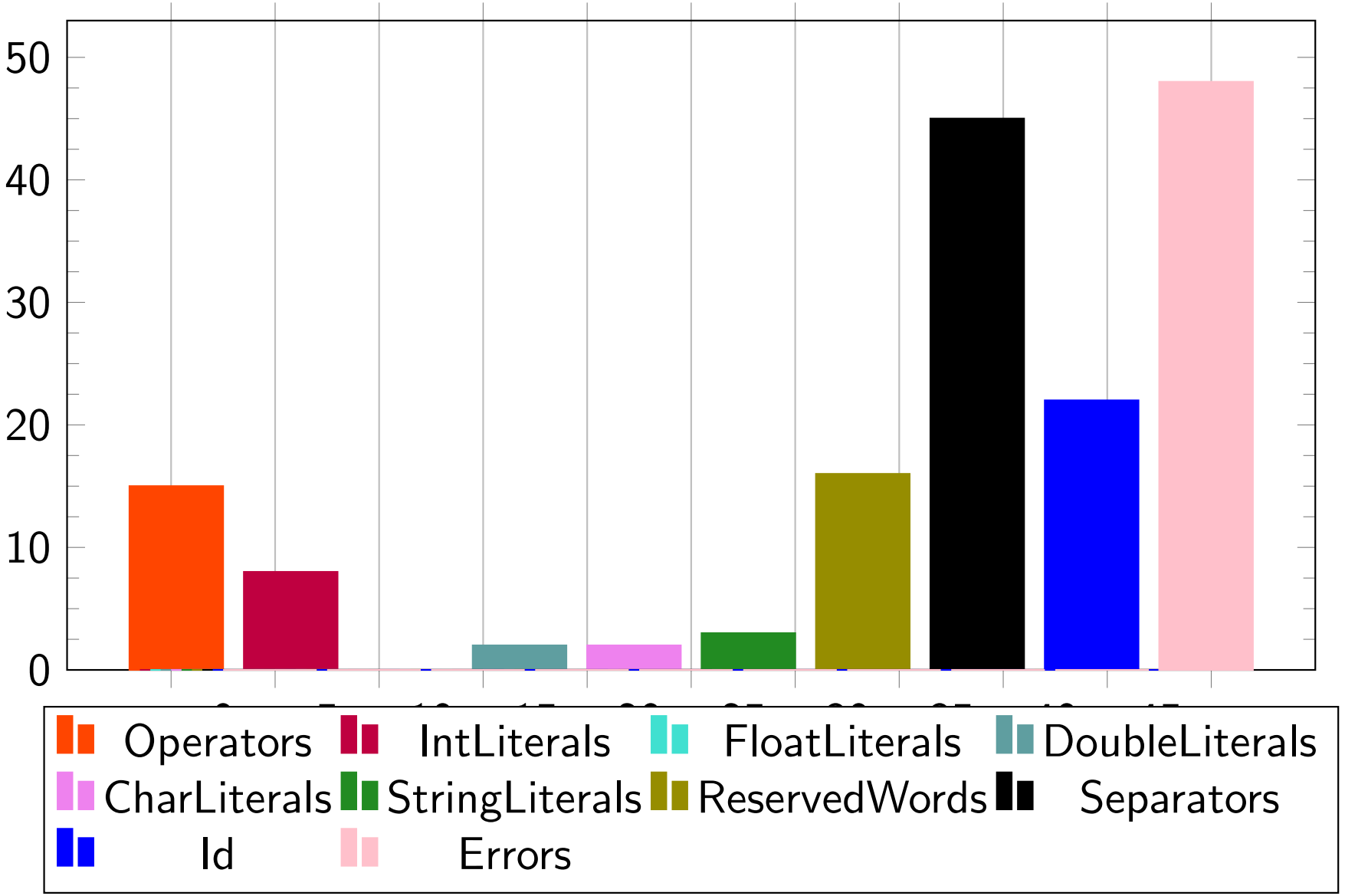
Font Lines

```
int _test(){
printf("%d", _1);
}
int _cinco_=_5;
int _main(){
int _t_=_100301;
int _r_=_1hola();
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


50
40
30
20
10
0
Operators
IntLiterals
FloatLiterals
DoubleLiterals
CharLiterals
StringLiterals
ReservedWords
Separators
Id
Errors