Autómatos e Linguagens Formais

Autómatos Finitos

M. Lurdes Teixeira
Dep. Matemática
Univ. Minho

2º semestre de 2019/2020

Os conceitos principais neste capítulo são:

- autómato finito,
- linguagem reconhecida por um autómato finito.

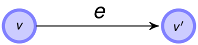
A relação entre linguagens regulares, estudadas no capítulo anterior, e linguagens reconhecidas por autómatos finitos é estabelecida pelo Teorema de Kleene, que é o resultado base de área de Teoria de Autómatos e Linguagens.

A prova do Teorema de Kleene é uma demonstração construtiva que conduz a dois algoritmos que permitem construir o autómato finito que reconhece a linguagem associada a uma certa expressão regular e vice-versa.

Definição

Um grafo orientado e etiquetado G (também designado por multi-grafo) é constituído por dois conjuntos: um conjunto V de vértices e um conjunto E de arestas (orientadas). Cada aresta é representada por um triplo ordenado (v, e, v') em que v e v' são vértices e e é uma etiqueta que pertence a um conjunto finito A. Um tal grafo representa-se por $G = (V, A, E)$.

EXEMPLO 1



Na figura está representada a aresta (v, e, v') cujo vértice inicial é v, o vértice final é v' e a etiqueta é e.

Definição

Duas arestas (v, e, w) e (v', e', w') de um grafo orientado e etiquetado dizem-se:

- consecutivas se $w = v'$ ou $w' = v$;

$$\begin{tikzcd} v \arrow[r, "e"] & v' \arrow[r, "e'"] & w' \end{tikzcd} \quad \text{ou} \quad \begin{tikzcd} v' \arrow[r, "e'"] & v \arrow[r, "e"] & w \end{tikzcd}$$

- coterminais se $v = v'$ e $w = w'$.

$$\begin{tikzcd} v \arrow[r, bend left, "e"] \arrow[r, bend right, "e'"'] & w \end{tikzcd}$$

Definições

Um caminho (orientado) é uma sequência de n arestas consecutivas ($n \in \mathbb{N}_0$), que se podem representar por:

$$v_1 \xrightarrow{e_1} v_2 \xrightarrow{e_2} v_3 \quad \cdots \quad v_n \xrightarrow{e_n} v_{n+1}$$

O vértice v_1 diz-se o vértice inicial ou origem do caminho e v_{n+1} diz-se o vértice final ou o fim do caminho. Alternativamente, diz-se que se trata de um caminho de v_1 para v_{n+1}.

No caso de $n = 0$, o caminho é representado apenas por um vértice, que é a origem e o fim do caminho.

A sequência

$$e_1 \cdots e_n$$

diz-se a etiqueta do caminho.

No caso de $v_1 = v_{n+1}$ o caminho diz-se fechado.

Definição

Um autómato (finito) é um quíntuplo $\mathcal{A} = (Q, A, \delta, i, F)$ onde

1. Q é um conjunto finito não vazio, chamado o conjunto de estados de $\mathcal{A}$;
2. A é um conjunto finito, chamado o alfabeto (de entrada) de $\mathcal{A}$;
3. δ é uma função de $Q \times A$ em $\mathcal{P}(Q)$, designada a função transição de $\mathcal{A}$. Cada triplo (p, a, q), em que $p, q \in Q$ e $a \in A$ são tais que $q \in \delta(p, a)$, diz-se uma transição de $\mathcal{A}$;
4. $i \in Q$ é dito o estado inicial de $\mathcal{A}$;
5. $F \subseteq Q$ é dito o conjunto de estados finais de $\mathcal{A}$.

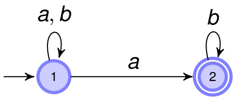
Um autómato finito representa-se por um grafo em que:

- os vértices são os estados;
- as arestas representam as transições;
- as etiquetas são letras do alfabeto;
- o estado inicial é assinalado através de uma seta;
- os estados finais são assinalados por círculos duplos.

EXEMPLO 2

Seja $\mathcal{A} = (\{1,2\}, \{a,b\}, \delta, 1, \{2\})$ onde δ é a função definida pela tabela seguinte:

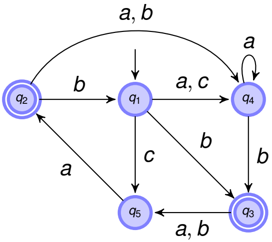
δ	1	2
a	$\{1,2\}$	$\emptyset$
b	$\{1\}$	$\{2\}$

$\mathcal{A} =$ 

Definição

Um caminho diz-se um caminho bem sucedido se a origem é o estado inicial e o fim é um estado final do autómato.

EXEMPLO 3

$\mathcal{A} =$ 

Exemplos de caminhos bem sucedidos:

$$q_1 \xrightarrow{c} q_4 \xrightarrow{b} q_3$$

$$q_1 \xrightarrow{c} q_4 \xrightarrow{a} q_4 \xrightarrow{b} q_3$$

$$q_1 \xrightarrow{c} q_4 \xrightarrow{a} q_4 \xrightarrow{b} q_3 \xrightarrow{b} q_5 \xrightarrow{a} q_2$$

Exemplos de caminhos que não são bem sucedidos:

$$q_1 \xrightarrow{c} q_5 \xrightarrow{a} q_2 \xrightarrow{b} q_1$$

$$q_1 \xrightarrow{c} q_5 \xrightarrow{a} q_2 \xrightarrow{b} q_4 \xrightarrow{b} q_3 \xrightarrow{a} q_5$$

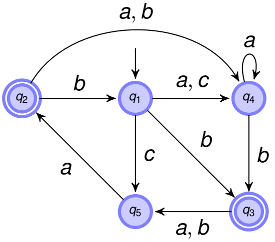
Definição

Dado um autómato $\mathcal{A}$, uma palavra sobre o alfabeto do autómato diz-se aceite ou reconhecida pelo autómato se é etiqueta de um caminho bem sucedido.

O conjunto de todas as palavras reconhecidas por um autómato finito $\mathcal{A}$ diz- se a linguagem reconhecida pelo autómato $\mathcal{A}$ e representa- se por $L(\mathcal{A})$.

EXEMPLO 3- continuação

$\mathcal{A} =$



As palavras cb, ca, ca^5b, ab^2a e cab^2a são palavras reconhecidas.

Qualquer palavra que termine na letra c ou que tenha prefixo b^3 não é reconhecida por este autómato.

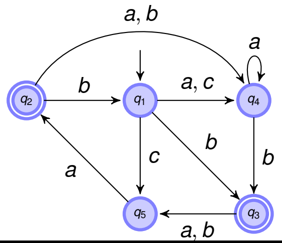
Dado um autómato $\mathcal{A} = (Q, A, \delta, i, F)$, a função $\delta : Q \times A \to \mathcal{P}(Q)$ pode ser estendida a uma função $\delta^* : Q \times A^* \to \mathcal{P}(Q)$ da seguinte forma:

1. $\delta^*(q, \varepsilon) = \{q\}$ qualquer que seja o estado q;
2. $\delta^*(q, aw) = \bigcup_{q' \in \delta(q,a)} \delta^*(q', w)$ quaisquer que sejam $a \in A$, $w \in A^*$ e $q \in Q$.

$$L(\mathcal{A}) = \left\{ w \in A^* \mid \delta^*(i, w) \cap F \neq \emptyset \right\}.$$

EXEMPLO 3 - continuação

$\mathcal{A} =$



$\delta^*(q_1, cb) = \{q_3\}$
$\delta^*(q_3, cb) = \emptyset$
$\delta^*(q_1, ab^2a) = \{q_2\}$
$\delta^*(q_1, cab^2a) = \{q_2, q_5\}$

Definição

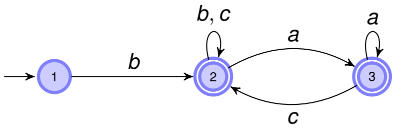
Dois autómatos $\mathcal{A}$ e $\mathcal{B}$ dizem-se equivalentes se $L(\mathcal{A}) = L(\mathcal{B})$.

Definição

Uma linguagem L diz-se reconhecível se existe um autómato $\mathcal{A}$ que reconhece L, ou seja, tal que $L = L(\mathcal{A})$.

EXEMPLO 4

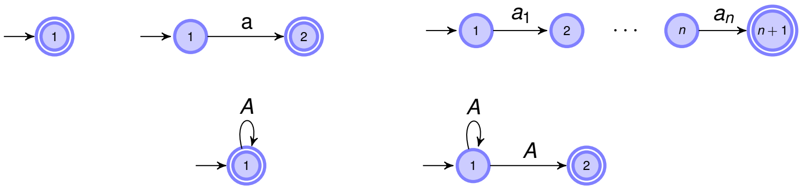
A linguagem $L = bA^* \setminus A^*abA^*$ sobre o alfabeto $A = \{a, b, c\}$ é reconhecível, porque é reconhecida pelo autómato seguinte:



Note-se que $L = \mathcal{L}\left(b(b+c)^*\left(aa^*c(b+c)^*\right)^*(\varepsilon + aa^*)\right)$ é uma linguagem regular.

EXEMPLOS 5

Outros exemplos de linguagens reconhecíveis sobre um alfabeto A são: $\{\varepsilon\}$, $\{a\}$ para qualquer $a \in A$, $\{u\}$ para qualquer $u = a_1 \cdots a_n \in A^+$, A^* e A^+. Os autómatos que as reconhecem são, respetivamente,



O resultado mais relevante da Teoria de Autómatos e Linguagens é o seguinte teorema.

Teorema de Kleene (1954)

Uma linguagem é regular se e só se é reconhecível.

Seja A um alfabeto qualquer.

- Recorde-se que $\mathcal{P}(A^*)$, o conjunto de todas as linguagens sobre A, é infinito não numerável.
- E qual o cardinal do conjunto $\mathcal{R}ec(A)$ de todas as linguagens, sobre um alfabeto A, que são reconhecíveis?

Lema

O conjunto $\mathcal{R}ec(A)$ é infinito numerável.

Prova

Fixemos um conjunto numerável $\mathcal{Q} = \{q_1, q_2, \ldots\}$. Cada linguagem reconhecível é reconhecida por algum autómato finito $\mathcal{A} = (Q, A, \delta, i, F)$ de alfabeto A e conjunto de estados $Q = \{q_1, \ldots, q_n\}$. Note-se que Q, A e F são finitos. Então, fixados Q, F e i, o número de maneiras de escolher as arestas de um autómato também é finito. Assim, o conjunto destes autómatos tem cardinal $\aleph_0$. Portanto o cardinal de $\mathcal{R}ec(A)$ também é igual a $\aleph_0$.

Corolário

Existe um conjunto infinito não numerável de linguagens não reconhecíveis.

O resultado seguinte apresenta uma condição necessária para que uma linguagem infinita seja reconhecível.

Lema da Iteração (ou da Bombagem)

Seja L uma linguagem reconhecível infinita, sobre um alfabeto A. Então existe uma constante $n \in \mathbb{N}$ tal que, para toda a palavra $u \in L$, se $|u| \geq n$, então existem palavras $x, y, z \in A^*$ tais que:

1. $|xy| \leq n$ e $y \neq \varepsilon$;
2. $u = xyz$;
3. $\forall\, k \in \mathbb{N}_0,\ xy^k z \in L$.

EXEMPLO 6

$$L = \{a^m b^m \mid m \in \mathbb{N}_0\} \subseteq \{a, b\}^*$$

Sejam $n \in \mathbb{N}$ e $u = a^n b^n \in L$. Então $|u| > n$.

Sejam $x, y, z \in A^*$ tais que:

- $u = xyz$,
- $|xy| \leq n$
- $y \neq \varepsilon$.

a^n		b^n
x	y	z

Assim, $y = a^\ell$, com $1 \leq \ell \leq n$, e

$$xy^kz = a^{n-\ell}a^{k\ell}b^n = \begin{cases} a^{n-\ell}b^n & \text{se } k = 0 \\ a^n b^n & \text{se } k = 1 \\ a^n a^{(k-1)\ell} b^n & \text{se } k > 1 \end{cases}$$

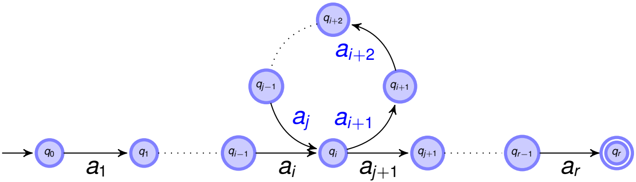
Logo, $xy^kz \notin L$ se $k \neq 1$.

PROVA do Lema da Iteração

Seja $\mathcal{A} = (Q, A, \delta, i, F)$ um autómato que reconhece L e seja $n = |Q|$. Então, para cada $u = a_1 a_2 \cdots a_r \in L$, com $r \geq n$, existe um caminho bem sucedido de $\mathcal{A}$, de etiqueta u:

$$i = q_0 \xrightarrow{a_1} q_1 \xrightarrow{a_2} q_2 \cdots q_{r-1} \xrightarrow{a_r} q_r$$

Dado que $r \geq n$, existem inteiros i e j, com $0 \leq i < j \leq n$, tais que $q_i = q_j$, e a palavra $a_{i+1} \cdots a_j$ é a etiqueta de um circuito com origem q_i:



Sejam

$$x = \begin{cases} \varepsilon & \text{se } i = 0 \\ a_1 \cdots a_i & \text{se } i > 0 \end{cases}, \quad y = a_{i+1} \cdots a_j, \quad z = \begin{cases} \varepsilon & \text{se } j = r \\ a_{j+1} \cdots a_r & \text{se } j < r \end{cases}.$$

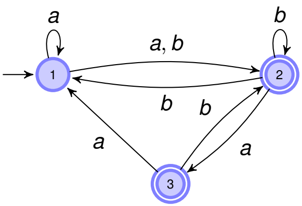
Então $|xy| = j \leq n$, $y \neq \varepsilon$, $u = xyz$ e, para todo o $k \in \mathbb{N}_0$, $xy^k z \in L$ pois $xy^k z$ é a etiqueta de um caminho bem sucedido de $\mathcal{A}$.

Definição

Um autómato $\mathcal{A} = (Q, A, \delta, i, F)$ diz-se completo se para cada par $(q, a) \in Q \times A$ existe uma transição (q, a, q') para algum $q' \in Q$, ou seja, $\delta(q, a) \neq \emptyset$.

EXEMPLO 7

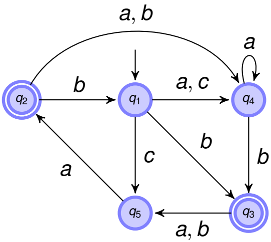
Sendo $A = \{a, b\}$, o seguinte autómato é completo



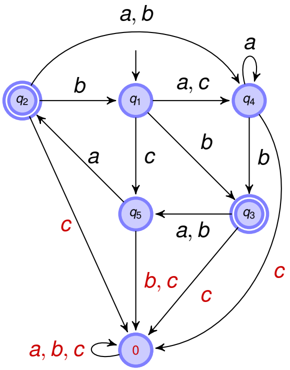
EXEMPLO 8

Seja $A = \{a, b, c\}$.

$\mathcal{A} =$



$\mathcal{B} =$



O autómato $\mathcal{A}$ não é completo, mas o autómato $\mathcal{B}$ é completo.

Notar que $\mathcal{A}$ e $\mathcal{B}$ são equivalentes.

Teorema

Todo o autómato é equivalente a um autómato completo.

PROVA

Seja que $\mathcal{A} = (Q, A, \delta, i, F)$ um autómato incompleto.
Define-se um novo autómato $\mathcal{A}' = (Q \cup \{p\}, A, \delta', i, F)$ em que:

- p é um novo estado;
- δ' é a extensão de δ a $Q \cup \{p\}$ tal que:
  - se $a \in A$ e $\delta(q, a) = \emptyset$, então $\delta'(q, a) = \{p\}$;
  - $\delta(p, a) = \{p\}$ para qualquer $a \in A$.

Logo, $\mathcal{A}'$ é completo.

Notar que qualquer caminho que passe no vértice p termina no vértice p e não é um caminho bem sucedido, porque $p \notin F$. Logo, os caminhos bem sucedidos de $\mathcal{A}$ e de $\mathcal{A}'$ coincidem, ou seja, $L(\mathcal{A}) = L(\mathcal{A}')$. Então $\mathcal{A}'$ é um autómato completo equivalente a $\mathcal{A}$.

Definição

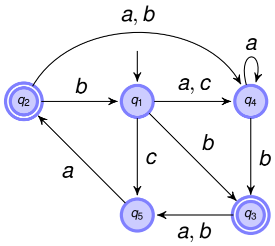
Um autómato $\mathcal{A} = (Q, A, \delta, i, F)$ diz-se:

- acessível se todos os estados são acessíveis, *i.e.*, se para cada estado $q \in Q$ existe um caminho de i para q.
- co-acessível se todos os estados são co-acessíveis, *i.e.*, se para cada $q \in Q$ existe um caminho de q para um estado final.

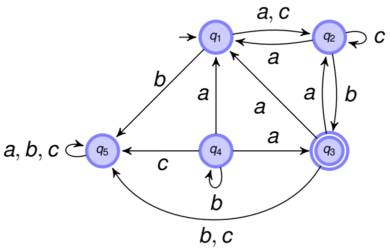
EXEMPLO 9

O autómato $\mathcal{A}$ é acessível e co-acessível.

$\mathcal{A} =$



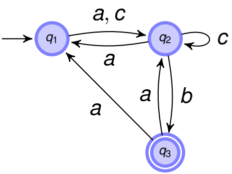
EXEMPLO 10



Este autómato

- não é acessível, porque o estado q_4 não é acessível,
- nem co-acessível, porque o estado q_5 não é co-acessível.

O autómato seguinte é acessível e co-acessível e é equivalente ao anterior.



Um vértice faz parte de um caminho bem sucedido se e só se é um vértice acessível e co-acessível.

Teorema

Todo o autómato $\mathcal{A}$, tal que $L(\mathcal{A}) \neq \emptyset$, é equivalente a um autómato acessível e co-acessível.

PROVA

Sejam $\mathcal{A} = (Q, A, \delta, i, F)$ e P o subconjunto de Q constituído por todos os vértices acessíveis e co-acessíveis de $\mathcal{A}$. Seja $\mathcal{A}' = (P, A, \delta', i, F \cap P)$ onde, para qualquer $(q, a) \in P \times A$, $\delta'(q, a) = \delta(q, a) \cap P$. Então,

1. para qualquer $p \in P$, existe um caminho C bem sucedido em $\mathcal{A}$ que passa por p, sendo que todos os vértices de C são elementos de P e, em particular, o vértice final pertence a $F \cap P$;
2. qualquer aresta de um caminho bem sucedido de $\mathcal{A}$ é da forma (q, a, q') em que $q' \in \delta(q, a)$ e $q, q' \in P$, pelo que $q' \in \delta'(q, a)$;
3. os caminhos bem sucedidos de $\mathcal{A}$ e de $\mathcal{A}'$ coincidem, *i.e.*,

$$\begin{aligned} L(\mathcal{A}') &= \{u \in A^* \mid \delta'^*(i, u) \cap (F \cap P) \neq \emptyset\} \\ &= \{u \in A^* \mid \delta^*(i, u) \cap F \neq \emptyset\} \\ &= L(\mathcal{A}). \end{aligned}$$

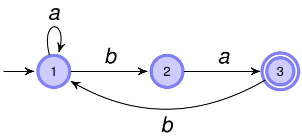
Então $\mathcal{A}'$ é acessível, co-acessível e equivalente a $\mathcal{A}$.

Definição

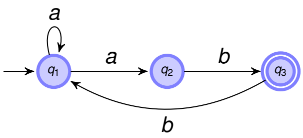
Um autómato $A = (Q, A, \delta, i, F)$ diz-se determinista se, para cada estado $q \in Q$ e cada letra $a \in A$, $\sharp\delta(q, a) \leq 1$, *i.e.*, existe no máximo uma transição de origem q e etiqueta a: $q \xrightarrow{a} q'$.

EXEMPLO 11

Autómato determinista



Autómato não determinista



Teorema

Todo o autómato $\mathcal{A}$ é equivalente a um autómato determinista $D(\mathcal{A})$.

PROVA

Seja $\mathcal{A} = (Q, A, \delta, i, F)$. Define-se $D(\mathcal{A}) = (Q', A, \delta', I, F')$ o autómato tal que:

- $Q' = \mathcal{P}(Q)$;
- $\delta' : \begin{array}{ccc} Q' \times A & \to & \mathcal{P}(Q') \\ (X, a) & \mapsto & \{\bigcup_{q \in X} \delta(q, a)\} \end{array}$;
- $I = \{i\}$ e $F' = \{X \in Q' \mid X \cap F \neq \emptyset\}$.

O autómato $D(\mathcal{A})$ é determinista (e completo) por construção.

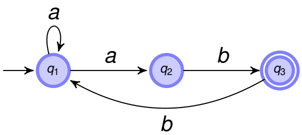
$$\begin{aligned} L(\mathcal{A}') &= \{u \in A^* \mid (\delta'^*(I, u) \cap F' \neq \emptyset\} \\ &= \{u \in A^* \mid \delta^*(i, u) = X \wedge (X \cap F) \neq \emptyset\} \\ &= \{u \in A^* \mid \delta^*(i, u) \cap F \neq \emptyset\} \\ &= L(\mathcal{A}) \end{aligned}$$

Logo, $D(\mathcal{A})$ é equivalente a $\mathcal{A}$.

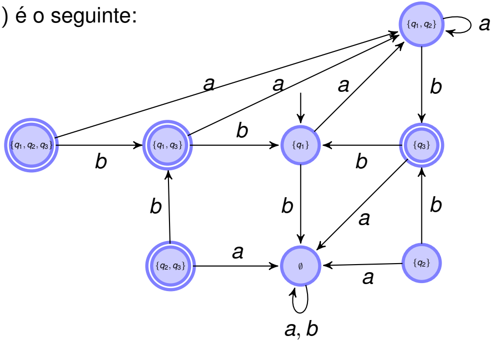
EXEMPLO 12

Autómato não determinista

$\mathcal{A} =$



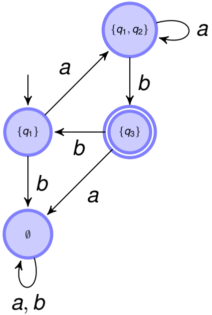
O autómato $D(\mathcal{A})$ é o seguinte:



Existem estados não acessíveis ou não co-acessíveis?

EXEMPLO 12 - continuação

Removendo os vértices não acessíveis de $D(\mathcal{A})$ obtém-se o seguinte autómato determinista, acessível e completo:

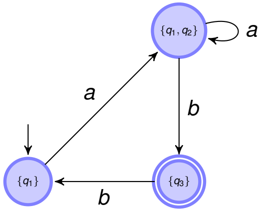


Teorema

Todo o autómato $\mathcal{A}$ é equivalente a um autómato determinista, completo e acessível.

EXEMPLO 12 - continuação

Removendo os vértices não acessíveis e não co-acessíveis de $D(\mathcal{A})$ obtém-se o seguinte autómato determinista, acessível e e co-acessível, mas não completo.



Teorema

Todo o autómato $\mathcal{A}$ é equivalente a um autómato determinista, acessível e co-acessível.

Definição

Um autómato com transições vazias (ou transições-ε) é um autómato que se representa por $\mathcal{A} = (Q, A, \delta, i, F)$ em que a função transição δ tem domínio $Q \times (A \cup \{\varepsilon\})$. Tais autómatos designam-se por autómatos assíncronos.

Uma transição vazia é uma transição etiquetada pela palavra vazia: (q, ε, q') com $q, q' \in Q$.

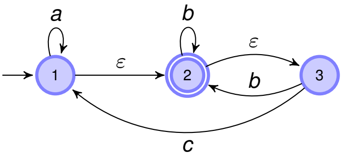
A designação 'assíncrono' resulta por oposição aos autómatos sem transições vazias que são também chamados síncronos, nos quais, se unidade de tempo corresponde a uma qualquer transição, o tempo de leitura de uma palavra é proporcional ao comprimento da palavra.

Se $\mathcal{A}$ é um autómato assíncrono, para cada estado q, denotamos por $\text{fecho}_\varepsilon(q)$ o conjunto formado pelo próprio q e pelos estados atingíveis a partir de q por um caminho de etiqueta ε*i.e.*

$$\text{fecho}_\varepsilon(q) = \{q\} \cup \delta^*(q, \varepsilon).$$

EXEMPLO 13

O grafo seguinte representa um autómato assíncrono.



$$\text{fecho}_\varepsilon(1) = \{1, 2, 3\}, \quad \text{fecho}_\varepsilon(2) = \{2, 3\} \quad \text{e} \quad \text{fecho}_\varepsilon(3) = \{3\}.$$

Teorema

Todo o autómato assíncrono é equivalente a um autómato síncrono.

PROVA

Sendo $\mathcal{A} = (Q, A, \delta, i, F)$ um autómato assíncrono qualquer, seja $\mathcal{A}' = (Q, A, \delta', i, F')$ o autómato tal que:

- $\delta' : Q \times A \to \mathcal{P}(Q)$ é a função definida, para cada $(q, a) \in Q \times A$, por

$$\delta'(q, a) = \bigcup_{p \,\in\, \text{fecho}_{\varepsilon}(q)} \delta(p, a).$$

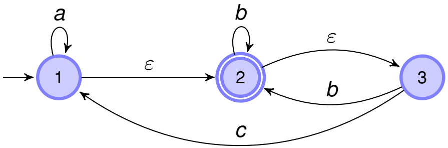
- $F' = \{q \in Q \mid \text{fecho}_{\varepsilon}(q) \cap F \neq \emptyset\}$.

Como se pode verificar, $\mathcal{A}'$ é um autómato síncrono e é equivalente a $\mathcal{A}$ porque

$$\begin{aligned} L(\mathcal{A}') &= \{u \in A^* \mid \delta'^*(i, u) \cap F' \neq \emptyset\} \\ &= \{u \in A^* \mid \delta^*(i, u) \cap F \neq \emptyset\} \\ &= L(\mathcal{A}) \end{aligned}$$

Exemplo 13- continuação

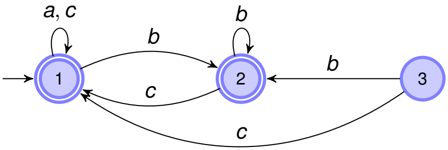
Consideremos o autómato com transições vazias $\mathcal{A}$ do exemplo anterior.



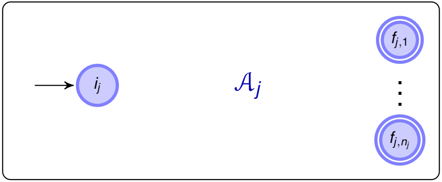
Como $\text{fecho}_\varepsilon(1) = \{1,2,3\}$, $\text{fecho}_\varepsilon(2) = \{2,3\}$ e $\text{fecho}_\varepsilon(3) = \{3\}$, então

$$\begin{array}{llllll}
\delta'(1,a) & = \delta(1,a) = \{1\} & \delta'(1,b) & = \cup_{i\in\{2,3\}}\delta(i,b) = \{2\} & \delta'(1,c) & = \delta(3,c) = \{1\} \\
\delta'(2,a) & = \emptyset & \delta'(2,b) & = \cup_{i\in\{2,3\}}\delta(i,b) = \{2\} & \delta'(2,c) & = \delta(3,c) = \{1\} \\
\delta'(3,a) & = \emptyset & \delta'(3,b) & = \delta'(3,b) = \{2\} & \delta'(3,c) & = \delta(3,c) = \{1\}
\end{array}$$

pelo que o autómato síncrono $\mathcal{A}'$ descrito na prova do teorema anterior é



Ao longo desta secção considera-se que cada linguagem L_j reconhecível é reconhecida por um autómato (síncrono ou assíncrono) do tipo $\mathcal{A}_j = (Q_j, A, \delta_j, i_j, F_j)$ em que $F_j = \{f_{j,1}, \ldots, f_{j,n_j}\}$, com $j \in \{1, 2\}$, que é representado como na figura abaixo.

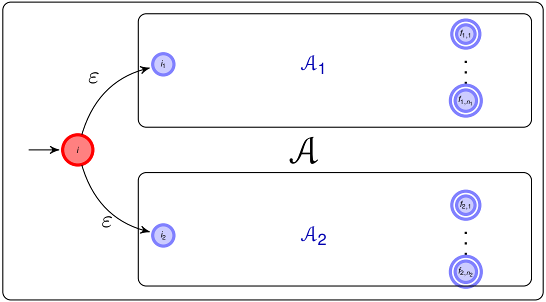


Lema A

Se L_1 e L_2 são linguagens reconhecíveis, então $L_1 \cup L_2$ é uma linguagem reconhecível.

PROVA

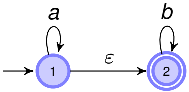
O autómato assíncrono $\mathcal{A} = (Q_1 \uplus Q_2 \uplus \{i\}, A, \delta, i, F_1 \cup F_2)$



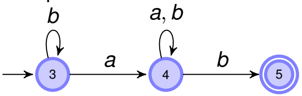
reconhece $L_1 \cup L_2$. Logo $L_1 \cup L_2$ é uma linguagem reconhecível.

EXEMPLO 14

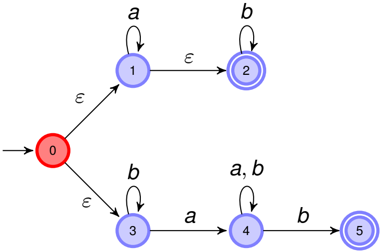
As linguagens $a^* b^*$ e $b^* a(a+b)^* b$ são reconhecidas pelos autómatos



e



Então, a linguagem $a^* b^* + b^* a(a+b)^* b$ é reconhecida pelo autómato seguinte (obtido a partir dos autómatos acima)

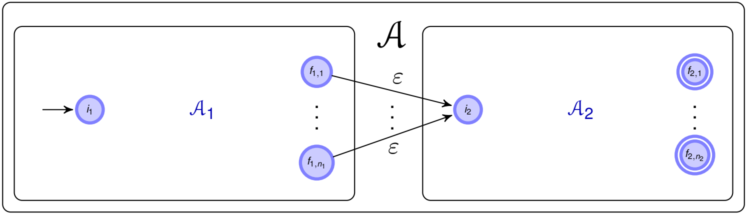


Lema B

Se L_1 e L_2 são linguagens reconhecíveis, então $L_1 L_2$ é uma linguagem reconhecível.

PROVA

O seguinte autómato $\mathcal{A} = (Q_1 \uplus Q_2, A, \delta, i_1, F_2)$



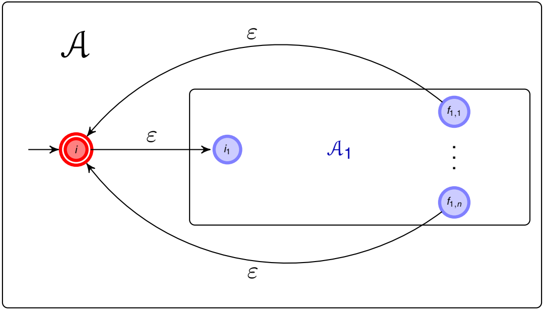
reconhece $L_1 L_2$. Portanto $L_1 L_2$ é uma linguagem reconhecível.

Lema C

Se L_1 é uma linguagem reconhecível, então L_1^* é uma linguagem reconhecível.

PROVA

O seguinte autómato $\mathcal{A} = (Q_1 \uplus \{i\}, A, \delta, i, \{i\})$



reconhece L_1^*, donde esta linguagem é reconhecível,

Proposição

Se $L \subseteq A^*$ é uma linguagem regular, então L é reconhecível.

PROVA

Por indução estrutural sobre $\mathcal{R}eg(A)$.

1. Se $L = \emptyset$ ou $L = \{\varepsilon\}$, então L é reconhecível.
2. Se $L = \{a\}$ em que $a \in A$, então L é reconhecível.
3. Seja $L = L_1 \cup L_2$ (resp., $L = L_1 L_2$) e suponhamos, por hipótese de indução, que L_1 e L_2 são linguagens reconhecíveis. Então, pelo Lema A (resp., Lema B) , L é reconhecível.
4. Seja $L = K^*$ e suponhamos, por hipótese de indução, que K é uma linguagem reconhecível. Então, pelo Lema C, L é reconhecível.

De (1)-(4) resulta, pelo Princípio de Indução Estrutural sobre $\mathcal{R}eg(A)$, que L é uma linguagem reconhecível

Para completar a demonstração do Teorema de Kleene, falta provar que toda a linguagem reconhecida por um autómato finito pode ser representada por uma expressão regular.

Definição

Seja $\mathcal{A} = (Q, A, \delta, i, F)$ um autómato finito. Suponhamos que os estados são representado por números de modo que $Q = \{1, \ldots, n\}$. O sistema de equações lineares à direita associado a $\mathcal{A}$ é o sistema

$$\begin{cases} X_1 = r_{11}X_1 + r_{12}X_2 + \cdots + r_{1n}X_n + s_1 \\ \quad\vdots \\ X_n = r_{n1}X_1 + r_{n2}X_2 + \cdots + r_{nn}X_n + s_n \end{cases}$$

onde, para cada $j, k \in Q$,

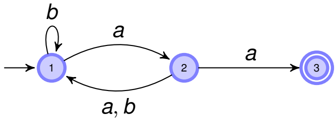
- r_{jk} é uma expressão regular que representa a linguagem

$$R_{jk} = \{a \in A \mid \text{existe uma transição } (j, a, k)\};$$

- $s_j = \varepsilon$ se $j \in F$, e $s_j = \emptyset$ caso contrário.

EXEMPLO 15

Considere-se o autómato $\mathcal{A}$.



O sistema associado a $\mathcal{A}$ é

$$\begin{cases} X_1 = bX_1 + aX_2 + \emptyset X_3 + \emptyset \\ X_2 = (a+b)X_1 + \emptyset X_2 + aX_3 + \emptyset \\ X_3 = \emptyset X_1 + \emptyset X_2 + \emptyset X_3 + \varepsilon \end{cases}$$

que pode ser escrito simplesmente na forma

$$\begin{cases} X_1 = bX_1 + aX_2 \\ X_2 = (a+b)X_1 + aX_3 \\ X_3 = \varepsilon \end{cases}$$

Lema

Seja $(t_1, t_2, \ldots, t_n) \in ER(A)^n$ a solução mínima do sistema associado ao autómato finito $\mathcal{A} = (Q, A, \delta, i, F)$. Então, para cada estado $j \in Q$,

$$\mathcal{L}(t_j) = \{u \in A^* \mid u \text{ é etiqueta de um caminho de } j \text{ para } q \in F\}.$$

Em particular,

$$L(\mathcal{A}) = \mathcal{L}(t_i),$$

ou seja, a linguagem reconhecida pelo autómato $\mathcal{A}$ é representada pela expressão regular t_i, onde i é o estado inicial de $\mathcal{A}$.

Notar que num sistema associado a um autómato síncrono se tem que $\varepsilon \notin R_{j,k}$ para quaisquer estados j, k, pelo que a solução do sistema é única.

Proposição

Se L é uma linguagem reconhecível, então L é regular.

Estamos então em condições de estabelecer o resultado fundamental desta área.

Teorema [Kleene'1954]

Uma linguagem é regular se e só se é reconhecível.

EXEMPLO 15 - continuação

Resolvendo o sistema obtém-se, sucessivamente,

$$\begin{cases} X_1 = bX_1 + aX_2 \\ X_2 = (a+b)X_1 + aX_3 \\ X_3 = \varepsilon \end{cases} \Leftrightarrow \begin{cases} X_1 = bX_1 + aX_2 \\ X_2 = (a+b)X_1 + a\varepsilon \\ X_3 = \varepsilon \end{cases}$$

$$\Leftrightarrow \begin{cases} X_1 = bX_1 + a((a+b)X_1 + a) \\ X_2 = (a+b)X_1 + a \\ X_3 = \varepsilon \end{cases} \Leftrightarrow \begin{cases} X_1 = (b + a^2 + ab)X_1 + a^2 \\ X_2 = (a+b)X_1 + a \\ X_3 = \varepsilon \end{cases}$$

$$\Leftrightarrow \begin{cases} X_1 = (b + a^2 + ab)^* a^2 \\ X_2 = (a+b)X_1 + a \\ X_3 = \varepsilon \end{cases} \Leftrightarrow \begin{cases} X_1 = (b + a^2 + ab)^* a^2 \\ X_2 = (a+b)(b+a^2+ab)^* a^2 + a \\ X_3 = \varepsilon \end{cases}$$

A solução mínima do sistema é portanto

$$\Big((b + a^2 + ab)^* a^2, \quad (a+b)(b+a^2+ab)^* a^2 + a, \quad \varepsilon\Big).$$

Dado que o estado inicial de $\mathcal{A}$ é 1, conclui-se que $L(\mathcal{A})$ é representada pela expressão regular $(b + a^2 + ab)^* a^2$.

Definição

Um autómato determinista, completo e que reconhece uma linguagem $L \subseteq A^*$ e que tem um número mínimo de estados, de entre os autómatos deterministas e completos que reconhecem L, diz-se um autómato minimal de L.

Seja L uma linguagem reconhecível. Notar que:

- Existe um autómato minimal que reconhece L. Mais, prova-se que que tal autómato é único (a menos da forma de identificar os esta- dos) e, usualmente, representa-se por $\mathcal{A}(L)$.
- $\mathcal{A}(L)$ é determinista, completo e acessível (abreviadamente, DCA).

Se $\mathcal{A} = (Q, A, \delta, i, F)$ é um autómato finito DCA, então a função transição δ é uma função em que, para qualquer par $(q, a) \in Q \times A$, a imagem $\delta(q, a)$ é um conjunto singular, digamos $\{q'\}$ com $q' \in Q$. Podemos então, neste caso, considerar que a função de transição é uma função total sobrejetiva $\delta : Q \times A \to Q$ e, com a notação anterior, diríamos que $\delta(q, a) = q'$.

Proposição

Seja $L \subseteq A^*$ uma linguagem. Então,

1. L é reconhecível se e só se o conjunto

$$Q_L = \{u^{-1}L \mid u \in A^*\}$$

   é finito;

2. $\mathcal{A}(L) = (Q_L, A, \delta, L, F_L)$ onde
   - $F_L = \{u^{-1}L \mid u \in L\} = \{K \in Q_L \mid \varepsilon \in K\}$;
   - a função de transição δ é definida, para cada $a \in A$, por

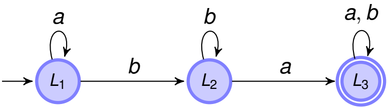
$$\delta(u^{-1}L, a) = a^{-1}(u^{-1}L) = (ua)^{-1}L.$$

EXEMPLO 16

Seja $L = A^* baA^*$ sobre o alfabeto $A = \{a, b\}$.

$$
\begin{aligned}
\varepsilon^{-1} L &= L = L_1 \\[1ex]
a^{-1} L_1 &= a^{-1}(A^*)baA^* \cup a^{-1}(baA^*) = L_1 \cup \emptyset = L_1 \\
b^{-1} L_1 &= b^{-1}(A^*)baA^* \cup b^{-1}(baA^*) = L_1 \cup aA^* = L_2 \\[1ex]
a^{-1} L_2 &= a^{-1} L_1 \cup a^{-1}(aA^*) = L_1 \cup A^* = A^* = L_3 \\
b^{-1} L_2 &= b^{-1} L_1 \cup b^{-1}(aA^*) = L_2 \cup \emptyset = L_2 \\[1ex]
a^{-1} L_3 &= L_3 \\
b^{-1} L_3 &= L_3.
\end{aligned}
$$

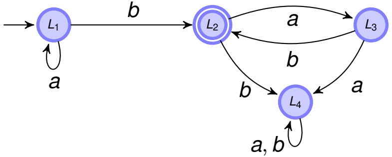
Assim, $Q_L = \{L_1, L_2, L_3\}$ e $\mathcal{A}(L) = (Q_L, A, \delta, L_1, \{L_3\})$ é o autómato seguinte.



EXEMPLO 17 Seja $L = a^*b(ab)^*$ sobre o alfabeto $A = \{a, b\}$.

$$
\begin{aligned}
\varepsilon^{-1}L &= L = L_1 \\
a^{-1}L_1 &= a^{-1}(a^*)b(ab)^* \cup a^{-1}(b(ab)^*) = L_1 \cup \emptyset = L_1 \\
b^{-1}L_1 &= b^{-1}(a^*)b(ab)^* \cup b^{-1}(b(ab)^*) = \emptyset \cup (ab)^* = (ab)^* = L_2 \\
a^{-1}L_2 &= a^{-1}(ab)(ab)^* = b(ab)^* = L_3 \\
b^{-1}L_2 &= b^{-1}(ab)(ab)^* = \emptyset = L_4 \\
a^{-1}L_3 &= L_4 \\
b^{-1}L_3 &= L_2 \\
a^{-1}L_4 &= L_4 \\
b^{-1}L_4 &= L_4.
\end{aligned}
$$

Assim, $Q_L = \{L_1, L_2, L_3, L_4\}$ e $\mathcal{A}(L) = (Q_L, A, \delta, L_1, \{L_2\})$ é o autómato:



Definição

Seja $\mathcal{A} = (Q, A, \delta, i, F)$ um autómato DCA. Define-se a relação $\sim$ sobre Q escrevendo $q \sim q'$ se a partir dos estados q e q' o autómato tem o mesmo "comportamento", isto é,

$$q \sim q' \quad \text{sse} \quad \forall u \in A^*, \delta^*(q, u) \in F \Leftrightarrow \delta^*(q', u) \in F.$$

Lema

Seja $\mathcal{A} = (Q, A, \delta, i, F)$ um autómato DCA. Então, a relação $\sim$ é uma equivalência sobre Q.

Se dois estados q e q' de um autómato DCA são tais que $q \sim q'$, então dizem-se estados equivalentes.

Lema

Seja $\mathcal{A} = (Q, A, \delta, i, F)$ um autómato DCA. Então, se $\overline{q}$ representa a classe-$\sim$ equivalência de $q \in Q$, e

$$\overline{Q} = Q/_\sim = \{\overline{q} : q \in Q\},$$

a correspondência

$$\begin{array}{rrcl} \overline{\delta}: & \overline{Q} \times A & \to & \overline{Q} \\ & (\overline{q}, a) & \mapsto & \overline{\delta(q, a)} \end{array}$$

é uma função.

PROVA

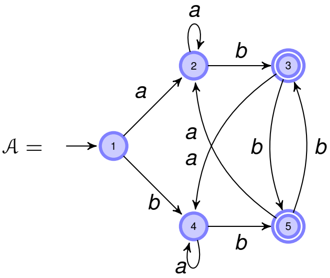
Para quaisquer $q_1, q_2 \in Q$ e $a \in A$,

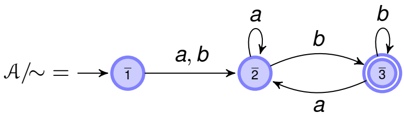
$$\begin{array}{rcl} \overline{q_1} = \overline{q_2} & \Longrightarrow & q_1 \sim q_2 \\ & \Longrightarrow & \forall u \in A^*,\ \delta^*(q_1, au) \in F \Leftrightarrow \delta^*(q_2, au) \in F \\ & \Longrightarrow & \forall u \in A^*,\ \delta^*(\delta(q_1, a), u) \in F \Leftrightarrow \delta^*(\delta(q_2, a), u) \in F \\ & \Longrightarrow & \delta(q_1, a) \sim \delta(q_2, a) \\ & \Longrightarrow & \overline{\delta(q_1, a)} = \overline{\delta(q_2, a)}. \end{array}$$

Definição

Seja $\mathcal{A} = (Q, A, \delta, i, F)$ um autómato DCA. Define-se o autómato quociente de $\mathcal{A}$ por $\sim$ como sendo o autómato $\mathcal{A}/_{\sim} = (\overline{Q}, A, \overline{\delta}, \overline{i}, \overline{F})$ onde $\overline{F} = \{\overline{f} : f \in F\}$ é o conjunto das classes-$\sim$ de equivalência dos elementos de F.

EXEMPLO 18





$\overline{Q} = Q/_{\sim} = \{\{1\}, \{2,4\}, \{3,5\}\} = \{\overline{1}, \overline{2}, \overline{3}\}.$

Dado um autómato DCA, como calcular a equivalência $\sim$?

Definição

Seja $\mathcal{A} = (Q, A, \delta, i, F)$ um autómato DCA. Para cada $k \in \mathbb{N}_0$, define-se a relação $\sim_k$ sobre Q, por: para quaisquer $q, q' \in Q$,

$$q \sim_k q' \quad \text{sse} \quad \forall u \in A^*, |u| \leq k \Rightarrow \left(\delta^*(q, u) \in F \Leftrightarrow \delta^*(q', u) \in F\right).$$

Notar que:

1. $\sim_{k+1} \subseteq \ \sim_k$ para todo o $k \in \mathbb{N}_0$;
2. $q \sim q'$ se e só se $q \sim_k q'$ para todo o $k \in \mathbb{N}_0$ (*i.e.*, $\sim = \bigcap_{i \geq 0} \sim_i$);
3. $q \sim_0 q'$ se e só se $q \in F \Leftrightarrow q' \in F$ (*i.e.*, $Q/_{\sim_0} = \{F, Q \setminus F\}$);
4. as relações $\sim_k$ podem ser definidas recursivamente da seguinte forma:
   1. $q \sim_0 q' \quad \Leftrightarrow \quad q, q' \in F \ \vee \ q, q' \in Q \setminus F$;
   2. para cada $k \in \mathbb{N}_0$,
   $$q \sim_{k+1} q' \quad \Leftrightarrow \quad \forall a \in A \left(q \sim_k q' \ \wedge \ \delta(q, a) \sim_k \delta(q', a)\right).$$

Lema

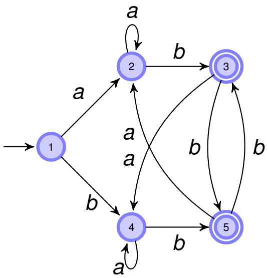
Seja $\mathcal{A} = (Q, A, \delta, i, F)$ um autómato DCA e seja n o número de estados de $\mathcal{A}$. Então,

1. $\sim_{r+1} = \sim_r$ para algum $r \in \{0, 1, \ldots, n-2\}$.
2. Se $k \in \mathbb{N}_0$ é tal que $\sim_{k+1} = \sim_k$, então $\sim = \sim_k$.

PROVA

1. Como, para todo o $k \in \mathbb{N}_0$, $\sim_{k+1} \subseteq \sim_k$, então $[q]_{\sim_{k+1}} \subseteq [q]_{\sim_k} \subseteq Q$ para todo o $q \in Q$. Dado que Q é finito, então existe $r \geq 0$ tal que $[q]_{\sim_{r+1}} = [q]_{\sim_r}$ para todo o $q \in Q$ e, como $\sharp Q = n$, então $r \leq n-2$.
2. Notar que se $k \in \mathbb{N}_0$ é tal que $\sim_{k+1} = \sim_k$, então $\sim_{k+\ell} = \sim_k$ para qualquer $\ell \geq 0$. Então,
$$\cdots = \sim_{k+\ell} = \cdots = \sim_{k+1} = \sim_k \subseteq \sim_{k-1} \subseteq \cdots \subseteq \sim_0$$
pelo que $\sim = \bigcap_{i \geq 0} \sim_i = \sim_k$.

EXEMPLO 18 - continuação

$\mathcal{A} =$ 

Usando a definição recursiva das relações $\sim_k$, obtém-se:

- $Q/_{\sim_0} = \{\{1,2,4\},\{3,5\}\}$.
- Notando que $\sim_1 \subseteq \sim_0$, para calcular $\sim_1$ basta verificar em cada classe-$\sim_0$ quais os elementos que são equivalentes módulo $\sim_1$.

Assim,

$$1 \not\sim_1 2 \quad \text{pois } \delta(1,b) = 4,\ \delta(2,b) = 3 \text{ e } 4 \not\sim_0 3;$$

$$2 \sim_1 4 \quad \text{pois } 2 \sim_0 4,\ \delta(2,a) = 2 \sim_0 4 = \delta(4,a) \text{ e } \delta(2,b) = 3 \sim_0 5 = \delta(4,b);$$

$$3 \sim_1 5 \quad \text{pois } 3 \sim_0 5,\ \delta(3,a) = 4 \sim_0 2 = \delta(5,a) \text{ e } \delta(3,b) = 5 \sim_0 3 = \delta(5,b).$$

Consequentemente, $\sim_1 \neq \sim_0$ e $Q/_{\sim_1} = \{\{1\},\{2,4\},\{3,5\}\}$.

EXEMPLO 18 - continuação

• Notando que $\sim_2 \subseteq \sim_1$, para calcular $\sim_2$ basta verificar em cada classe-$\sim_1$ quais os elementos que são equivalentes módulo $\sim_2$.

$$\begin{array}{ll} 2 \sim_2 4 & \text{pois } 2 \sim_1 4,\ \delta(2,a) = 2 \sim_1 4 = \delta(4,a) \ \text{ e } \ \delta(2,b) = 3 \sim_1 5 = \delta(4,b); \\ 3 \sim_2 5 & \text{pois } 3 \sim_1 5,\ \delta(3,a) = 4 \sim_1 2 = \delta(5,a) \ \text{ e } \ \delta(3,b) = 5 \sim_1 3 = \delta(5,b). \end{array}$$

Tem-se então $\sim_2 = \sim_1$, donde $\sim = \sim_1$ e, pelo lema anterior,

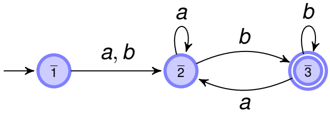
$$\overline{Q} = Q/_\sim = \{\{1\}, \{2,4\}, \{3,5\}\}.$$

Quanto à função $\overline{\delta}$, tem-se que:

$$\begin{array}{lll} \overline{\delta}(\overline{1},a) = \overline{\delta(1,a)} = \overline{2} & \text{e} & \overline{\delta}(\overline{1},b) = \overline{\delta(1,b)} = \overline{4} \\ \overline{\delta}(\overline{2},a) = \overline{\delta(2,a)} = \overline{2} & \text{e} & \overline{\delta}(\overline{2},b) = \overline{\delta(2,b)} = \overline{3} \\ \overline{\delta}(\overline{3},a) = \overline{\delta(3,a)} = \overline{4} & \text{e} & \overline{\delta}(\overline{3},b) = \overline{\delta(3,b)} = \overline{5} \end{array}$$

Logo, o autómato quociente $\mathcal{A}/_\sim$ é o seguinte:

$\mathcal{A}/\sim\ =$



Proposição

Se $\mathcal{A} = (Q, A, \delta, i, F)$ é um autómato DCA, então o autómato quociente $\mathcal{A}/_\sim$ é um autómato DCA equivalente a $\mathcal{A}$ e tem o menor número de estados entre os autómatos DCA equivalentes a $\mathcal{A}$.
Assim, $\mathcal{A}/_\sim$ é (a menos dos nomes dos estados) o autómato minimal da linguagem $L(\mathcal{A})$.