Github:
https://github.com/Aurelian-Iancu/UBB-Computer-Science/tree/main/Semester5/Formal%20Languages%20and%20Compiler%20Design/Lab2

Scanner:

The scanner class is responsible with building the program internal form and symbol table, while checking if the input program is lexically correct or not. The program internal form is a list composed from pairs of the form string and position in the symbol table, where string is the actual token, string const, int const or identifier. For the tokens, the position in the symbol table is considered to be (-1,-1). The scanner also has fields for the program string, a symbol table, an index and the current line of the program.

Operations:

- setProgram(string program) – setter for the program field
- readTokens() – read from the token.in file the tokens and separate them into reserved words and other tokens
- skipSpaces() – method to skip the spaces from the program and update the index accordingly
- treatStringConstant() -method to treat the case in which we have a string constant in the program; the method checks if the string is lexically correct (no invalid characters and quotes are closed correctly) and if it is, it adds it to the symbol table if it does not exist and to the program internal form, updates the index and returns true; if the string constant is invalid, the method returns false
- treatIntConstant() – method to treat the case in which we have an integer constant in the program; the method checks if the number is valid(no other characters except numbers, and it cannot start with 0 like 01) and if it is, it adds it to the symbol table if it does not exist and to the program internal form, updates the index and returns true; if the integer is invalid, the method returns false
- checkIfValid(string possible identifier, string programSubstring): - the method checks if a possible identifier is valid by checking if it is part of a declaration or if it is in the symbol table; if it does not satisfy any of the conditions, the possibleIdentifier is an invalid token and the method returns false; otherwise, the method returns true
- treatIdentifier() – method to treat the case in which we have an identifier in the program; the method checks if it is a valid identifier (starts with _ or letter, contains only letters, digits and _, is part of a declaration or a previously defined identifier) and if it is, it adds it to the symbol table if it does not exist and to the program internal form, updates the index and returns true; otherwise it returns false

- treatFromTokenList() – the method checks if the current element of the program is a reserved word or another token and if it is, it adds it to the program internal form and returns true; otherwise it returns false
- nextToken() – treats the current case and if there is no corresponding case, throw an exception that there is a lexical error at the current line and index
- scan(programFileName) – reads the program from the given file and checks for the next token until the end of the program; at the end, it writes the program internal form and symbol table to output files and displays a message that the program is lexically correct; if an exception is caught, it displays the message of that exception