

EXPRESIONES REGULARES

INTRODUCCIÓN

La búsqueda de caracteres o secuencias de caracteres en documentos forma parte de las tareas estándar y recurrentes de la tecnología de la información. Por lo general, el objetivo es modificar o sustituir fragmentos de texto o líneas de código, cuya complejidad aumenta en función de las veces que la secuencia de caracteres aparece en el documento. En la década de 1950 se encontró una solución basada en las lenguas formales de la informática teórica que sigue presente en el desarrollo actual de software y que permite simplificar estas tareas repetitivas mediante el uso de las denominadas **expresiones regulares**.

¿QUÉ ES UNA EXPRESIÓN REGULAR?

Las regex son las unidades de descripción de los lenguajes regulares, que se incluyen en los denominados lenguajes formales. Son un instrumento clave de la informática teórica, la cual, entre otras cosas, establece las bases para el desarrollo y la ejecución de programas informáticos, así como para la construcción del compilador necesario para ello. Para cada regex existe un denominado **autómata finito** que **acepta el lenguaje especificado por la expresión** y que, con ayuda de la construcción de Thompson, se desarrolla a partir de una expresión regular.

¿CÓMO FUNCIONA UNA EXPRESIÓN REGULAR?

Una expresión regular puede estar formada, o bien exclusivamente por caracteres normales [como abc], o bien por una combinación de caracteres normales y metacaracteres [como ab*c]. Los metacaracteres describen ciertas construcciones o disposiciones de caracteres: por ejemplo, si un carácter debe estar en el inicio de la línea o si un carácter solo debe o puede aparecer exactamente una vez, más veces o menos. Además, cada regex se puede vincular a una acción concreta, como la ya mencionada **“Reemplazar”**. Esta acción se ejecuta en todos los lugares en los que se detecta la expresión regular, es decir, **en todos los puntos en los que haya una coincidencia similar a la de los ejemplos**.

¿QUÉ RETOS PRESENTA EL USO DE EXPRESIONES REGULARES?

Para asegurar que se logrará el objetivo en todos los casos, pueden utilizarse instrucciones muy generales, pero, si se busca un resultado más exacto, entonces será inevitable formular un patrón regex específico. También vale la pena considerar la longitud: cuanto más compacta sea la expresión regular, menor será su tiempo de procesamiento. No obstante, no se debe descuidar la legibilidad. Si las instrucciones originales son demasiado complicadas y no tienen comentarios, modificarlas a posteriori es muy difícil.

¿QUÉ REGLAS SINTÁCTICAS SON VALIDAS PARA LAS EXPRESIONES REGULARES?

Las expresiones regulares se pueden aplicar en diversos lenguajes, tales como Perl, Python, Ruby, JavaScript, XML o HTML, por lo que los usos o funciones pueden llegar a ser muy diferentes. Así, una expresión regular puede estar formada por hasta tres partes, independientemente del lenguaje en el que se va a utilizar: **pattern** [patrón de búsqueda], **delimiter** [delimitador] y **modifier** [modificador].

CASOS DE USO

COINCIDENCIA SOLO CON LA FRASE EXACTA

Ejemplo de uso: Coincidencia con la frase recetas deliciosas.

Ejemplo de expresiones regex:

- `[\W|^]recetas\sdeliciosas[\W|$]`
- `[\W|^]recetas\s{0,3}deliciosas[\W|$]`
- `[\W|^]receta[s]{0,1}\s{0,3}deliciosa[s]{0,1}[\W|$]`

COINCIDENCIA CON LA PALABRA O CON LA FRASE DE UNA LISTA

Ejemplo de uso: Coincidencia con cualquier palabra o frase de la lista siguiente:

- tontería
- maldito
- caray
- caramba
- madre mía
- ostras

Expresión regex: `[?i][\W|^](tontería|maldito|caray|caramba|madre\s mía|ostras)[\W|$]`

COINCIDENCIA DE UNA PALABRA CON DISTINTAS GRAFÍAS O CON CARACTERES ESPECIALES

Ejemplo de uso: Coincidencia con las palabras dinero rápido y algunas de las alteraciones que utilizan los spammers, como:

- dinero r@pido
- din€ro r@pido
- din€ro rápido

Expresión regex: `din[eE€]ro +r[á4@]pido`

COINCIDENCIA CON CUALQUIER DIRECCIÓN DE CORREO DE UN DETERMINADO DOMINIO

Ejemplo de uso: Coincidencia con cualquier dirección de correo electrónico de los dominios yahoo.com, hotmail.com y gmail.com.

Expresión regex: `[\W|^][\w.\-]{0,25}@[yahoo|hotmail|gmail]\.com[\W|$]`

CONCLUSIONES

Las expresiones regulares son una herramienta versátil y poderosa con un amplio rango de aplicaciones en diversos campos. Su dominio puede aumentar significativamente la eficiencia y productividad en tareas relacionadas con el manejo de texto.