The used regex for identifiers is ^[a-zA-Z]([a-zA-Z]|[0-9])*$

and it only accepts strings that have at least

one letter on the first position and the any sequence of letter and digit characters.

The used regex for constants is ^0|[+-]?[1-9][0-9]*$|^\'.*\'$

and it only accepts constants that are 0 alone or a sequence of digits that has in front '+' or '-' sign or

neither. And then any sequence of digits or any as another string any sequence of any character, the latter

being supposed to emulate any string that can be given to the program.

As for the algorithm, the main idea is that at first the file with the code is parsed and each ant it is

tokenized. After all the tokens are found, the list with them is parsed and each and every one of them is

classified in identifiers, constant, reserved words, separators or operators, and if the token cannot

be classified in any of these categories, it will be classified as an lexical error.