Benjamin Bancala

CS214

PA1

Readme

My program takes in a string from the command line and stores it into a char\* that has been malloc’ed to handle up to 500 characters. TKCreate is called using the char\* as its only parameter. TKCreate creates a TokenizerT token structure which has properties curr (the char\*), word (the tokenized word), and state (the current state used for determining what type the tokenized word is). TKGetNextToken is then called which begins to access the original char\*. Depending on the first character, the state is then guessed and the appropriate method is called to figure out what type the word. This is continued with every character until a delimiter is hit. In my program, a delimiter is any type of whitespace or any character that would call a different method. When a delimiter is hit, the program returns the word, stores it, and prints it. If the delimiter happened to be whitespace, it is deleted, and the program continues. If the delimiter was another character signaling a different type the program calls TKGetNextToken on the new set of characters.

The program can identify alpha words, alphanumeric words, digits, decimals, hexadecimal, octal, floats, C operators, and unknowns:

Alpha word tokens are just letters.

Alphanumeric tokens are numbers without a decimal that are not classified as decimal, hex, octal, or float.

Decimal tokens have a decimal somewhere in the number.

Hexadecimal tokens starts 0x then the remaining characters are between 0-F.

Octal tokens are sets of digits between 0-7.

Floats tokens have an e followed by a number, a + ,or a – and then some numbers.

C operator tokens are the different operations that C can perform.

Unknown tokens are certain cases such as 5. (a decimal without numbers after).

When the program finishes, TKDestroy is called freeing the allocated memory.

Example of the program:

./tokenizer “hello m4 4.5 0555 0x32F3 444 [ ] 4.34e-23”

Output:

Word = hello

Alphanumeric = m4

Decimal = 4.5

Octal = 0555

Hex = 0x32F3

Digit = 444

C operator = [

C operator = ]

Float = 4.34e-23