Pontificia Universidad Católica de Chile
Escuela de Ingeniería
Departamento de Ciencia de la Computación

IIC2223/IIC2224 – Teoría de Autómatas y Lenguajes Formales – 2' 2024

Tarea 2

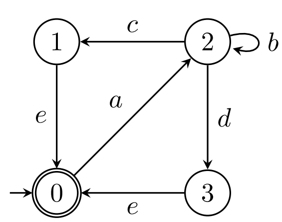
Publicación: Viernes 6 de septiembre.
Entrega: **Jueves 12 de septiembre hasta las 23:59 horas.**

Indicaciones

- Debe entregar una solución para cada pregunta (sin importar si está en blanco).
- Cada solución debe estar escrita en LaTeX. No se aceptarán tareas escritas a mano ni en otro sistema de composición de texto.
- Responda cada pregunta en una hoja separada y ponga su nombre en cada hoja de respuesta.
- Debe entregar una copia digital por el buzón del curso, antes de la fecha/hora de entrega.
- **Se penalizará con 1 punto en la nota final de la tarea por cada regla que no se cumpla.**
- La tarea es individual.

Pregunta 1

1. Para el siguiente DFA, realice el método de eliminación de estados y encuentre una expresión regular equivalente. Detalle cada uno de los pasos, como también el autómata finito no determinista generalizado resultante después de cada etapa.

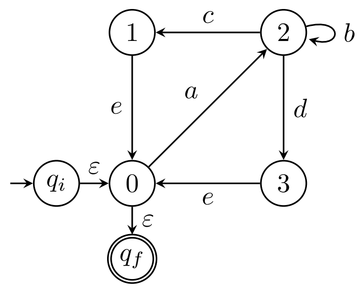


2. Sea $\mathcal{A} = (Q, \Sigma, \delta, q_0, F)$ un DFA. El método de eliminación de estados para construir una expresión regular desde $\mathcal{A}$ depende del orden $\preceq$ sobre Q escogido para eliminar estados. O sea, para dos ordenes distintos $\preceq_1$ y $\preceq_2$, podemos generar distintas expresiones regulares R_1 y R_2 si aplicamos la eliminación de estados siguiendo $\preceq_1$ o $\preceq_2$, respectivamente. Para un orden $\preceq$ de los estados Q, sea $R^{\preceq}_{\mathcal{A}}$ la expresión regular resultante de seguir el método de eliminación de estados sobre $\mathcal{A}$ según el orden $\preceq$.

   Demuestre una familia de DFAs $\{\mathcal{A}_n\}_{n\in\mathbb{N}}$ con $\mathcal{A}_n = (Q_n, \Sigma, \delta_n, q_0^n, F_n)$ sobre el mismo alfabeto Σ tal que $|Q_n| \in \Theta(n)$ (esto es, el número de estados crece lineal con respecto a n) y, para cada n, existe un orden $\preceq_n$ de Q_n tal que $R^{\preceq_n}_{\mathcal{A}_n}$ tiene tamaño $|R^{\preceq_n}_{\mathcal{A}_n}| \in \Omega(2^n)$ donde $|R|$ corresponde al número de letras y operadores en la expresión regular R.

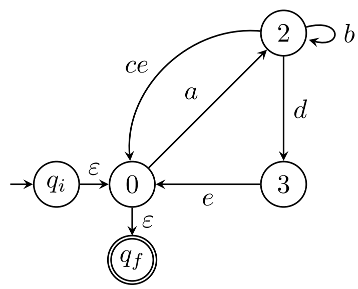
Solución

Problema 1.1 Antes de comenzar con el método de eliminación de estados, primero se debe convertir el DFA a un GNFA. Para ello, debemos agregar 2 nuevos estados q_i y q_f y conectarlos al estado 0 mediante ε-transiciones, resultando en el siguiente GNFA:

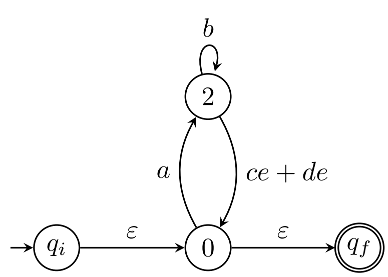


A continuación, iniciamos el método de eliminación de estados. Por simplicidad, las expresiones regulares se irán simplificando a medida que se avanza en el método, pero no se mostrará explícitamente. Por ejemplo, (εa) se simplificará a (a) y $(\varnothing + a)$ se simplificará a (a).

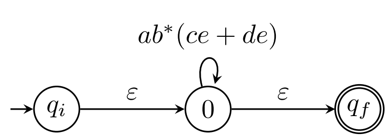
Paso 1: Eliminar el estado 1.



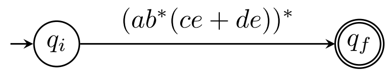
Paso 2: Eliminar el estado 3.



Paso 3: Eliminar el estado 2.



Paso 4: Eliminar el estado 0.



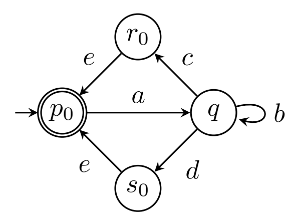
Dado que solo tenemos 2 estados, el inicial y el final, la expresión regular resultante es la expresión correspondiente a la transición entre estos dos estados, es decir, $(ab^*(ce+de))^*$.

Distribución de puntaje

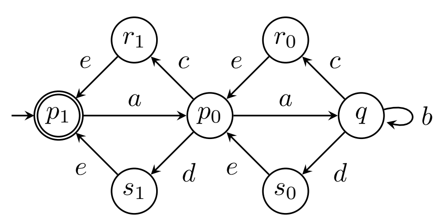
- 1 punto por convertir el DFA a GNFA.
- 1 punto por eliminar los estados 1, 3 y unir las expresiones ce y de.
- 1 punto por eliminar el estado 2 junto a su *loop*.
- 1 punto por concluir con la eliminación de estados y obtener la expresión regular.

Problema 1.2 Definimos la familia de DFAs $\{\mathcal{A}_n\}_{n\in\mathbb{N}}$ de forma inductiva de la siguiente manera:

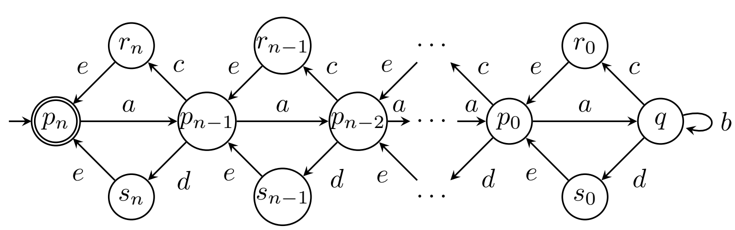
$\mathcal{A}_0$:



$\mathcal{A}_1$:



$\mathcal{A}_n$:



Formalmente, $\mathcal{A}_0 = (Q_0, \Sigma, \delta_0, q_0^0, F_0)$ donde:

- $Q_0 = \{p_0, r_0, s_0, q\}$
- $\Sigma = \{a, b, c, d, e\}$
- δ_0 es tal que:
  - $\delta_0(p_0,\ a) = q$
  - $\delta_0(q,\ b) = q$
  - $\delta_0(q,\ c) = r_0$
  - $\delta_0(q,\ d) = s_0$
  - $\delta_0(r_0,\ e) = p_0$
  - $\delta_0(s_0,\ e) = p_0$
- $q_0^0 = p_0$
- $F_0 = \{p_0\}$

Y para $n > 0$, $\mathcal{A}_n = (Q_n, \Sigma, \delta_n, q_0^n, F_n)$ donde:

- $Q_n = Q_{n-1} \uplus \{p_n, r_n, s_n\}$
- δ_n es tal que posee las mismas transiciones que δ_{n-1}, a las que se les añaden:
  - $\delta_n(p_n,\ a) = p_{n-1}$
  - $\delta_n(p_{n-1},\ c) = r_{n-1}$
  - $\delta_n(p_{n-1},\ d) = s_{n-1}$
  - $\delta_n(r_n,\ e) = p_n$
  - $\delta_n(s_n,\ e) = p_n$
- $q_0^n = p_n$
- $F_n = \{p_n\}$

Tamaño del DFA $\mathcal{A}_n$

Para mostrar que $|Q_n| \in \Theta(n)$, primero notamos que $|Q_0| = 4$, luego, como $Q_n = Q_{n-1} \uplus \{p_n, r_n, s_n\}$, tenemos que $|Q_n| = |Q_{n-1}| + 3$ para $n > 0$. Por lo tanto, $|Q_n| = 4 + 3n$, es decir, $|Q_n| \in \Theta(n)$.

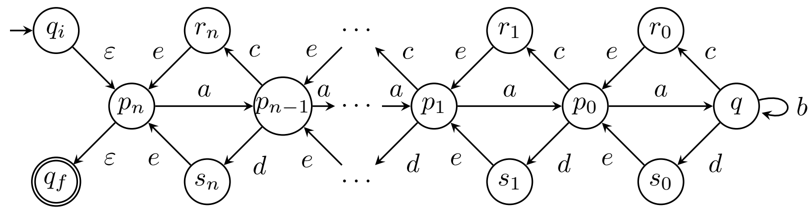
Orden $\preceq_n$ y eliminación de estados

El orden que seguiremos para eliminar los estados de $\mathcal{A}_n$ será el siguiente:

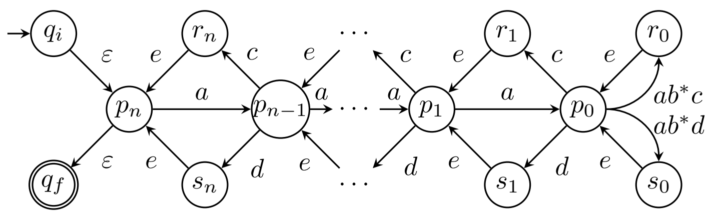
$$q \preceq_n r_0 \preceq_n s_0 \preceq_n p_0 \preceq_n r_1 \preceq_n s_1 \preceq_n p_1 \preceq_n \cdots \preceq_n r_n \preceq_n s_n \preceq_n p_n$$

Este proceso de eliminación de estados se ilustra de la siguiente manera en el algortimo de eliminación de estados para $\mathcal{A}_n$:

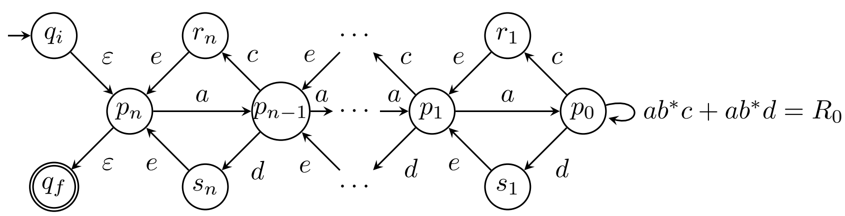
Convertir a GNFA:



Eliminar q:

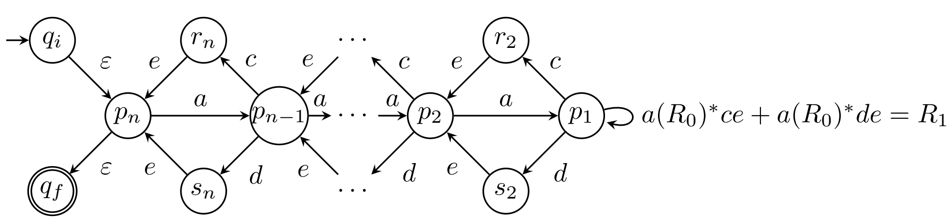


Eliminar r_0 y s_0:

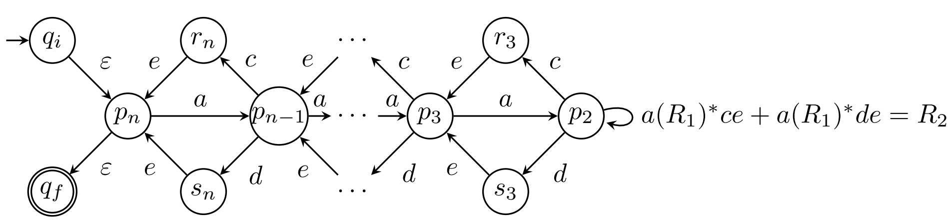


En este punto, definimos $R_0 = ab^*c + ab^*d$. Luego, notamos que debido a como se construyó $\mathcal{A}_n$, la expresión regular que describe el lenguaje aceptado por $\mathcal{A}_0$ es $R_{\mathcal{A}_0}^{\preceq 0} = (R_0)^*$.

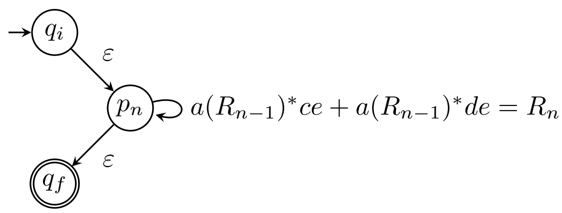
Eliminar p_0, r_1 y s_1:



Eliminar p_1, r_2 y s_2:



Eliminar hasta r_n y s_n:



Notamos que cada R_i es de la forma $a(R_{i-1})^*ce + a(R_{i-1})^*de$, por lo tanto, la cantidad de caracteres en R_i es al menos el doble de la cantidad de caracteres en R_{i-1}, lo que implica un crecimiento exponencial en la cantidad de caracteres de R_n con respecto a n, es decir, $|R_n| \in \Omega(2^n)$.

Distribución de puntaje

- 1 punto por definir correctamente la familia de DFAs.
- 1 punto por demostrar que $|Q_n| \in \Theta(n)$.
- 1 punto por presentar el orden de eliminación de los estados.
- 1 punto por demostrar el orden genera que $|R^{\preceq_n}_{\mathcal{A}_n}| \in \Omega(2^n)$.

Pregunta 2

Sean $a, b, c \in \mathbb{N}$ tal que $a > 0$. Para cada uno de los siguientes lenguajes sobre el alfabeto $\Sigma = \{\mathtt{1}\}$, diga si el lenguaje es regular o no. Demuestre su afirmación.

1. $L_1 = \{\mathtt{1}^{b \cdot n + c} \mid n \geq 0\}$
2. $L_2 = \{\mathtt{1}^{a \cdot n^2 + b \cdot n + c} \mid n \geq 0\}$

Solución

Problema 2.1 El lenguaje L_1 es regular. Para demostrarlo construiremos una expresión regular que lo genere. Al tener $\mathtt{1}^{b \cdot n + c}$ es claro que esto significa $(\mathtt{1}^n)$ b veces, seguido de c veces $\mathtt{1}$. Ejemplo de esto es:

- $\mathtt{1}^c$, $n = 0$
- $\mathtt{1}^{b+c}$, $n = 1$
- $\mathtt{1}^{2b+c}$, $n = 2$
- $\mathtt{1}^{3b+c}$, $n = 3$

Y así sucesivamente.
Este tipo de patrón se puede reconocer con la expresión regular $R_1 = (\mathtt{1}^b)^* \cdot \mathtt{1}^c$. Por lo tanto, L_1 es regular.

Demostración.

- $L_1 \subseteq \mathcal{L}(R_1)$.

  Sea $w \in L_1$, entonces $w = \mathtt{1}^{b \cdot n + c}$ para algún $n \geq 0$. Por lo tanto, $w = \mathtt{1}^b \cdot \mathtt{1}^b \cdot \mathtt{1}^b \cdot \ldots \cdot \mathtt{1}^b \cdot \mathtt{1}^c$, donde $\mathtt{1}^b$ se repite n veces. Esto es equivalente a $(\mathtt{1}^b)^* \cdot 1^c$. Por lo tanto, $w \in \mathcal{L}(R_1)$.

- $\mathcal{L}(R_1) \subseteq L_1$.

  Sea $w \in \mathcal{L}(R_1)$, entonces $w = (\mathtt{1}^b)^* \cdot \mathtt{1}^c$. Por lo tanto, $w = \mathtt{1}^b \cdot \mathtt{1}^b \cdot \mathtt{1}^b \cdot \ldots \cdot \mathtt{1}^b \cdot \mathtt{1}^c$, donde $\mathtt{1}^b$ se repite 0 o más veces. Esto es equivalente a $\mathtt{1}^{b \cdot n + c}$. Por lo tanto, $w \in L_1$.

Por lo tanto, $L_1 = \mathcal{L}(R_1)$ y L_1 es regular.

Distribución de puntaje

- 1 punto por decir que el lenguaje es regular.
- 2 puntos por la expresión regular o autómata que genera el lenguaje.
- 1 punto por argumentar o demostrar correctitud de la expresión regular o autómata.

Problema 2.2 Necesitamos demostrar que el lenguaje que consiste en todas las palabras de longitud cuadrática de la forma $a \cdot n^2 + b \cdot n + c$ no es regular. Para esto, podemos utilizar el contrapositivo del lema de bombeo.

- Para todo $N > 0$
- Existe una palabra $x \cdot y \cdot z \in L_2$ con $|y| \geq N$ tal que
- Para todo $u \cdot v \cdot w = y$ con $v \neq \epsilon$
- Existe un $i \geq 0$ tal que $x \cdot u \cdot v^i \cdot w \cdot z \notin L_2$

Sea $N > 0$ cualquiera. Considere $n \in \mathbb{N}$ lo suficientemente grande tal que $N < n < a \cdot n^2 + b \cdot n + c$. Con este n, definimos la palabra $w = \mathtt{1}^{a \cdot n^2 + b \cdot n + c}$. Es claro ver que $w \in L_2$.

Después, elegimos la descomposición $w = x \cdot y \cdot z \in L_2$ con $x = \mathtt{1}^{a \cdot n^2 + b \cdot n + c - N}$, $y = 1^N$ y $z = \epsilon$. Lo anterior está bien definido ya que $a \cdot n^2 + b \cdot n + c - N > 0$ por definición de n.

Ahora tomamos una descomposición cualquiera de la forma $y = 1^k \cdot 1^l \cdot 1^m$ con $l > 0$ y $k + l + m = N$. Si bombeamos con $i = 2$, obtendremos la palabra $x \cdot u \cdot v^2 \cdot w \cdot z = \mathtt{1}^{a \cdot n^2 + b \cdot n + c - N} \cdot 1^k \cdot 1^{2 \cdot l} \cdot 1^m$.

Observando el exponente de la palabra, tenemos que

$$a \cdot n^2 + b \cdot n + c - N + k + l + m + l = a \cdot n^2 + b \cdot n + c - N + N + l = a \cdot n^2 + b \cdot n + c + l$$

Ahora, necesitamos argumentar que esta palabra no pertenece a L_2. Como $0 < l \leq N < n$, entonces:

$$\begin{aligned} a \cdot n^2 + b \cdot n + c &< a \cdot n^2 + b \cdot n + c + l \\ &\leq a \cdot n^2 + b \cdot n + c + N \\ &< a \cdot n^2 + b \cdot n + c + n \end{aligned}$$

Luego, como sabemos que $n \leq 2 \cdot a \cdot n + b + a$ dado que $a \geq 1$ y $b \geq 0$, tenemos que

$$\begin{aligned} a \cdot n^2 + b \cdot n + c + n &\leq a \cdot n^2 + b \cdot n + c + 2 \cdot a \cdot n + b + a \\ &\leq a \cdot (n+1)^2 + b \cdot (n+1) + c \end{aligned}$$

De ambas desigualdades concluimos que:

$$a \cdot n^2 + b \cdot n + c < a \cdot n^2 + b \cdot n + c + l < a \cdot (n+1)^2 + b \cdot (n+1) + c$$

Por lo tanto, la palabra $x \cdot u \cdot v^i \cdot w \cdot z$ ya no corresponde a un número de la forma $a \cdot n^2 + b \cdot n + c$ para ningún $n \in \mathbb{N}$, lo que contradice la suposición de que L_2 es regular. Por lo tanto, L_2 no es regular.

Distribución de puntaje

- 1 punto por escoger la correctamente la división de la palabra.
- 1 punto por escoger un i que haga fallar la propiedad de bombeo.
- 1 punto por la desigualdad que muestra que la palabra bombeada no pertenece al lenguaje.
- 1 punto por identificar que b puede ser 0.

Evaluación y puntajes de la tarea

Cada item de cada pregunta se evaluará con un puntaje de 0, 1, 2, 3 o 4 puntos. Todas las preguntas tienen la misma ponderación en la nota final y cada item tiene la misma ponderación en cada pregunta.