

NAME

tokenizer – utility for determining and printing token types

SYNOPSIS

tokenizer [ARGUMENT] ...

DESCRIPTION

Print out the tokens of ARGUMENT. Tokens are determined by the type they correspond to and are separated by newlines. Tokens include common data types along with operators and keywords that are a part of the C programming language. Whitespace is ignored.

TOKEN TYPES

word, decimal integer, octal integer, hexadecimal integer, floating point, operator, keyword.

EXAMPLES

```
./tokenizer "123.5E-10abc.12 0xf"
```

floating point: "123.5E-10"

word: "abc"

structure member: "."

decimal integer: "12"

hexadecimal integer: "0xf"

```
./tokenizer "I love examples"
```

word: "I"

word: "love"

word: "examples"

DESIGN

The argument inputted is inspected one character at a time. The initial type of the token is assumed based on the first few characters. After the type is assumed, characters are read until a character is encountered that is not allowed in the assumed type. In certain cases, a type may be ‘promoted’ into another if the situation arises. Information about each token is collected and subsequently printed.

BEHAVIOR

Tokens are matched greedily, ex: “1230xF” results in
decimal integer: "1230"

word: "xF"

instead of

decimal integer: “123”

hexadecimal integer: “0xF”

ERROR HANDLING

In the case that malloc() fails:

“Memory allocation error” is printed

exit() is called with parameter EXIT_FAILURE