

## COMPILER DESIGN LAB 4

```
import re

def is_id(tok):
    return re.match(r'[_a-zA-Z][_a-zA-Z0-9]*', tok) != None

def is_num(tok):
    return re.match(r'\b\d+\b', tok) != None

def tokenizer(s):
    kw = {'int', 'float', 'char', 'if', 'else', 'while', 'for'}
    op = {'+', '-', '*', '/', '=', '==', '++', '--', '%', '&&', '||', '!', '>', '<', '>=',
    '<='}
    dl = {';', ',', '(', ')', '{', '}' }
    id_limit = 31

    s = re.sub(r'//.*', '', s)
    s = re.sub(r'/\*.*?\*/', '', s, flags=re.DOTALL)
    s = re.sub(r'\s+', ' ', s).strip()

    tokens = re.findall(r'==|!=|<=|>=|\+|\+|--|[\+\/\*\/\%=&!<>]=?|[\;\(\)\{\}]\|\b\d+\b|[_a-zA-Z][_a-zA-Z0-9]*', s)

    for tok in tokens:
        if tok in kw:
            print("Keyword:", tok)
        elif is_id(tok):
            if len(tok) > id_limit:
                print("Warning!!!!")
                print("Identifier:", tok[:id_limit], "(truncated)")
            else:
                print("Identifier:", tok)
        elif is_num(tok):
            print("Number:", tok)
        elif tok in op:
            print("Operator:", tok)
        elif tok in dl:
            print("Delimiter:", tok)
        else:
            print("Unrecognized token:", tok)

# Example usage
s=input("Enter string : ")
tokenizer(s)
```

```
PS C:\Users\Samar Mittal\Desktop\Compiler
Enter string : float y=x+5;
Keyword: float
Identifier: y
Identifier: x
Operator: +
Number: 5
Delimiter: ;
```

TEST CASE 1

```
PS C:\Users\Samar Mittal\Desktop\Compi
Enter string : int x=42;
Keyword: int
Identifier: x
Operator: =
Number: 42
Delimiter: ;
```

TEST CASE 2

```
PS C:\Users\Samar Mittal\Desktop\Com
Enter string : /* This is a test */i
Keyword: int
Identifier: a
Operator: =
Number: 5
Delimiter: ;
Keyword: if
Delimiter: (
Identifier: a
Unrecognized token: &
Identifier: gt
Delimiter: ;
Number: 3
Delimiter: )
Delimiter: {
```

TEST CASE 3

TEST CASE 4

```
PS C:\Users\Samar Mittal\Desktop\Compiler Lab\lab4> python -u "c:\Users\Samar Mittal
Enter string : int this_is_a_very_long_identifier_name_that_needs_truncation = 10;
Keyword: int
Warning!!!!
Identifier: this_is_a_very_long_identifier_ (truncated)
Operator: =
Number: 10
Delimiter: ;
PS C:\Users\Samar Mittal\Desktop\Compiler Lab\lab4> █
```