Lenguajes y Autómatas

Carlos Barrón Romero

Notación y recordatorios:

- Conjuntos por extensión: Es una lista de elementos bien distinguidos entre llaves. Ejemplo: $A = \{1,2,a\}$.

- Conjunto por regla: Es un conjunto denotado por una regla o proposición lógica para elementos de un conjunto dado. Ejemplo: $B = \{s_1 s_2 \mid s_1, s_2 \in \Sigma\}$.

- Un alfabeto es un conjunto finito de símbolos. Ejemplo de un alfabeto: $\Sigma = \{0,1\}$.

Ejercicio. Con los datos anteriores encontrar B. Respuesta: $B = \{00,01,10,11\}$.

- La Cardinalidad de un conjunto es su número de elementos. Se denota por **| |**, $|\cdot| : P(\Omega) \to \mathbb{N}$**.** Donde $|P(\Omega)$ denota el conjunto potencia de Ω. Otra forma de denotar al conjunto potencia es 2^{Ω}. El conjunto potencia consiste de todos los subconjuntos de un conjunto dado.

Ejemplos: Sea $\Omega = \{0,1\}$. Entonces $2^{\Omega} = \{\phi, \{0\}, \{1\}, \{0,1\}\}$.
Note que los subconjuntos corresponden a las combinaciones (sin repetición, la fórmula es $\binom{n}{k} = \frac{n!}{(n-k)!k!}$): $\binom{2}{0} = 1, \binom{2}{1} = 1, \binom{2}{2} = 1$.
$|\phi| = 0$.
$|\{a\}| = 1$.
$|2^{\{0,1\}}| = |\{\phi, \{0\}, \{1\}, \{0,1\}\}| = 4$.
Note que se cumple la fórmula $|P(\Omega)| = 2^{|\Omega|}$, para $|\Omega| < \infty$.

Expresiones Regulares (ER).
Operaciones

- Concatenación: Se refiere a las cadenas de símbolos que se pegan en el orden que aparecen. Ejemplo: concatenar 0 y 1, resulta 01. Concatenar 1 y 0, resulta10.

Como 01 ≠ 10, se tiene que la concatenación no es una operación conmutativa.

El conjunto vacío **ϕ** (en negrita) es la ER que corresponde con el conjunto vacio {}.

El símbolo nulo, ε, es un símbolo especial que no tiene longitud y si aparece no altera a una cadena. Por ejemplo: ε0ε0 = 00, ε01εε00 = 100 (aparece y desaparece en cualquier cadena).

La longitud de Cadena es el número de símbolos en la cadena. Se denota $| \, |$, $|\cdot|: \Sigma^* \to \mathbb{N}$, donde Σ^* (Cerradura de Kleene, la explicación viene después) es cualquier cadena de símbolos de incluyendo al símbolo nulo. Ejemplos. |ε0ε0| = |00|=2, |ε01εε00| = |100|=3, |01|=2.

Note que $\varepsilon \notin \Sigma$, para cualquier alfabeto.

Expresiones Regulares (ER) y sus conjuntos.
Dado un Σ (alfabeto) todos los subconjuntos de sus símbolos son ER y se denotan por letra negrita o bold o letra gorga).

Ejemplo. Sea Σ={a,b,0,1}.

ER	Conjunto
a	{a}
0	{0}
Ø	{ }, conjunto regular de cero elementos.
Σ	{a, b,0,1}

La concatenación de dos ER es unia ER y corresponde con el conjunto de todas las cadenas concatenadas entre la ER dadas.

ER	Conjunto
a0	{a0}
0b	{0b}
Ø1	{1}
aΣ	{aa, ab, a0,a1}

La unión de ER es una ER y corresponde con la unión de los conjuntos regulares. Se denota por +.

ER	Conjunto
a+**0**	{a, 0}
0+**b**	{0, b}
Ø+**1**	{1}
a+**Σ**	{a, b, 0, 1}

La cerradura de Kleene es la ER con un * y corresponde con el conjunto regular que es la unión de todas las potencias de concatenación de los símbolos del conjunto regular.
Por ejemplo:

ER	Conjunto
$\mathbf{0}^*$	{ε, 0, 00, 000, ...}

Note que $\mathbf{0}^* = 0^0 + 0^1 + 0^2 + 0^3 + \ldots$
Sea Σ_2 = {0, 1}
Σ_2^* = {ε, 0, 1, 00, 01, 10, 11, 000, …}
$\Sigma_2^* = (\mathbf{0} + \mathbf{1})^*$

Observación: Las ER denotan tipos de palabras (por ejemplo, número natural) y son comúnmente usados en análisis léxico (Lex) en los compiladores.

Sea Σ_3 = {0,1,2}, dígitos de los números en base 3.

O sea: $3^0 d_0 + 3^1 d_1 + 3^2 d_2 + \cdots + 3^n d_n =$ $d_n \cdots d_2 d_1 d_{0_3}$ (número en base 3), donde $d_i \in \Sigma_3, i = 0,1,2,\ldots,n$. Por ejemplo $121_3 = 3^2*1 + 3^1*2 + 3^0*1 = 16$.

Todos los números naturales en la base 3 corresponden (con algunas repeticiones) con Σ_3 (Σ_3)*

Nota: El subíndice 3 no se incluye, pero se puede incluir.

En forma equivalente con la ER, **d** = {0, 1, 2}.

d* = d^0 + **d**^{1} + **d**^{2} + d^3 + **d**^{4} + … son los números naturales en la base 3.

Solo los números naturales en base 3 de longitud 4 se representan con conjunto regular: **dddd** (que corresponde a la potencia de concatenación 4, d^4 . Note que d^4 no es ER.). **dddd** ={0000, 0001, 0002, 0010, ..., 2222}.

Teorema de Klene

Toda ER es reconocida por autómatas finitos determinísticos (AFD), recíprocamente, los AFD solo reconocen expresiones regulares.

Lenguajes son todo conjunto de cadenas o palabras. **L**= {cadenas}

Una representación de un lenguaje se puede hacer por ER.

Ejemplo. Definir el lenguaje I como la expresión regular de un tipo variable o identificador de variable que comienza en letra y continua con letras o dígitos.

Sea Σ={a ,b, c, …z ,0, 1, …, 9}

L = **a** + **b** + **c** +…+ **z** (**L** de letra)

D = **0**+**1**+…+**8**+**9** (**D** de dígito)

El lenguaje pedido es I = (**L**+**D**) (**L**+**D**)* , ya que (**L**+**D**) (**L**+**D**)* = {a, b, c, d, ..., z, 0, 1, 2, 3, ..., 9, aa, ab, ac, ad, ..., az, a0, a1, a2, a3, ..., a9, ba, bb, bc, ..., bz, b0, b1, b2, b3, ..., b9, ...}.

Sea **C** = **b** + **c** + **d** y **V** = **a** + **e** + **i** + **o** + **u**. Calcular el número de palabras de **CVCVCV.**

Note **|CV|** = (3*5), y como son tres se tiene, **|CVCVCV| =** 15^3 **= 3375.**

Sea Σ_2 = {0,1}

(**0**+ **1**)* **0** representa a los números pares.

(**0**+ **1**)* **1** representa a los números impares.

AFD reconocedor de los números pares en binario.

AFD=$(Q, a, \Sigma, \delta, F)$ donde Q = {a, b}, Σ={0, 1}, F = {b}.

Tabla de transición de $\delta : Q \times \Sigma \to Q$		
Q	Σ	Q
$\to$a	0	<u>b</u>
a	1	a
b	0	<u>b</u>
b	1	a

Demostrar si se acepta o rechaza 001.

a001 → b01 → b1 → a se rechaza.

Demostrar si se acepta o rechaza 10.

a10 → a0 → b se acepta.

Expresiones regulares (ER) vs Autómatas finitos determinísticos

Proposición. Un AFD siempre genera una ER.

Explicación. Un AFD para un símbolo, s, su ER es **s**

Sea un AFD con 2 estados, Q={a,b}, Σ={s} y un estado final, F={b}.

Tabla de transición de $\delta : Q \times \Sigma \to Q$		
Q	Σ	Q
$\to$a	s	<u>b</u>

solo reconoce la ER **s**.

Sea un AFD con 2 estados, Q={a,b}, Σ={0, 1} y un estado final, F={b}.

Con

Tabla de transición de $\delta : Q \times \Sigma \to Q$		
Q	Σ	Q
$\to$a	0	<u>b</u>
b	1	a

Reconoce 0, 010, 01010, 0101010,

Note que el ejemplo anterior muestra que si un AFD de 2 estados reconoce una cadena de tamaño 3 entonces reconoce la ER **0**(**10**)*.
Este resultado se conoce como el lema del bombeo. A fuerzas repite (una, dos, tres o mas veces) una subcadena.

Otro ejemplo:
Sea un AFD con 2 estados, Q={a,b}, Σ={s} y un estado final, F={b}. Con

Tabla de transición de $\delta : Q \times \Sigma \to Q$		
Q	Σ	Q
$\to$a	s	<u>b</u>
b	s	<u>b</u>

Este AFD acepta la ER: **ss***
Otro ejemplo:
Sea un AFD con un estados, Q={a}, Σ={s} y un estado final, F={a}. Con

Tabla de transición de $\delta : Q \times \Sigma \to Q$		
Q	Σ	Q
$\to$a	s	<u>a</u>

Este AFD acepta la ER: **s***

En conclusión un AFD genera una ER.

El teorema de Kleene en su forma moderna, muestra el ciclo siguiente: L(AFD) → L(ER) → L(AFN-ε) →L(AFN) → L(AFD).

Notas finales de clase:

1. Hoy vimos ER y L(AFD) → L(ER).

2. Anteriormente, vimos la parte L(AFN) → L(AFD), que corresponde con el Método de la Potencia. El reciproco: L(AFD) → L(AFN) vimos que era elemental, cambiando en la tabla de transición de δ_N del AFN del estado resultante al conjunto con el estado de resultante δ de la tabla del AFD. O sea, si se tiene (q,s,r) en el AFD se cambia a (q,s,{r}) en el AFN.

3. La próxima clase: L(AFN-ε) →L(AFN). Y se termina el Teorema de Kleene.