

DOCUMENTATION SCANNER:

*PIF

-uses a list of tuples, which contain the element and -1 if they are a separator, operator or reserved word, and the position in the symbol table if they are a constant or an identifier

*REGEX:

-to check if one is an identifier it can be either one letter, or a sequence that has to start with a letter, then it can contain other letters or digits or the '_' sign

-to check if one is constant, it can either be 0 or be a digit or number that starts with a nonzero digit, that contains + or - optionally at the beginning of it. It can also be a string, which can be anything between the two ' ' signs

*FUNCTIONS:

-read_tokens reads the separators, operators, alphabet, reserved words from the tokens.in file

-is_in_operator checks if the element is in any of the operator's elements

-is_operator checks if the given string is an operator even if it's composed

-check_operators is for the cases when someone inputs something such as a string of operators (<!=+*=->), in which case the function checks each character one by one, if they are an operator themselves, they get appended into the operator list. If they are not an operator but they appear in a composed operator they can only be "!=" , so the function checks if they have an "=" sign following and if they do then the operator "!=" is appended to the list. If it is not even contained in one of the operands, it is added to the error list

-read_file creates a list with every line from the written program

-check_all_digits checks if the given string contains only digits using the tokens list of digits

-get_elements goes through the read file and appends to a list strings by category. if it finds an operator it keeps appending to the same string the following characters until it finds something that is not an operator. It does the same for the separators, words and digits (even if they are followed by words). If it is none of them it gets appended separately into the list

-get_by_category gets the elements from the read program, and goes through them checking if they are a separator, operator or reserved word, and adding them into a pif_list as a tuple of the string and -1. If they are a constant they get added in the pif_list as a tuple with -2, if they are an identifier they get added as a tuple with -3 and if it is an error it appends it to pif_list as a tuple with the members -4 and the position of the error.

-the main goes through each element got from calling the function get_by_category from the read file. In case the value on the second position of the tuple is -2, the element is appended to the constant's symbol table. In case it's -3, it gets appended to the identifier's symbol table. And in case it is -4, it is appended into the lexical error's output file. If it is -1 it gets appended to the pif list. Then the pif and each symbol table is written in its own output file

