

Expresiones Regulares

Definición: Una expresión regular es una forma de describir cadenas de caracteres en términos generales, es decir, definir patrones de texto que se puedan buscar, comparar, seleccionar, editar o verificar en otra cadena.

El patrón se aplica a una cadena de izquierda a derecha y una vez un caracter de la cadena se ha usado, no puede reutilizarse, ejemplo el patrón **“oso”**, se presenta en la cadena **“ososososo”**, sólo dos veces.

Los constructores de una expresión regular incluyen caracteres, clases de caracteres, operadores, cuantificadores, metacaracteres y delimitadores de contorno.

Caracteres: en una expresión regular se usan caracteres alfanuméricos normales y caracteres especiales. Los caracteres especiales se escriben como secuencias de escape.

Caracteres

x	el caracter x
<i>alfanumericos</i>	Todos los Caracteres alfanuméricos coinciden con sí mismos. Por ejemplo, la expresión regular “6 meses de cuerentena” debe coincidir con la cadena “6 meses de cuerentena”
\\	El caracter de barra invertida
\t	Coincide con un caracter de tabulación.
\n	El caracter de nueva línea (avance de línea) ('\ u000A') Coincide con un carácter de nueva línea.
\r	El caracter de retorno de carro ('\ u000D')
\v	Coincide con un carácter de tabulación vertical.
\f	El caracter de avance de página ('\ u000C')
\a	El caracter de alerta (campana) ('\ u0007')
\e	El caracter de escape ('\ u001B')
\cx	Coincide con el carácter de control indicado por x. El valor de x debe ser un valor que se encuentre en el rango de A-Z o a-z. \cM coincide con Control-M o con un carácter de retorno de carro .

El \ Marca el caracter siguiente como un caracter especial, un literal, una referencia inversa o un escape octal.

La secuencia '\\ ' coincide con "\"

"\" coincide con "\".

n** coincide con el carácter **n

***\n** coincide con un carácter de nueva línea*

Clase de caracteres

La función de los corchetes en el lenguaje de las expresiones regulares es representar "clases de caracteres", o sea, agrupar caracteres en grupos o clases.

Clases de caracteres	
[abc]	a, b, o c (clase simple) man[ie]ja coincide con manija o con maneja [abc] coincide con la 'a' de "plato".
[^abc]	Cualquier caracter excepto a, b o c (negación) 1[^02] coincide con el número 13 u 11 pero no con el 10 o el 12 [^abc]' coincide con la 'p' de " p lato".
[a-zA-Z]	Desde a a la z o desde A hasta Z, incluidos [1-9] coincide con cualquier dígito de una sola cifra EXCEPTO 0 ' [a-z] ' coincide con cualquier caracter alfabético en minúscula en el rango de la 'a' a la 'z'
[a-d[m-p]]	Desde a hasta d, o desde m a la p: [a-dm-p] (unión)
[a-z&&[def]]	La d, la e, o la f (intersección)
[a-z&&[^bc]]	Desde la a hasta la z, excepto la b y la c: [ad-z] (resta)
[a-z&&[^m-p]]	Desde la a hasta la z, excepto desde la m hasta la p: [a-lq-z](resta)

Metacaracteres	
.	Cualquier caracter (puede o no coincidir con terminadores de línea) bar. coincide con barT y con bar2 pero no con barcito Nota Para establecer una coincidencia con cualquier caracter, incluido '\n', utilice un patrón como '[\s\S]'.
\d	Un dígito: [0-9]
\D	Un no dígito: [^ 0-9]
\s	Un caracter de espacio en blanco: [\t\n\r\f] Puede combinarse de la misma manera como [\d\s]; de este modo, coincide con un caracter que sea un dígito o espacio en blanco
\S	Un caracter que no sea un espacio en blanco: [^ \s] Coincide con cualquier carácter no correspondiente a un espacio en blanco. Equivale a [^ \f\n\r\t\v]
\w	Cualquier caracter alfanumérico: [a-zA-Z_0-9] Coincide con cualquier carácter de palabra, incluido el guión bajo. Equivale a '[A-Za-z0-9_]'
\W	Cualquier caracter no alfanumérico: [^ \w] Coincide con cualquier caracter no correspondiente a una palabra. Equivale a '[^A-Za-z0-9_]'
\h	Un caracter de espacio en blanco horizontal: [\t \xA0 \u1680 \u180e \u2000- \u200a \u202f \u205f \u3000]
\H	Un caracter de espacio en blanco no horizontal: [^ \h]
\v	Un caracter de espacio en blanco vertical: [\n \x0B \f \r \x85 \u2028 \u2029]
\V	Un caracter de espacio en blanco no vertical: [^ \v]

Operadores	
XY	X seguido de Y
X Y	X o Y (primero 1ero 1) de julio coincidirá con 1ero de julio pero no con 2 de julio 'z comida' coincide con "z" o con "comida". '(z f)omida' coincide con "zomida" o con "comida".
(X)	X, como <u>grupo de captura</u> . Los paréntesis crean una subcadena o elemento a los que se les pueden aplicar metacaracteres. a(beja)?t coincide con a o con abeja pero no con abej

Cuantificadores

Cuantificadores	
X?	<p>X cero o una vez abc? coincide con ab o con abc</p> <p>colo?r coincide con color pero no con colorrr</p> <p>escribir(les)? coincide con el escribir de escribir o de escribirles.</p>
X*	<p>X cero o más veces</p> <p>be* coincide con b o beo beeeeeeeeeee</p> <p>zo* coincide con z y con zoo.</p>
X+	<p>X una o más veces</p> <p>be+ coincide con be o con bee pero no con b</p>
X{n}	<p>X exactamente n veces</p> <p>[0-9]{3} coincide con cualquier número de tres dígitos</p> <p>r{2} no coincide con la r de pera, pero coincide con las dos r de perro.</p> <p>b{4} coincide con bbbb</p>
X{n, }	<p>X al menos n veces</p> <p>[0-9]{3,} coincide con cualquier número de tres o más dígitos</p> <p>r{2,} no coincide con la "r" de "pera", pero coincide con todas las r de "perrrrro". 'o{1,}' equivale a 'o+'. 'o{0,}' equivale a 'o*'.</p>
X{n,m }	<p>X mínimo n veces pero no más de m veces</p> <p>[0-9]{3,5} coincide con cualquier número de tres, cuatro o cinco dígitos</p> <p>"r{1,3}" coincide con las tres primeras r de "perrrrro". 'o{0,1}' equivale a 'o?'.</p> <p>c{2, 4} coincide con cc, ccc, cccc</p>

Delimitadores		Ejemplos
^	La coincidencia debe empezar al principio de una cadena o línea. Ej: ^http: coincide con "http": solo cuando se encuentra al principio de una línea.	“^hola” coincide con “hola SAra” pero no con “dijo hola”
		^gato coincide con las cadenas que comienzan con gato

\$	La coincidencia se debe producir al final de la cadena o antes de \n al final de la cadena.	"ola\$" coincide con "hola" pero no a "olas" gato\$ coincide con cualquier cadena que finalice con gato
\b	La coincidencia se debe producir en un límite de palabras Coincide con un límite de palabra: la posición entre una palabra y un espacio	Ola\b coincide sólo con la primera "ola" en "hola le dijo a las olas" 'er\b' coincide con el 'er' de "beber" pero no con el 'er' de "servicio".
\B	La coincidencia no se debe producir en un límite de palabras	\Bola coincide sólo con la primera "ola" en "hola le dijo a las olas" 'er\B' coincide con el 'er' de "servicio" pero no con el 'er' de "beber".
\A	La coincidencia se debe producir solo al principio de la cadena (no se admiten varias líneas)	
\G	La coincidencia se debe iniciar en la posición en que finalizó la coincidencia anterior.	\Gola coincide dos veces en "olaolas"
\Z	Coincide con el final de la cadena a la que se aplica la expresión regular.	.\Z coincide con la k de jol\hok
\z	La coincidencia se debe producir solo al final de la cadena.	

Clases de caracteres POSIX (solo US-ASCII)

\p{Lower}	Un carácter alfabético en minúscula: [a-z]
\p{Upper}	Un carácter alfabético en mayúsculas: [A-Z]
\p{ASCII}	Todo ASCII: [\ x00- \ x7F]
\p{Alpha}	Un carácter alfabético: [\ p {Lower} \ p {Upper}]
\p{Digit}	Un dígito decimal: [0-9]
\p{Alnum}	Un carácter alfanumérico: [\ p {Alpha} \ p {Digit}]
\p{Punct}	Puntuación: uno de ! " # \$ % & ' () * + , - . / : ; < = > ? @ [\] ^ _ ` { } ~
\p{Graph}	Un carácter visible: [\ p {Alnum} \ p {Punct}] Coincide únicamente con caracteres visibles, es decir, con cualquier carácter excepto con espacios, caracteres de control, etc.
\p{Print}	Un carácter imprimible: [\ p {Graph} \ x20]
\p{Blank}	Un espacio o un tabulador: [\ t]
\p{Cntrl}	Un carácter de control: [\ x00- \ x1F \ x7F]
\p{XDigit}	Un dígito hexadecimal: [0-9a-fA-F]
\p{Space}	Un carácter de espacio en blanco: [\ t \ n \ x0B \ f \ r]