Mini Assignment - 2

Roll No : AI20BTECH11006

10 features that you notice in the Lex specification of C language.

1. `comment` function is defined, which is called for block level comments
2. keywords such as `auto`, `break`, `for`, `int` etc all have associated tokens declared in `y.tab.h`, which is generated using yacc.
3. The keywords are listed below

   > `auto`, `break`, `case`, `char`, `continue`, `default`, `do`, `else`, `enum`, `extern`, `float`, `for`, `goto`, `if`, `int`, `long`, `register`, `return`, `short`, `signed`, `sizeof`, `static`, `struct`, `switch`, `typedef`, `union`, `unsigned`, `void`, `volatile`, `while`

4. There is a `check_type` function, albeit only the pseudocode is given. It checks a string if it is some identifier(variables, functions etc) after it has already checked that it is not a reserved keyword.
5. There is a `count` function, which is used to determine the length of a line.
6. Tab width is 8 in the count function
7. An identifier must begin with a letter or underscore as it is `{L}({L}|{D})*`, where L is `[a-zA-Z_]`
8. Integers, or floats both are given the same token.
9. There are few different representations of both int and floats
10. int : it can be represented in hexadecimal form, it can be represented in 0 followed by 1 or more digits followed by one or more character from u, U, l, L. It can be represented without 0 in beginning in previous format.
11. float : it can be represented in a few forms as well, one or more digits followed by E or e, followed by optional + or - and then one or more digits. It can also be represented as zero or more digits followed by decimal, followed by one or more digits, followed by optional E or e, followed by optional f, F, l, L.
12. There are token for arithmetic operators, logical operators, ellipsis etc.
13. It ignores bad characters, bad characters are those which do not match any c lex specification.

10 features that you notice in the Yacc specification of C language.

1. There are around around 59 different tokens in yacc specification, some of which are `INT`, `EXTERN`, `REGISTER` etc.

2. After these tokens, we have a `translation_unit`

3. The expressions are divided into

   > primary_expression, postfix_expression, argument_expression_list, unary_expression, cast_expression, multiplicative_expression, additive_expression, shift_expression, relational_expression, equality_expression, and_expression, exclusive_or_expression, logical_or_expression, conditional_expression, assignment_expression, assignment_expression, constant_expression.

4. We have 6 unary operators, namely

> &, *, +, -, ~, !

5. multiplicative_epression has *, %, /, while additive_expression has +,-. We also have shift_expressions, relational_expressions, equality_expressions, logical expressions which are all arithmetic expressions.

6. We have assignment_expression for assignment, the operators include

   > =, `MUL_ASSIGN`, `MUL_ASSIGN`, `DIV_ASSIGN`, `MOD_ASSIGN`, `ADD_ASSIGN`, `SUB_ASSIGN`, `LEFT_ASSIGN`, `RIGHT_ASSIGN`, `AND_ASSIGN`, `XOR_ASSIGN`, `OR_ASSIGN`.

7. We have a selection statement which defines `if-else if` block, `switch-case` block

8. We have initializer, to initialize identifiers.

9. We have iteration statements which deine loops like `for`, `while`, `do while`

10. We have a jump statement, which has the following tokens

    > GOTO, CONTINUE, BREAK, RETURN etc

    These are to exit and enter a loop

11. We have a yyerror function, which will give the error if there is a mistake in the program