Tema 1: Lenguajes Regulares

Lección 1.4

Expresiones regulares y lenguajes regulares

Jordi Bernad

Índice

1. Expresiones regulares
2. Equivalencia entre e.r. y AFD
3. Construir AFD equivalente a e.r.
4. Construir e.r equivalente a AFD

Recordad definición expresión regular

Definición

Una expresión r se dice que es una **expresión regular (e.r.)** si es igual a:

- $r = a$, donde a es un símbolo para algún alfabeto Σ.
- $r = \epsilon$
- $r = \emptyset$
- $r = (r_1 + r_2)$, donde r_1 y r_2 son expresiones regulares.
- $r = (r_1 \cdot r_2)$, donde r_1 y r_2 son expresiones regulares.
- $r = (r_1^*)$, donde r_1 es una expresión regular.

Recordad lenguaje representado por una e.r.

Definición

Dada una expresión regular r, el **lenguaje representado** por r se denota por $L(r)$ y es igual a

- $L(a) = \{a\}$, si a es un símbolo de un alfabeto Σ.
- $L(\epsilon) = \{\epsilon\}$
- $L(\emptyset) = \emptyset$
- $L(r_1 + r_2) = L(r_1) \cup L(r_2)$
- $L(r_1 \cdot r_2) = L(r_1) \cdot L(r_2)$
- $L(r^*) = L(r)^*$

Relación entre e.r. y lenguajes regulares

- ¿Qué relación existe entre los lenguajes representados por e.r. y los lenguajes regulares?
- Demostraremos que son los mismos.

Teorema

- Si r es una e.r. regular, entonces $L(r)$ es regular.
- Si L es un lenguaje regular, entonces existe una e.r. r tal que $L(r) = L$

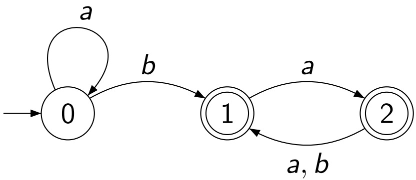
r e.r. $\Rightarrow L(r)$ regular

Denotemos por $|r|$ el número de caracteres que forman la e.r.
Demostremos por inducción en el número de caracteres:

1. Si $|r| = 1$, entonces $r = a$, $r = \epsilon$, $r = \emptyset$. Por tanto, $L(a) = \{a\}$, $L(\epsilon) = \{\epsilon\}$, $L(\emptyset) = \emptyset$ son regulares: trivial, son conjuntos finitos.
2. Si $|r| = n > 1$, $r = (r_1 + r_2)$, $r = (r_1 \cdot r_2)$, $r = (r_1^*)$
   - $L(r_1 + r_2)$ es regular: como $|r_1|, |r_2| < n$, $L(r_1)$ y $L(r_2)$ son regulares, y la unión de lenguajes regulares es regular $L(r_1 + r_2) = L(r_1) \cup L(r_2)$ es regular
   - $L(r_1 \cdot r_2)$ es regular: como $|r_1|, |r_2| < n$, $L(r_1)$ y $L(r_2)$ son regulares, y la concatenación de lenguajes regulares es regular $L(r_1 \cdot r_2) = L(r_1) \cdot L(r_2)$ es regular
   - $L(r_1^*)$ es regular: como $|r_1| < n$, $L(r_1)$ es regular. , y la estrella de Kleene de lenguajes regulares es regular $L((r_1)^*) = L(r_1)^*$ es regular.

L regular $\Rightarrow$ existe r e.r. tal que $L = L(r)$

- Tenemos un AFD $M = \{Q, \Sigma, \delta, q_0, F\}$

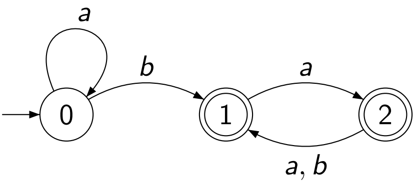


- Para cada estado $q \in Q$, L_q es el conjunto de cadenas que se aceptan empezando en el estado q

$$L_q = \{w \mid \delta(q, w) \in F\}$$

- Notar que el lenguaje reconocido por M es $L_0 = L(M)$.

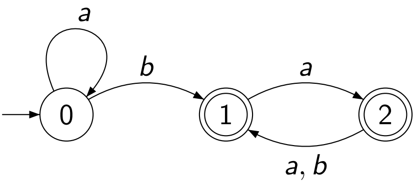
L regular $\Rightarrow$ existe r e.r. tal que $L = L(r)$



- Las cadenas aceptadas desde el estado 0 son de la forma:
  - una a seguida de cualquier cadena aceptada desde el estado 0
  - una b seguida de cualquier cadena aceptada desde el estado 1
- Podemos expresar esta igualdad de conjuntos

$$L_0 = aL_0 + bL_1$$

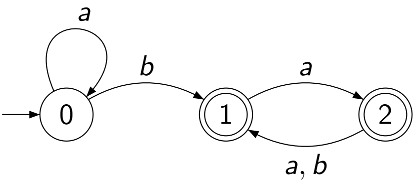
L regular $\Rightarrow$ existe r e.r. tal que $L = L(r)$



- Las cadenas aceptadas desde el estado 1 son de la forma:
  - una a seguida de cualquier cadena aceptada desde el estado 2
  - Como 1 es un estado final, también se acepta ϵ
- Podemos expresar esta igualdad de conjuntos

$$L_1 = aL_2 + \epsilon$$

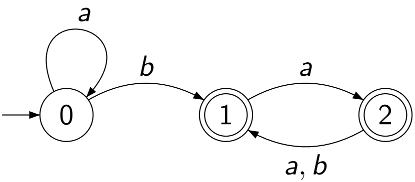
L regular $\Rightarrow$ existe r e.r. tal que $L = L(r)$



- Las cadenas aceptadas desde el estado 2 son de la forma:
  - una a seguida de cualquier cadena aceptada desde el estado 1
  - una b seguida de cualquier cadena aceptada desde el estado 1
  - Como 2 es un estado final, también se acepta ϵ
- Podemos expresar esta igualdad de conjuntos

$$L_2 = aL_1 + bL_1 + \epsilon$$

L regular $\Rightarrow$ existe r e.r. tal que $L = L(r)$



- Planteamos un "sistema de ecuaciones"

$$\begin{aligned} L_0 &= aL_0 + bL_1 \\ L_1 &= aL_2 + \epsilon \\ L_2 &= aL_1 + bL_1 + \epsilon \end{aligned}$$

- ¿Cómo resolvemos el sistema? Tenemos que "despejar" L_0

L regular $\Rightarrow$ existe r e.r. tal que $L = L(r)$

- Podemos sustituir L_2 en la ecuación $L_1 = aL_2 + \epsilon$

$$\begin{aligned} L_0 =& aL_0 + bL_1 \\ L_1 =& a((aL_1 + bL_1 + \epsilon) + \epsilon \end{aligned}$$

- Equivalentemente

$$\begin{aligned} L_0 =& aL_0 + bL_1 \\ L_1 =& a(a + b)L_1 + (a + \epsilon) \end{aligned}$$

- Para despejar L_1 necesitamos el siguiente resultado.

Lema de Arden

Sean A, B dos lenguajes con $\epsilon \notin A$. Si X es un lenguaje desconocido tal que $X = AX + B$, entonces la única solución es $X = A^*B$.

L regular $\Rightarrow$ existe r e.r. tal que $L = L(r)$

- En nuestro sistema de ecuaciones

$$\begin{aligned} L_0 =& aL_0 + bL_1 \\ L_1 =& a(a+b)L_1 + (a+\epsilon) \end{aligned}$$

- Aplicando el lema de Arden podemos obtener L_1

$$L_1 = (a(a+b))^*(a+\epsilon)$$

- Sustituyendo L_1 en la primera ecuación obtenemos

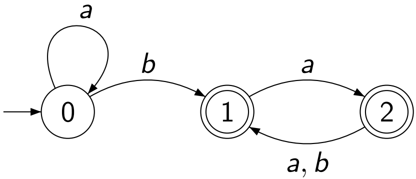
$$L_0 = aL_0 + b\big((a(a+b))^*(a+\epsilon)\big)$$

- Aplicando de nuevo el lema de Arden

$$L_0 = a^*(b((a(a+b))^*(a+\epsilon)))$$

Conclusión

- Para hallar una e.r. equivalente a un AFD, planteamos el sistema de ecuaciones y resolvemos utilizando el lema de Arden



- Expresión regular equivalente: $a^*(b((a(a+b))^*(a+\epsilon)))$

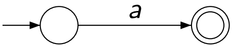
Observaciones

- ¿Cómo saber si dos e.r. representan el mismo lenguaje?
- Construimos el AFnD para cada una de las expresiones.
- Determinizamos los AFnD.
- Minimizamos los AFD, y vemos si son iguales.

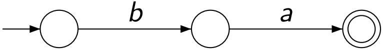
Ejemplo: ¿$(a+ba)^* = (a^*(ba)^*)^*$?

- Construir AFnD equivalente $(a+ba)^*$

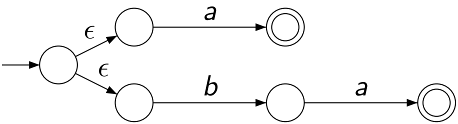
$L(a)$



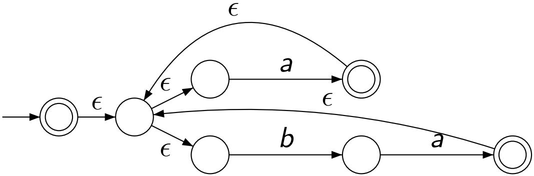
$L(ba)$



$L(a+ba)$

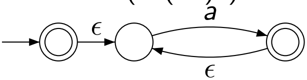


$L((a+ba)^*)$

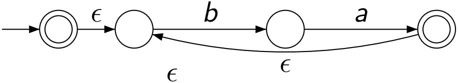


- Construir AFnD equivalente a $(a^*(ba)^*)^*$

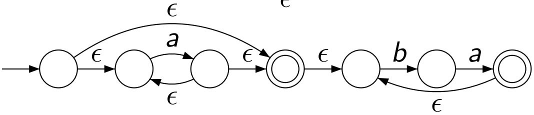
$L(a^*)$



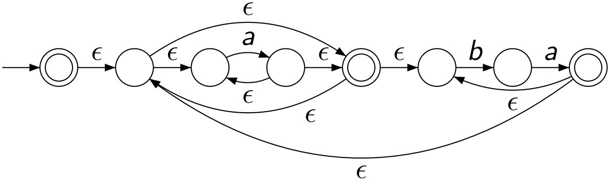
$L((ba)^*)$



$L(a^*(ba)^*)$



$L((a^*(ba)^*)^*)$



Ejemplo: ¿$(a + ba)^* = (a^*(ba)^*)^*$?

- Ejercicio: determinizar y minimizar los dos AFnD

Referencias

- Kelly: Capítulo 2, sección 2.8
- Sipser: Capítulo 1, sección 1.3