https://github.com/915-Munteanu-Tudor/flcd/tree/main/lab3

In scanner class I keep a list for pif and another list for tokens and also two symbol tables: one for constants and one for identifiers.
I also have 3 methods that use regex: is_identifier, tokenize and is_constant. The first one matches all the identifiers, the "words" which begin with or small or capital characters then followed by the same or digits (0 or more). The words True and False are excepted (with a negative lookahead) because they are reserved to be constants. For constants, we match 0, +/- any number which doesn't start with 0, characters and strings and booleans. Tokenize method is used to split a string using anything that is not a sequence of letters, digits, quotes or strings, gets rid of ‘ ‘ and None, and puts together the composed relational operators like '<='.

The scan method opens the program file and parses it line by line which is split by ' ' ,tokenized then every element resulting from these splits is checked if it is token, identifier or constant (in this order) and is put in pif and in the corresponding symbol table.
If the respective element does not belong to any of these categories, it is raised a lexical error. After the scan is done, pif, and symbol table files are written if there is no error.