Source code:
https://github.com/donpaul999/Assignments_UBB/tree/master/3rd%20Year/FLCD/Lab_3

The Symbol table is represented using a HashTable.

The PIF is kept as a simple list having pairs consisting of (token, position), where position is another pair representing the position of said token in the Symbol Table.

The tokenizing algorithm goes character by character on each line and checks whether the current we have so far is part of an operator, is a separator, begins a string or is building a constant or identifier, and then appends the tokens to a list which is returned.

The scanning algorithm splits each line of the program into tokens, and for each token if it's a constant or identifier, look up its position in the ST, if it's an operator/separator/reserved word, its position is (-1,-1).

Also, if it's a constant or identifier, instead of keeping the variable name/ constant value, it will be added into the PIF with the code "constant" or "id", respectively.

If the token is none of the above, that means we have a lexical error at that line, and the error is appended to the message.