

Referencia Rápida de Expresiones Regulares

Patrones, cuantificadores, grupos, lookaheads, banderas

Patrones Básicos

Metacaracteres

- Cualquier carácter (excepto nueva línea)
- ^ Inicio de cadena / línea
- \$ Fin de cadena / línea
- * 0 o más del anterior
- + 1 o más del anterior
- ? 0 o 1 del anterior (opcional)
- \ Escapar metacarácter

Coincidencia Literal

```
hello # matches "hello" exactly
a.c   # matches "abc", "alc", "a-c", etc.
.txt  # matches literal ".txt"
```

Clases de Caracteres

Expresiones de Corchetes

```
[abc]   Coincidir a, b, o c
[^abc]  Coincidir cualquier cosa excepto a, b, c
[a-z]   Letra minúscula
[A-Z]   Letra mayúscula
[0-9]   Dígito
[a-zA-Z0-9] Alfanumérico
```

Clases Abreviadas

```
\d Dígito [0-9]
\D No dígito [^0-9]
\w Carácter de palabra [a-zA-Z0-9_]
\W Carácter que no es de palabra
\s Espacio en blanco [\t\n\r\f]
\S Sin espacio en blanco
```

Cuantificadores

Cuantificadores Codiciosos

```
* 0 o más (codicioso)
+ 1 o más (codicioso)
? 0 o 1 (codicioso)
{n} Exactamente n veces
{n,} n o más veces
{n,m} Entre n y m veces
```

Cuantificadores Perezosos

```
*? 0 o más (perezoso / no codicioso)
+? 1 o más (perezoso)
?? 0 o 1 (perezoso)
{n,m}? Entre n y m (perezoso)
```

Los cuantificadores perezosos coinciden con el mínimo de caracteres posible

Codicioso vs Perezoso

```
<.+> # greedy: "<b>bold</b>"
<.+?> # lazy: "<b>"
```

Anclas

```
^ Inicio de cadena (o línea con bandera m)
$ Fin de cadena (o línea con bandera m)
\b Límite de palabra
\B No límite de palabra
\A Inicio de cadena (no afectado por m)
\Z Fin de cadena (no afectado por m)
```

Ejemplos de Anclas

```
^Hello # starts with "Hello"
world$ # ends with "world"
\bword\b # "word" as whole word
\Bword\B # "word" inside another word
```

Grupos y Alternación

Grupos de Captura

```
(abc) # capture group: match "abc"
(a|b|c) # alternation: a or b or c
(cat|dog) # match "cat" or "dog"
(\d{3})-(\d{4}) # groups: "123-4567"
```

Tipos de Grupos

```
(pattern) Grupo de captura
(?:pattern) Grupo sin captura
(?P<name>pat) Grupo con nombre (Python)
(?<name>pat) Grupo con nombre (JS, .NET)
\1 \2 Referencia inversa al grupo 1, 2
a|b Alternación: a o b
```

Lookahead y Lookbehind

```
(?=pattern) Lookahead positivo
(?!pattern) Lookahead negativo
(?<=pattern) Lookbehind positivo
(?<!pattern) Lookbehind negativo
```

Ejemplos de Lookaround

```
\d+(?= USD) # digits followed by " USD"
\d+(?! USD) # digits NOT followed by " USD"
(?<=\$)\d+ # digits preceded by "$"
(?!\$)\d+ # digits NOT preceded by "$"
```

Los lookarounds coinciden con una posición sin consumir caracteres

Patrones Comunes

```
\d{1,3}(\.\d{1,3}){3} Dirección IPv4 (básica)
[\\w.+-]+@[\\w-]+\\.([\\w.]+) Email (básico)
https?://[\\w./-?&#]=]+ URL (básica)
\\(\\d{3}\\)\\?[-.\\s]?\\d{3}[-.\\s]?\\d{4} Número de teléfono de EE.UU.
\\d{4}-\\d{2}-\\d{2} Fecha (AAAA-MM-DD)
#[0-9a-fA-F]{6} Código de color hexadecimal
```

Estos son patrones simplificados; el uso en producción puede requerir validación más estricta

Banderas

```
g Global: encontrar todas las coincidencias, no solo la primera
i Sin distinción de mayúsculas
m Multilínea: ^ / $ coinciden con límites de línea
s Dotall: . también coincide con nueva línea
x Detallado: ignorar espacios en blanco, permitir comentarios
u Unicode: soporte completo de Unicode
```

Uso de Banderas por Lenguaje

```
/pattern/gi # JavaScript
re.compile(r"pat", re.I | re.M) # Python
grep -iE "pattern" # grep (extended)
```