

Expresiones regulares

4. Expresiones de repetición



Expresiones de repetición

- Metacaracteres de repetición
- Cuantificadores de repetición
- Expresiones “codiciosas” (Greedy)
- Expresiones “flojas” (Lazy)
- Utilizar las repeticiones en forma eficiente

Expresiones regulares

Metacaracteres de repetición



Metacaracteres de repetición

* cero o más veces del elemento precedente

+ una o más veces del elemento precedente

? cero o una vez (opcional) del elemento precedente

Metacaracteres de repetición

/aguas*/ empata aguasssss, aguas, agua

/aguas+/ aguas, aguassssssss, no agua

/aguas?/ aguas, agua, no aguassssss

Metacaracteres de repetición

`/\d\d\d\d*/` tres dígitos o más

`/\d\d\d+/` tres dígitos o más

`/Jes+ica/` Jesica, Jessica

`/\w+s/` palabras que terminen en ese

Metacaracteres de repetición

* soportado por todos los motores regex

+, ? No soportados por BRE (Unix viejo)

`[a-z]+\d[a-z]*`

Expresiones regulares

Cuantificadores de repetición



Cuantificadores de repetición

- {} lo que está antes de las llaves debe aparecer exactamente esa cantidad de veces
- {min,max} --> lo que está antes tiene que aparecer entre esas cantidades (ambas inclusive)
- {min,} definimos un mínimo, no un máximo, tiene que aparecer desde esa cantidad de veces.

Cuantificadores de repetición

- min y max deben ser positivos
- min siempre es incluido y puede ser cero
- max es opcional

$\backslash d\{4,8\}$

$\backslash d\{4\}$

$\backslash d\{4,\}$

$\backslash d\{3\}-\backslash d[3]-\backslash d\{2\}$

Cuantificadores de repetición

`\d{0,}` es igual a `\d*`

`\d{1,}` es igual a `\d+`

`A{1,3}` empata con A, AA, AAA pero no AAAA

Expresiones regulares

Expresiones “codiciosas”



Expresiones “codiciosas”

Las expresiones regulares son “codiciosas” (greedy) porque siempre van a seleccionar la cadena más grande.

`<p>(.*)</p>`

`<p>en el agua clara, que brota en la fuente</p> <p> un lindo pescado, sale de repente</p>`

Expresiones “codiciosas”

Codicioso

Perezoso

(.*)

(.*?)

(.+)

(.+?)

(?)

(??)

Expresiones regulares

Expresiones perezosas



Expresiones perezosas

Para hacer a las expresiones regulares ya no sean “codiciosas”, las podemos hacer perezosas:

Codicioso	Perezoso
(.*)	(.*?)
(.+)	(.+?)
(?)	(??)
{min.max}	{min,max}?

Expresiones perezosas

Las expresiones regulares “perezosas” no son necesariamente más rápidas o más lentas.

Las sentencias “perezosas” no son soportadas por todas las herramientas de UNIX.

Expresiones regulares

Utilizar las repeticiones en forma eficiente



Utilizar las repeticiones en forma eficiente

`/\w+s/`

El metacaracter “+” es más eficiente que “*”

Las llaves {} son más eficientes que “+” y “*”

Delimitar las búsquedas con delimitadores (^\$) es más eficiente