EXPRESIONESREGULARES

24 AGOSTO

Grupo conformado por:

• Juan Fernando Alzate Gómez

Diseño de expresiones regulares

Entero, Real, Notación Científica e Email

❖ Entero

^[-+]?[0-9]+\$

La anterior expresión regular describe, cualquier número entero. La parte inicial con el símbolo ^ inidica que la primera expresión consecuente a éste, deberá estar obligatoriamente al inicio de cualquier cadena de entrada que le sea suministrada; dicha expresión es [-+]?. La expresión regular antes mencionada significa que puede estar el símbolo + o el símbolo -, 0 o 1 veces, cantidad de veces que están referenciadas después de la expresión por el símbolo ?.

Seguido de la expresión inicial sigue [0-9]+\$, que es una ocurrencia de un número entre el 0 al 9; como después de esta, va el signo +, esto indica que como mínimo para ser un número entero deben haber 1 o más ocurrencias de un número de entre el 0 al 9. Cabe resaltar que en la primera parte de la expresión regular en general, la subexpresión [+-]? permite que el espacio vacio se dé, que es otra forma de representar a un número positivo a parte del +. Esto ya que si ocurre 0 veces representa a la cadena vacia. Y el símbolo \$ consecuente a la anterior expresión regular indica que debe al final terminar con un número del 0 al 9.

❖ Real

^[-+]?[0-9]+\.[0-9]+\$

Esta expresión regular hace referencia a los números reales. En ella se observa que al principio está la subexpresión ^[-+]?, que al igual que en la expresión regular de los números enteros, esta parte de la expresión regular indica que al inicio por el símbolo ^ debe haber ya sea un símbolo + o un símbolo – o ninguno de los dos por el símbolo ? (0 o 1 veces), luego la subexpresión [0-9]+ refiere a que debe haber 1 o más veces números de entre el 0 al 9. La posterior expresión

\. Indica con el carácter de escape \ que debe ser consecuente a la cantidad de números puesta anteriormente un solo carácter punto. Luego, al final se repite la expresión [0-9]+ pero con un \$ al final, que indica que la cadena debe terminar con un número.

❖ Notación Científica

La expresión de notación científica se compone de varias partes: ^[+-]?[0-9]+(\.([0-9])+)? y [eE][+-]?([0-9]+)\$. La primera expresión es la misma de los numeros reales, con la excepción de que la subexpresión (\.([0-9])+)?, admite que después de la parte entera pueda haber o no parte decimal del número, lo que convertiria al número, en real. La segunda supexpresión indica que se debe posterior a la primera subexpresión, poner un carácter e o E, que refleja lo que en notación científica es la potencia de 10. Seguido de esta, va la misma expresión regular que hay para los números enteros, ya que las potencias son enteras (de 1 a más números de entre el 0 al 9). El símbolo del final \$ recalca que la última expresión antes del símbolo deber ser con la que finalice una cadena del tipo notación científica, que en este caso es un número.

❖ Email

La primera subexpresión, que va antes del arroba (@), ^[a-z](((\.|_)?([a-z]|[0-9]){1,3}){3,6}), indica que el correo electrónico debe iniciar con una letra (condición que establece el símbolo ^) y posteriormente puede o no tener separadores (.,_) pero si obligatoriamente caracteres alfanuméricos (números o letras) siendo una cantidad máxima de 18 caracteres alfanuméricos si es que no se pone ningun separador, y con separadores máximo 21 caracteres bajo la condición de que los caracteres alfanumericos después de los separadores sean entre 3 a 18. Teniendo en cuenta lo anterior además admite un máximo de 3 veces de uso de separadores. La subexpresión que sigue después del arroba [a-z]{4,8}((\.[a-z]{3,9})(\.[a-z]{2})?)\$ hace referencia al nombre del tipo de correo (gmail, outlook, yahoo, hotmail, etc) con un mínimo de 4 y máximo de 8 letras, y

a dos posibles extensiones que como máximo se le pueden agregar a un correo electrónico (.edu.co, .net, .com, .com.co, etc) en donde la primera es de entre 3 a 9 caracteres de longitud después del punto, y la segunda extensión usualmente hace referencia a la abreviatura del país de un correo (.co, .uk, .us, .mx, etc) por lo cual solo admite 2 caracteres alfabéticos después del punto; sin embargo esta segunda extensión es opcional, situación que está determinada por el símbolo ?. Adicionalmente el símbolo \$ indica que la primera extensión de correo o la segunda van al final de la cadena enviada.