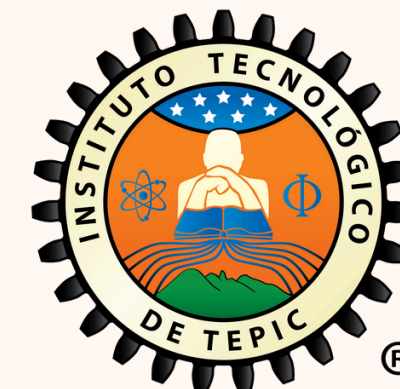




TECNOLÓGICO
NACIONAL DE MÉXICO®



Haz tu Propio Compilador con Python (y Sobrevive en el Intento)

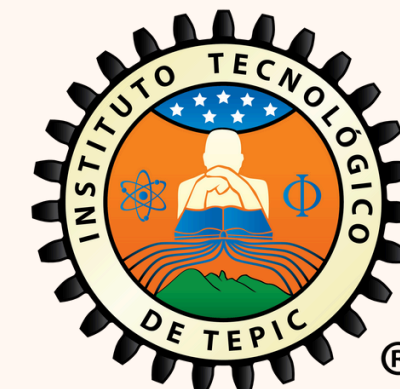


Análisis Léxico

By Jorge Ismael Betancourt Espericueta



TECNOLÓGICO
NACIONAL DE MÉXICO®



¿Qué es una **expresión regular**?

Una **expresión regular** (regex, por sus siglas en inglés: regular expression) es una secuencia de caracteres que define un patrón de búsqueda. Se utiliza principalmente para encontrar, validar, reemplazar o manipular texto en diversas tareas de programación.



/regex?/



5variable

12a

4AsiNo



$r' \backslash d + [a-zA-Z\tilde{n}\tilde{N}][a-zA-Z0-9\tilde{n}\tilde{N}]^*$



Representa uno
o más dígitos.
Cualquier
dígito"
(equivalente a
[0-9])



Esta es una clase de
caracteres que incluye
letras mayúsculas y
minúsculas del alfabeto
inglés, además de las
letras ñ y Ñ



Similar a la clase de
caracteres anterior, pero
esta también incluye
dígitos (0-9)

Pregunta de comprensión

¿Cuál de las siguientes cadenas es válida según la expresión regular? $r' \backslash d+[a-zA-ZñÑ][a-zA-Z0-9ñÑ]^*$

A) abc123

B) 123a

C) 12ñ34

D) 456!



Pregunta de comprensión

¿Cuál de las siguientes cadenas es válida según la expresión regular? $r'\backslash d+[a-zA-ZñÑ][a-zA-Z0-9ñÑ]^*$

A) 7890XYZ

B) 12-abc

C) 1234.567

D) 99ñ



r' \ # . * ? \ #'

Busca un #

**El punto
significa
"cualquier
carácter"**

**El asterisco
significa "cero o
más"
repeticiones**

**Este signo
transforma la
búsqueda en no
codiciosa**

#Esto es una cadena#

Pregunta de comprensión

¿Cuál de las siguientes cadenas es válida según la expresión regular? $r'\backslash\#.*?\backslash\#'$

A) #Hola Mundo#

B) #1234

C) Hola#Mundo#

D) # #



Pregunta de comprensión

¿Cuál de las siguientes cadenas es válida según la expresión regular? $r'\backslash\# \cdot *? \backslash\#'$

A) ##

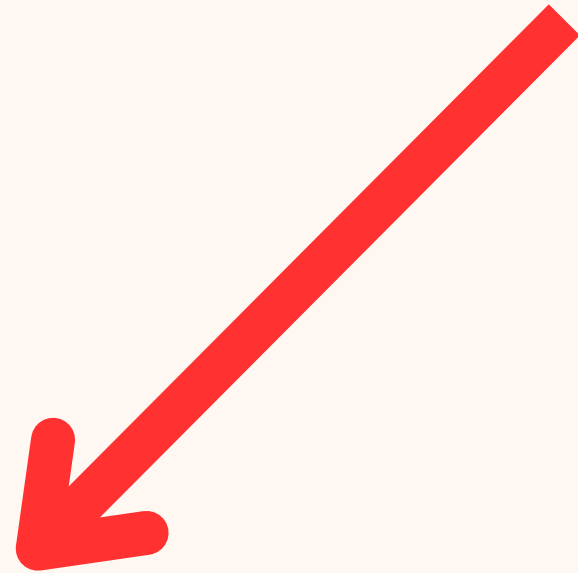
B) #Esta no es#

C) #Mundo

D) #1 2 3#abc



r' \ / / (. * ?) \ / / '



**Busca una
barra
diagonal
literal.**



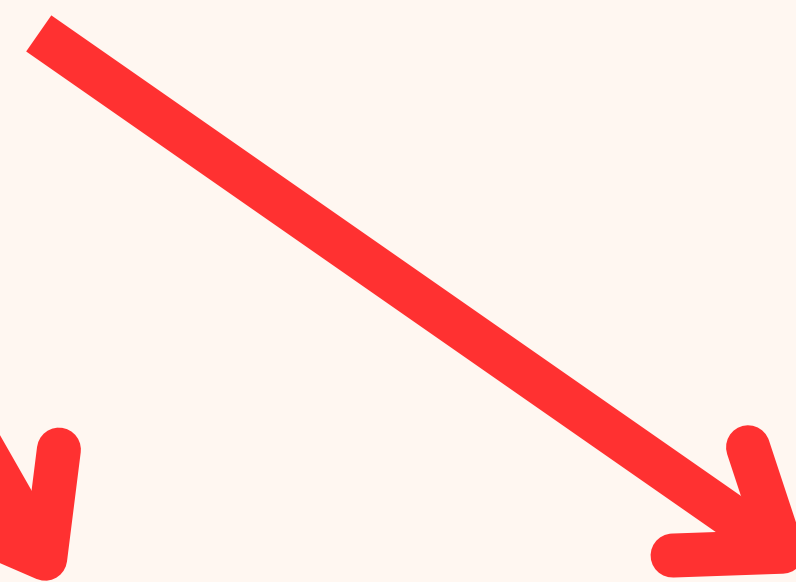
**Los paréntesis
indican una
subexpresión**



**El punto
significa
"cualquier
carácter"**



**El asterisco
significa "cero o
más"
repeticiones**



**Este signo
transforma la
búsqueda en no
codiciosa**

//Esto es un comentario//

Pregunta de comprensión

¿Cuál de las siguientes cadenas es válida según la expresión regular? $r'\backslash//(. * ?)\backslash//'$

A) **// Esto es un comentario //**

B) **//Comentarios sin cierre**

C) **// 12345 //**

D) **Esto no es un comentario //**



Pregunta de comprensión

¿Cuál de las siguientes cadenas es válida según la expresión regular? $r'\backslash//(. * ?)\backslash//'$

A) $// \text{ contenido } //$

B) $////$

C) No es un comentario $//$

D) $//\text{abc} //$



R

34.99

0.7

9.5

Es un or

Números Reales

$r'(\backslash d+\backslash.\backslash d+|\backslash.\backslash d+)'$

Representa uno o
más dígitos.
Cualquier dígito"
(equivalente a [0-9])

Un punto literal

.27

.4

Pregunta de comprensión

¿Cuál de las siguientes cadenas es válida según la expresión regular? $r'(\backslash d+\backslash.\backslash d+ | \backslash.\backslash d+)$

A) 3.14

B) .45

C) 5.

D) abc.5



Pregunta de comprensión

¿Cuál de las siguientes cadenas es válida según la expresión regular? $r' \backslash // (\cdot * ?) \backslash // '$

A) 10.0

B) 0.123

C) 5.25

D) 12.34.56



```
def analisis(cadena):  
    lexer.input(cadena)  
    tokens = []  
    # Inicia el número de línea en 1  
    lexer.lineno = 1  
    for tok in lexer:  
        column = tok.lexpos - cadena.rfind('\n', 0, tok.lexpos)  
        tokens.append((tok.value, tok.type, tok.lineno, column))  
    return tokens
```

