Cheatography

Regular Expressions: Regexes in Python Cheat Sheet

by schmidtlady16 via cheatography.com/208683/cs/44907/

Módulo Re

`import re`	Modulo que contiene funciones y métodos Regex

Función re.search()

`re.sea rch (<r ege x>, <st rin g >)`

Escanea la cadena <string> en busca de la primera ubicación donde <regex> coincide

Si encuentra una coincidencia, devuelve un objeto de coincidencia; si no, devuelve None

Metacaracteres

.	Coincide con cualquier carácter excepto el salto de línea	***r"a.b" coincide con "-acb", "-a_b", etc.***
^	Ancla la coincidencia al inicio de una cadena; dentro de una clase de caracteres, complementa la clase	***r"^Hello" coincide con "-Hello World"***
$	Ancla la coincidencia al final de una cadena	***r"world$" coincide con "-hello world"***
*	Coincide con cero o más repeticiones	***r"a"* coincide con "-aaa" ***
+	Coincide con una o más repeticiones	***r"a+" coincide con "a", "aaa"***
?	Coincide con cero o una repetición; también crea una coincidencia de longitud mínima para *, + y ?	***r"colou?r" coincide con "color", "colour"***
{}	Coincide con un número específico de repeticiones indicado en {m,n}	***r"a{2,3}" coincide con "aa", "aaa"***

Metacaracteres (cont)

\	Escapa un metacaracter para usarlo como literal	***r"\$5" coincide con "-$5"***
[]	Especifica una clase de caracteres	***r"[aeiou]" coincide con vocales***
|	Designa una alternación	***gato|perro busca cualquiera de las dos palabras en el texto, devolviendo ambas coincidencias***
()	Crea un grupo de coincidencia	***r"(ab)c" coincide con "-abc"***

Anclas

^ o \A	Ancla al inicio de la cadena	***r"^The" coincide con "The cat" pero no con "A cat The"***
$ o \Z	Ancla al final de la cadena	***r"end$" coincide con "This is the end"***
\b	Ancla al inicio o final de una palabra	***r"\bcat\b" coincide con "cat", no con "scatter"***
\B	Ancla en una posición que no es límite de palabra	***r"\Bcat" coincide con "scatter" pero no con "cat"***

Clases de Caracteres

Clases de Caracteres (cont)

\S	Coincide con cualquier carácter que no sea espacio en blanco	***r"\S+" coincide con "HelloWorld"***

Cuantificadores

*	Cero o más veces del carácter o grupo anterior	***r"abc"* coincide con "ac", "abc", "abbc" ***
+	Una o más veces del carácter o grupo anterior	***r"ab+c" coincide con "abc", "abbc", pero no con "-ac"***
?	Cero o una vez del carácter o grupo anterio	***r"colou?r" coincide con "-color" y "colour"***
{m}	Exactamente m repeticiones	***r"a{3}" coincide con "aaa"***
{m,n}	Entre m y n repeticiones	***r"a{2,4}" coincide con "aa", "aaa", "aaaa"***

Modificadores de Expresión Regular (Flags)

`re.I` `re.IGN ORECASE`

Hace que la coincidencia no distinga entre mayúsculas y minúsculas

re.search(r"hello", "Hello World", re.I)

`re.M` `re.MUL TILINE`

Permite que ^ y $ coincidan al principio o fin de cada línea

re.search(r"^cat", "cat\ncat", re.M)

`re.S` `re.DOTALL`

Permite que . coincida con un salto de línea (\n)

re.search(r"hello.world", "hello\nworld", re.S)

`re.X` `re.VERBOSE`

Permite incluir espacios en blanco y comentarios en la expresión

re.search(r"(\d{3}) # area", "123-456", re.X)

`re.A` `re.ASCII`

Especifica codificación ASCII para la clasificación de caracteres

re.search(r'\w+', 'café_123', re.ASCII)

\d	Coincide con cualquier dígito decimal	*r"\d+" coincide con "123", "456-7"*
\D	Coincide con cualquier carácter que no sea dígito	*r"\D+" coincide con "abc", "-XYZ"*
\w	Coincide con cualquier carácter de palabra	*r"\w+" coincide con _ a-zA-Z0-9*
\W	Coincide con cualquier carácter que no sea de palabra	*r"\W+" coincide con "@ # $"*
\s	Coincide con cualquier espacio en blanco	*r"\s+" coincide con " ", "\t", "\n"*