

# Expresiones regulares

## 3. Conjunto de caracteres



# Conjunto de caracteres

---

- Definir un conjunto de caracteres
- Rango de caracteres
- Conjunto negativo de caracteres
- Metacaracteres dentro de los conjuntos de caracteres
- Abreviaciones conjunto de caracteres
- Expresiones de corchetes POSIX

# Expresiones regulares

Definir un conjunto de caracteres



# Definir un conjunto de caracteres

---

Podemos definir un conjunto de caracteres por medio de los “corchetes” (Square Brackets) [ ]

/[aeiou]/

/[áéíóú]/

# Definir un conjunto de caracteres

---

Sólo evalúa caracteres, no palabras.

El orden de los caracteres no importa.

/m[aeiu]sa/ empata con “masa”, “mesa”, “misa” y “musa” no con “mosa”

# Expresiones regulares

Rango de caracteres



# Rango de caracteres

---

Por medio del renglón medio” (-) podemos abreviar los rangos de caracteres, por ejemplo:

[A-Z]

Representan todos los caracteres entre ambos.

Debes escribirse en orden. No es válido

[Z-A]

# Rango de caracteres

---

Sólo utilizamos caracteres, no conjunto de caracteres:

[40-45]

No representa de 40 a 45, sino 4, de 0 a 4 y 5.

Podemos escribir diferentes grupos de caracteres:

[A-Za-z0-9]

No incluye vocales acentuadas, eñes, etc., sólo las letras del abecedario inglés.



# Rango de caracteres

---

Prueba [á-ü]

[50-60]mal

[A-Z]ola

[0-9][0-9]-[0-9][0-9]-[0-9][0-9]-[0-9][0-9]-[0-9][0-9]

[0-9][0-9][0-9][0-9][0-9]

# Expresiones regulares

Conjunto negativo de caracteres



# Conjunto negativo de caracteres

---

El caracter ^ (circunflejo) dentro de los corchetes es una negación de uno de los conjuntos de caracteres.

“Ninguno de los caracteres marcados”

Se escribe al inicio, dentro del conjunto de caracteres:

`[^aeiou]`

`[^"#$%&/()=*]`

# Expresiones regulares

Metacaracteres dentro de los  
conjuntos de caracteres



# Metacaracteres dentro de los conjuntos de caracteres

---

Los metacaracteres dentro de los conjuntos de caracteres ya están “escapadas”.

No hay necesidad de “escaparlos” nuevamente.

`m[ao.xcv]`lo malo, m.lo, pero no mulo

Los siguientes metacaracteres si hay que “escaparlos”

`] - ^ \`

`var[([0-9][\]])` por ejemplo `var(9)` `var[9]`

# Metacaracteres dentro de los conjuntos de caracteres

---

Foto[0\_-/]nueva.jpg

foto-nueva.jpg foto\_nueva.jpg foto/nueva.jpg foto1nueva.jpg

Foto[0\_\\-/]nueva.jpg

# Expresiones regulares

Abreviaciones conjunto de caracteres



# Abreviaciones conjunto de caracteres

---

<code>\d</code>	Dígitos	<code>[0-9]</code>
<code>\w</code>	Caracteres	<code>[a-zA-Z0-9_]</code> (no incluye el punto)
<code>\s</code>	Espacios en blanco	<code>[ \t\r\n]</code>
<code>\D</code>	No dígito	<code>[^0-9]</code>
<code>\W</code>	No caracteres	<code>[^a-zA-Z0-9_]</code>
<code>\S</code>	No espacios en blanco	<code>[^ \t\r\n]</code>



# Abreviaciones conjunto de caracteres

---

`/\d\d\d\d\d\d/`

`/\w\w\w/`

`/\w\s\w\w/`

`/[\w\-/`

# Abreviaciones conjunto de caracteres

---

`[\d\s]`

`[^\d\s]` Diferente a...

`[\D\S]` Diferente a...

Soportado por PERL

No soportado por Unix

# Expresiones regulares

Expresiones de corchetes POSIX



# Expresiones de corchetes POSIX

---

`[alpha:]`

`[dogit:]`

`[alnum:]`

`[lower:]`

`[upper:]`

# Expresiones de corchetes POSIX

---

[ :punct: ]

[ :space: ]

[ :blank: ]

[ :print: ]

[ :graph: ]

[ :cntrl: ]

# Expresiones de corchetes POSIX

---

`[:cntrl:]`

`[:xdigit:]` (hexadecimal)

Se utilizan dentro de corchetes

`[[:alpha:]]`

(Perl, PHP, Ruby y UNIX)

No soportados en Java, JavaScript, .NET, Python y HTML5