



EXPRESIONES REGULARES

Las expresiones regulares son, como su nombre lo indica, expresiones que llegan en la forma de cadenas de texto. Estas expresiones son utilizadas dentro del mundo de la computación para determinar si un texto cumple o no con un patrón dado. Este patrón se lo escribe como una expresión regular.

En este curso las expresiones regulares nos ayudarán para hacer reglas dentro de Scrapy. En realidad, no necesitaremos tener un conocimiento profundo sobre las expresiones regulares. Ya que las que haremos van a ser simples. Lo importante es saber que estas expresiones me sirven para detectar si una cadena cumple o no con un patrón.

Las expresiones regulares toman en cuenta mayúsculas, minúsculas, dígitos, símbolos, puntuación y espacios.

Algunos ejemplos de los patrones más comunes escritos con expresiones regulares para el Web Scraping son los siguientes:

Expresión Regular (Patrón)	DESCRIPCIÓN	HACEN MATCH	NO HACEN MATCH
<code>r 'https '</code>	Hace match con cualquier cadena que tenga los caracteres "https"	<code>https</code> <code>https://www.google.com</code> <code>111https000</code>	<code>http</code> <code>http://www.google.com</code> <code>hola mundo</code> <code>htt1ps</code>
<code>r 'page= '</code>	Hace match con cualquier cadena que tenga los caracteres "page="	<code>page=</code> <code>google.com?page=8</code> <code>google.com?page=uno</code>	<code>Page=</code> <code>google.com?p=8</code> <code>page</code>
<code>r 'page=\d+ '</code>	Hace match con cualquier cadena que tenga los caracteres "page=" seguido de cualquier número.	<code>google.com?page=8</code> <code>/page=800&q=hola</code> <code>/page=0</code>	<code>page=uno</code> <code>page=</code> <code>?p=4</code> <code>page=\d+</code>



Puedes encontrar más información sobre expresiones regulares en los siguientes enlaces:

<https://www.petefreitag.com/cheatsheets/regex/>

<https://www.rexegg.com/regex-quickstart.html>

<https://regexr.com/>