Unidad II: Expresiones regulares

2.1. Definición formal de una ER

Una expresión regular (ER) sobre un alfabeto finito Σ se define recursivamente como sigue:

1. Para todo c ∈ Σ, c es una ER
2. Φ es una ER
3. Si E1 y E2 son ERs, E1 | E2 es una ER
4. Si E1 y E2 son ERs, E1 · E2 es una ER
5. Si E1 es una ER, E1 ⋆ es una ER
6. Si E1 es una ER, (E1) es una ER

Cuando se lee una expresión regular, hay que saber qué operador debe leerse primero.

Esto se llama precedencia. Por ejemplo, la expresión a | b · c ⋆, ¿debe entenderse como (1) la "⋆" aplicada al resto? (2) ¿la "|" aplicada al resto? (3) ¿la "·" aplicada al resto? La respuesta es que, primero que nada se aplican los "⋆", segundo los "·", y finalmente los "|".

Esto se expresa diciendo que el orden de precedencia es ⋆, ·, |. Los paréntesis sirven para alterar la precedencia. Por ejemplo, la expresión anterior, dado el orden de precedencia que establecimos, es equivalente a a | (b · (c ⋆)). Se puede forzar otro orden de lectura de la ER cambiando los paréntesis, por ejemplo (a | b) · c ⋆.

Asimismo, debe aclararse cómo se lee algo como a|b|c, es decir ¿cuál de los dos "|" se lee primero? Convengamos que en ambos operadores binarios se lee primero el de más a la izquierda (se dice que el operador "asocia a la izquierda"), pero realmente no es importante,

por razones que veremos enseguida.

Observar que aún no hemos dicho qué significa una ER, sólo hemos dado su sintaxis pero no su semántica.

2.2. Operaciones

Ofrecen algo más que os autómatas no:

- Manera declarativa de expresar las cadenas que queremos aceptar.

Tipos:

UNION.- Si a y b son expresiones regulares, a|b es una expesión regular ta que: {a y b} = a|b, es decir que puede aparecer o no indistintamente.

Símbolo: **|**

CONCATENACION: Si a y b son expresiones regulares, ab es una expresión regular tal que:

{a y b} = {a}{b}

Símbolo: **.**

CIEREE U OPERACIÓN ESTRELLA: Si a es una expresión regular, entones a* es una expresión regular que denota {a}*. Es decir que denota as cadenas:

a

aa

aaa…a

Es decir puede no venir por lo cual seria un conjunto vacio o venir repetidamente

Símbolo: *

CIERRE POSITIVO: Si es una expresión regular, entonces a+ es una expresión regular que denota {a}+. Es decir denota las cadenas:

a

aa

aaa….a

Símbolo: +

2.3. Aplicaciones en problemas reales

Facilitar las construcciones de un compilador.

Validar la sintaxis de un programa

Algunos generadores lexicográficos toman como entrada una sucesión de expresiones regulares y produce un autómata finito que reconozca cualquier expresión ahí descritos.

Algunos editores de texto y programas similares permiten la substitución de una cadena por otra cualquiera que cumpla con la expresión regular dada.

Correctores ortográficos para encontrar patrones en el genoma humano.