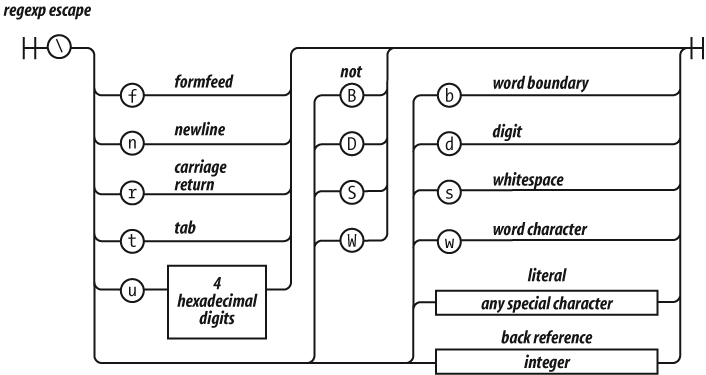
Elements

## Escape Regexp



El carácter barra invertida indica escape tanto en los factores regexp como en las cadenas, pero en los factores regexp funciona de forma un poco diferente.

Como en las cadenas, \f es el carácter de salto de línea, \n es el carácter de nueva línea, \r es el carácter de retorno de carro, \t es el carácter de tabulación y \u permite especificar un carácter Unicode como una constante hexadecimal de 16 bits. En los factores regexp, \b*is not* es el carácter de retroceso.

\d es igual que [0-9]. Coincide con un dígito. \D es lo contrario: [^0-9].

\s es lo mismo que [\f\n\r\t\u000B\u0020\u00A0\u2028\u2029]. Es un conjunto parcial de caracteres Unicode de espacio en blanco. \S es lo contrario: [^\f\n\r\t\u000B\u0020\u00A0\u2028\u2029].

\w es lo mismo que [0-9A-Z\_a-z]. \W es lo contrario: [^0-9A-Z\_a-z]. Se supone que representa los caracteres que aparecen en las palabras. Por desgracia, la clase que define es inútil para trabajar con prácticamente cualquier lenguaje real. Si necesitas emparejar una clase de letras, debes especificar tu propia clase.

Una clase de letra sencilla es [A-Za-z\u00C0-\u1FFF\u2800-\uFFFD]. Incluye todas las letras de Unicode, pero también miles de caracteres que no son letras. Unicode es grande y complejo. Una clase de letras exacta del Plano Básico Multilingüe es posible, pero sería enorme e ineficaz. Las expresiones regulares de JavaScript proporcionan un soporte extremadamente pobre para la internacionalización.

\b pretendía ser un anclaje de límites de palabras que facilitara la búsqueda de coincidencias de texto en los límites de las palabras. Por desgracia, utiliza \w para encontrar los límites de las palabras, por lo que es completamente inútil para las aplicaciones multilingües. Esta parte no es buena.

\1 es una referencia al texto que fue capturado por el grupo 1, de modo que se pueda cotejar de nuevo. Por ejemplo, podrías buscar palabras duplicadas en el texto con:

var doubled\_words = /([A-Za-z\u00C0-\u1FFF\u2800-\uFFFD]+)\s+\1/gi;

doubled\_words busca apariciones de palabras (cadenas que contienen 1 o más letras) seguidas de espacios en blanco seguidos de la misma palabra.

\2 es una referencia al grupo 2, \3 es una referencia al grupo 3, y así sucesivamente.

[anterior](ch07s03_4.html)[Subtema 5 de 9: (Ver todo)](ch07s03.html)[siguiente](ch07s03_6.html)