

# Systems Programming Asst0

## NAME

**tokenizer**

breaks up a single string into types of inputs

## LIBRARY USED

```
stdio, stdlib, string, ctype
```

## USAGE

```
./tokenizer <string to analyze>
```

## DESCRIPTION

This program takes a single string of an arbitrary length and breaks them up into types of tokens, including

```
word, decimal integer, octal integer, hexadecimal integer, floating point, C operator
```

Additionally, the program can recognize the reserved keywords in C as distinct tokens, skip `//` and

```
/* */
```

 comments, and recognize strings in quotes as single tokens.

The program reads the input string char by char from the start to the end, and prints out every single token in the form: `<token type>: "<token name>"`.

## FEATURES

This program runs fast in  $O(n)$  time and read each char only once in most cases. It's implemented with the idea of state machine using switch cases, hence the decision process of the type of a token is straight forward and won't run into irrelevant types.