

Arturo Corro
Stanley Chu

System programming -Assignment 1- Tokenizer

Tokenizer is a program that runs by taking an input string from the User and processes the input into different categories of "Tokens". Tokens can be denoted by Decimal integer, Float, Word, Hexadecimal, Octal, C operators, and bad tokens. The Tokenizer program is smart enough to only take in 1 String per run time command. If no string is inputted, Tokenizer will detect that nothing was inputted. Spaces, tabs, new line, return in between the string itself is also a determining factor of different tokens. Bad tokens are denoted by any symbol that is not covered by any of the above following definitions in tokenizer.

The program utilizes many sub functions such as gethex, getFloat, getOctal, getWord, getOp as easy helper methods that make TKGetnexttoken easier to manipulate instead of a long list of if conditionals. Each function is situation specific in accordance with its own definition. This program utilizes a first come first serve basis in that, given a word "hello0.123" it will detect the word "hello0" and realize "." as a structured member in the C operators library, followed by the decimal "123". One of its key features as a program is specificity, and being able to detect and break down token by token.

**** Note****

When user inputs the string, it is highly encouraged to test it with '<user string input> '. The ' ' is important because bash by default uses ! and \$ for its original commands.

Example: ./tokenizer 'hello world'