

Lexical Analysis

C Language

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

Project 1, Compilers and Interpreters course, I 2022 Semester

oday

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.



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

- The Definition Section: This section 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**