

Formális nyelvek és a fordítóprogramok alapjai

3. előadás

Előadó: Nagy Sára, mesteroktató
Algoritmusok és Alkalmazásaik Tanszék

Emlékeztető

Definíció: Grammatikának a következő négyest nevezzük:

$G=(N,T,P,S)$

- N a nemterminális ábácé,
- T a terminálisok ábécéje,
- P az átírási szabályok véges halmaza,
- S a kezdőszimbólum.

Emlékeztető

Nyelvtan által generált nyelv:

$$L(G) := \{ u \in T^* \mid S \xRightarrow[G]{*} u \}$$

Ekvivalens nyelvtanok:

A G_1 es G_2 nyelvtanok ekvivalensek, ha $L(G_1) = L(G_2)$.

Nyelvtani transzformáció

A nyelvtani transzformáció olyan eljárás, amely egy G grammatikából egy másik G' grammatikát készít.

Ekvivalens transzformációról beszélünk, ha minden G grammatikára és az ő G' transzformáltjára igaz, hogy $L(G)=L(G')$.

ϵ -mentesítés

Tétel:

Minden $G=(N,T,P,S)$ környezetfüggetlen (2-es típusú) grammatikához megkonstruálható egy vele ekvivalens $G'=(N',T,P',S')$ környezetfüggetlen grammatika úgy, hogy P' -ben **nincs** $A \rightarrow \epsilon$ alakú szabály, kivéve, ha $\epsilon \in L(G)$, mert akkor $S' \rightarrow \epsilon \in P'$, de ekkor S' nem szerepelhet szabály jobboldalán.

ϵ -mentesítés

Első lépésben meghatározzuk, hogy mely nemterminálisokból vezethető le az üres szó.

$$H := \{ A \in N \mid A \xRightarrow[G]{*} \epsilon \}$$

Ehhez definiáljuk a H_i ($i \geq 1$) halmazokat:

$$H_1 := \{ A \in N \mid \exists A \rightarrow \epsilon \in P \}$$

$$H_{i+1} := H_i \cup \{ A \in N \mid \exists A \rightarrow w \in P \text{ és } w \in H_i^* \}$$

$$H_1 \subseteq H_2 \subseteq \dots \subseteq H_k = H_{k+1} \exists k \text{ és legyen } H := H_k.$$

ε -mentesítés

Ekkor látható,
ha $A \in N$ és $A \xRightarrow[G]{*} \varepsilon$, akkor, és csak akkor, ha $A \in H$.

Ennek következménye, hogy
 $\varepsilon \in L(G)$, akkor, és csak akkor, ha $S \in H$.

ϵ -mentesítés

Második lépésben átalakítjuk H ismeretében a grammatika szabályait a kellő alakúra.

$S \notin H$ esetén:

$A \rightarrow v' \in P'$, akkor, és csak akkor, ha $v' \neq \epsilon$ és $\exists A \rightarrow v \in P$ úgy, hogy v' -t v -ből úgy kapjuk, hogy elhagyunk nulla vagy több H -beli nemterminálist v -ből.

ϵ -mentesítés

$S \in H$ estén:

A korábbi szabályokhoz hozzá vesszük még a következő két szabályt:

$$S' \rightarrow \epsilon \text{ és } S' \rightarrow S$$

,ahol $S' \notin N$ a G' grammatika új kezdőszimbóluma.

Megjegyzés: Az átalakítás megőrzi a 2. és 3. típust.

Példa

$$L = \{ u \in \{a,b\}^* \mid \ell_a(u) = \ell_b(u) \geq 0 \}$$

$$G = (\{S\}, \{a,b\}, P, S)$$

$$P: S \rightarrow aSbS$$

$$S \rightarrow bSaS$$

$$S \rightarrow \varepsilon$$

és $L(G) = L$. (Ezt korábban bizonyítottuk.)

Példa

$L = \{ u \in \{a,b\}^* \mid \ell_a(u) = \ell_b(u) \}$, azaz ugyanannyi „a” és „b” van a szavakban.

$G = (\{S\}, \{a,b\}, P, S)$ $G' = (\{S', S\}, \{a,b\}, P', S')$

$P: S \rightarrow aSbS$

$P': S \rightarrow aSbS$

$S \rightarrow abS$

$S \rightarrow aSb$

$S \rightarrow ab$

$S \rightarrow bSaS$

$S \rightarrow bSaS$

$S \rightarrow baS$

$S \rightarrow bSa$

$S \rightarrow ba$

$S \rightarrow \varepsilon$

$S' \rightarrow \varepsilon, S' \rightarrow S$

és $L(G) = L(G')$.

Példa

u=abba szó levezetése:

$$S \xRightarrow{G} aSbS \xRightarrow{G} aSbbSaS \xRightarrow{G} abbSaS \xRightarrow{G} abbaS \xRightarrow{G} abba$$

$$S' \xRightarrow{G'} S \xRightarrow{G'} abS \xRightarrow{G'} abba$$

Példa

$G = (\{S, A, B\}, \{a, b\}, P, S)$

$P: S \rightarrow aAS$

$S \rightarrow AaB$

$S \rightarrow AB$

$A \rightarrow BB$

$B \rightarrow bA$

$B \rightarrow \epsilon$

$H_1 = \{B\}$

$H_2 = H_1 \cup \{A\} = \{A, B\}$

$H_3 = H_2 \cup \{S\} = \{A, B, S\} = N$

$H = H_3$

$G' = (\{S', S, A, B\}, \{a, b\}, P', S')$

$P': S \rightarrow aAS \mid aS \mid aA \mid a$

$S \rightarrow AaB \mid aB \mid Aa \mid a$

$S \rightarrow AB \mid B \mid A$

$A \rightarrow BB \mid B$

$B \rightarrow bA \mid b$

$S' \rightarrow \epsilon$

$S' \rightarrow S$

Emlékeztető

Reguláris műveletek:

- **unió**
- **konkatenáció**
- **(iteratív) lezárás**

Nyelvosztályok zártsága a reguláris műveletekre

Tétel:

Az \mathcal{L}_i , $i = 0, 1, 2, 3$ nyelvosztályok mindegyike zárt a reguláris műveletekre nézve.

Nyelvosztály zártsága nyelvi műveletre

Legyen φ n -változós nyelvi művelet, azaz
ha L_1, \dots, L_n nyelvek, akkor $\varphi(L_1, \dots, L_n)$ is nyelv.

Az \mathcal{L} nyelvcsalád zárt a φ műveletre nézve,
ha $L_1, \dots, L_n \in \mathcal{L}$ estén $\varphi(L_1, \dots, L_n) \in \mathcal{L}$.

Unió

$G=(N,T,P,S)$ legyen az L nyelvhez tartozó grammatika és
 $G'=(N',T,P',S')$ legyen az L' nyelvhez tartozó grammatika és
 $N \cap N' = \emptyset$ és G, G' azonos típusúak.

$i=0,2,3$ esetén, legyen S_0 új szimbólum, azaz $S_0 \notin (N \cup N')$.

$G_U = (N \cup N' \cup \{S_0\}, T, P \cup P' \cup \{S_0 \rightarrow S, S_0 \rightarrow S'\}, S_0)$.

Látható, hogy G_U típusa megegyezik G, G' típusával,
és $L(G) \cup L(G') = L(G_U)$

Unió

$i=1$ estén, ha $\varepsilon \in (L \cup L')$, akkor az előbbi módon elkészített grammatikában nem teljesül a KES.

Ezért tekintsük az $L_1 = L \setminus \{\varepsilon\}$ és $L_2 = L' \setminus \{\varepsilon\}$ nyelveket, amelyeket G_1 és G_2 1-es típusú grammatikák generálnak. Készítsük el G_U -t az előbbi módon, majd vezessünk be egy S_1 új kezdőszimbólumot és adjuk a szabályhalmazhoz az $S_1 \rightarrow \varepsilon$ és $S_1 \rightarrow S_0$ szabályokat.

Emlékeztető

$G=(N,T,P,S)$ grammatika **3-as típusú**, ha szabályai

$A \rightarrow uB$ alakúak, ahol $A,B \in N$, $u \in T^*$ vagy

$A \rightarrow u$ alakúak, ahol $A \in N$, $u \in T^*$

Konkatenáció

(Megjegyzés: Most csak a 3-as típusra bizonyítjuk.)

Legyen $i = 3$.

A P szabályhalmazból megkonstruálunk egy P_1 szabályhalmazt úgy, hogy minden $A \rightarrow u$ alakú szabályt felcserélünk egy $A \rightarrow uS'$ alakú szabályra, a többi szabályt változatlanul hagyjuk.

A $G_c = (N \cup N', T, P_1 \cup P', S)$ grammatika 3-as típusú és generálja az $L(G)L(G')$ nyelvet.

Lezárás

(Megjegyzés: Most csak a 3-as típusra bizonyítjuk.)

Legyen $i = 3$.

Legyen S_0 új szimbólum, azaz $S_0 \notin N$.

Definiáljuk a P_1 szabályhalmazt úgy, hogy minden

$A \rightarrow u$ alakú szabályt felcserélünk egy $A \rightarrow uS_0$ alakú szabályra és ezek legyenek a P_1 elemei.

$G_* = (N \cup \{S_0\}, T, P_1 \cup P \cup \{S_0 \rightarrow \varepsilon, S_0 \rightarrow S\}, S_0)$

grammatika generálja az L^* nyelvet.

Emlékeztető

Reguláris műveletek:

- ▶ unió, $(L_1 \cup L_2)$
- ▶ konkatenáció, $(L_1 L_2)$
- ▶ (iteratív) lezárás. $(L^* = L^0 \cup L^1 \cup L^2 \cup \dots)$

3-as nyelvcsalád leírásai

A 3-as nyelvcsalád nyelveit leírhatjuk

- 3-as típusú grammatikával,
- reguláris kifejezéssel,
- véges determinisztikus automatával,
- véges nemdeterminisztikus automatával.

Bizonyítható, hogy

$$\mathcal{L}_3 = \mathcal{L}_{\text{reg}} = \mathcal{L}_{\text{VDA}} = \mathcal{L}_{\text{VNDA}} \cdot$$

Megjegyzés: A programozási nyelvek lexikális egységei a 3-as nyelvcsaládba tartoznak.

Reguláris nyelvek (rekurzív definíció)

- az elemi nyelvek: \emptyset , $\{\varepsilon\}$, $\{a\}$, ahol $a \in U$, azaz egy tetszőleges betű
- azon nyelvek, melyek az elemi nyelvekből az unió, a konkatenáció és a lezárás műveletek véges számú alkalmazásával állnak elő;
- nincs más reguláris nyelv

Példa: $\{\{a\} \cup \{b\}\}^*\{b\} = \{ub \mid u \in \{a,b\}^*\}$

Reguláris nyelvek

Tétel: Minden L reguláris nyelvhez megadható egy G 3-as típusú grammatika, amelyre $L=L(G)$. ($\mathcal{L}_{\text{reg}} \subseteq \mathcal{L}_3$)

Bizonyítás:

Elemi nyelvekhez adható 3-as típusú grammatika.

$G=(\{S\},\{a\},\{S\rightarrow aS\},S)$	$L(G)=\emptyset$
---	------------------

$G=(\{S\},\{a\},\{S\rightarrow \varepsilon\},S)$	$L(G)=\{\varepsilon\}$
--	------------------------

$G=(\{S\},\{a\},\{S\rightarrow a\},S)$	$L(G)=\{a\}$
--	--------------

Bizonyítás folytatása

Korábban láttuk, hogy az \mathcal{L}_3 nyelvcsalád zárt a reguláris műveletekre nézve.

Az elemi nyelvek grammatikáiból kiindulva megkonstruálható a reguláris műveletekhez tartozó grammatika konstrukciókkal a megfelelő 3-as típusú grammatika bármely összetett reguláris nyelvhez.

Reguláris kifejezések:

Definíció:

- az elemi regularis kifejezések: \emptyset , ε , a , ahol $a \in U$
- ha R_1 és R_2 és R regularis kifejezések akkor
 $(R_1 \mid R_2)$;
 $(R_1 R_2)$;
 $(R)^*$ is reguláris kifejezések.
- a reguláris kifejezések halmaza a legszűkebb halmaz, melyre a fenti két pont teljesül.

Reguláris kifejezések:

Jelölje L_R az R reguláris kifejezéshez tartozó nyelvet.

$$L_{\emptyset} = \emptyset, L_{\varepsilon} = \{\varepsilon\}, L_a = \{a\}$$

Ha Q és R reguláris kifejezések, akkor

$$L_{(Q|R)} = L_Q \cup L_R$$

unió

$$L_{(QR)} = L_Q L_R$$

konkatenáció

$$L_{(R)^*} = (L_R)^*$$

lezárás

Reguláris kifejezések:

A műveletek prioritási sorrendje növekvően:
unió, konkatenáció, lezárás.

A zárójelek elhagyhatók a reguláris kifejezésekből a prioritásoknak megfelelően.

Példák reguláris kifejezésekre

- $(a|b)^*b$, ahol
 $L_{(a|b)^*b} = \{\{a\} \cup \{b\}\}^*\{b\} = \{ub \mid u \in \{a,b\}^*\}$
- $L_{aa^*b^*} = \{a, aa, ab, aaa, aab, abb, \dots\}$
 $aba \notin L_{aa^*b^*}$

Példák reguláris kifejezésekre

$0|1(0|1)^*0$ $T=\{0,1\}$

A fenti reguláris kifejezésnek a páros bináris számoknak felelnek meg, vezető nullák nélkül.

Érdekes helyek, ahol gyakorolhatók a reguláris kifejezések.

<https://regexone.com/>

<https://regexcrossword.com/>

Megjegyzés: A megadott kifejezések a Flex programgenerátor kifejezéseinek részét képezik.

3-típusú grammatikák normál formája

Tétel:

Minden 3-as típusú, nyelv generálható egy olyan grammatikával, amelynek szabályai

$A \rightarrow aB$, ahol $A, B \in N$ és $a \in T$ vagy

$A \rightarrow \varepsilon$ alakúak, ahol $A \in N$.

Megjegyzés: A 3-as normál forma alakítható majd át könnyen automatává.

3-as típusú grammatikák normálformára hozása

Legyen $G=(N,T,P,S)$ 3-as típusú grammatika.

Megkonstruálunk egy $G'=(N',T,P',S)$ 3-as normál formájú grammatika, melyre $L(G)=L(G')$.

Lépesei:

1. hosszredukció
2. befejező szabályok átalakítása
3. láncmentesítés

Hosszredukció

Elhagyjuk az $A \rightarrow a_1 \dots a_k B$ alakú szabályokat,
ahol $k \geq 2$ és $\forall i \in [1, k]: a_i \in T$ és $A \in N$, $B \in N$ vagy $B = \epsilon$.

Helyettesítjük a következő szabályokkal:

$A \rightarrow a_1 Z_1$, ahol $Z_1 \notin N$, azaz új nemterminális,

$Z_1 \rightarrow a_2 Z_2$, ahol $Z_2 \notin (N \cup Z_1)$

...

$Z_{k-1} \rightarrow a_k B$

Megjegyzés: Minden szabályra új nemterminálisokat vezetünk be.

Befejező szabályok átalakítása

Elhagyjuk az $A \rightarrow a$ alakú szabályokat, ahol $a \in T$ és $A \in N$.

Legyen E egy új nemterminális.
(Ez lehet közös minden befejező szabály esetén.)

Vegyük fel P' -be az

$A \rightarrow aE$ és a $E \rightarrow \epsilon$ szabályokat az előbbiek helyett.

Megjegyzés: A 3-as normál forma alakítható majd át könnyen automatává.

Láncmentesítés

Elhagyjuk az $A \rightarrow B$ alakú szabályokat P -ből, ahol $A, B \in N$.

Első lépésben meghatározzuk minden $A \in N$ esetén a

$H(A) := \{ B \in N \mid A \xRightarrow[G]{*} B \}$ halmazokat.

Ehhez definiáljuk a H_i ($i \geq 1$) halmazokat:

$H_1(A) = \{ A \}$

$H_{i+1}(A) = H_i(A) \cup \{ B \in N \mid \exists C \in H_i(A) \text{ és } C \rightarrow B \in P \}$

$H_1(A) \subseteq H_2(A) \subseteq \dots \subseteq H_k(A) = H_{k+1}(A) \exists k$ és legyen $H(A) := H_k(A)$.

Ezután P' -be felvesszük az $A \rightarrow X$ szabályokat,

ha $\exists B \in H(A)$ és $B \rightarrow X \in P$, ahol $X \in (T \cup N)^*$ és X nem csak egyetlen nemterminális.

Példa

P: $S \rightarrow abS$

$S \rightarrow B$

$B \rightarrow bB$

$B \rightarrow V$

$V \rightarrow aa$

hosszredukció után:

P': $S \rightarrow aZ$

$Z \rightarrow bS$

$S \rightarrow B$

$B \rightarrow bB$

$B \rightarrow V$

$V \rightarrow aY$

$Y \rightarrow aE$

$E \rightarrow \varepsilon$

Példa

$P': S \rightarrow aZ$

$Z \rightarrow bS$

$S \rightarrow B$

$B \rightarrow bB$

$B \rightarrow V$

$V \rightarrow aY$

$Y \rightarrow aE$

$E \rightarrow \varepsilon$

$H(S) = \{S, B, V\}$

$H(B) = \{B, V\}$

$H(V) = \{V\}$

láncmentesítés után:

$P': S \rightarrow aZ$

$Z \rightarrow bS$

$S \rightarrow bB \mid aY$

$B \rightarrow bB$

$B \rightarrow aY$

$V \rightarrow aY$

$Y \rightarrow aE$

$E \rightarrow \varepsilon$

Véges determinisztikus automata (VDA)

Definíció:

$A = (Q, T, \delta, q_0, F)$ rendezett ötöst véges determinisztikus automatának nevezzük, ahol

- Q az állapotok nem üres véges halmaza,
- T az input szimbólumok ábécéje,
- $\delta: Q \times T \rightarrow Q$ leképezés az állapot-átmeneti függvény,
- $q_0 \in Q$ a kezdőállapot,
- $F \subseteq Q$ elfogadóállapotok halmaza.

Véges determinisztikus automata (VDA)

Véges determinisztikus automata esetén a

$\delta: Q \times T \rightarrow Q$ állapot-átmeneti függvény

$\forall (q,a)$ párra értelmezett, ahol

$(q,a) \in Q \times T$ és egyetlen olyan $p \in Q$ állapot van, amelyre $\delta(q,a) = p$.

Példa

Legyen $A = (Q, T, \delta, q_0, F)$ a következő, ahol

$Q = \{q_0, q_1, q_2, q_3\}$, $T = \{a, b\}$, $F = \{q_0\}$ és

$\delta(q_0, a) = q_2$, $\delta(q_0, b) = q_1$,

$\delta(q_1, a) = q_3$, $\delta(q_1, b) = q_0$,

$\delta(q_2, a) = q_0$, $\delta(q_2, b) = q_3$,

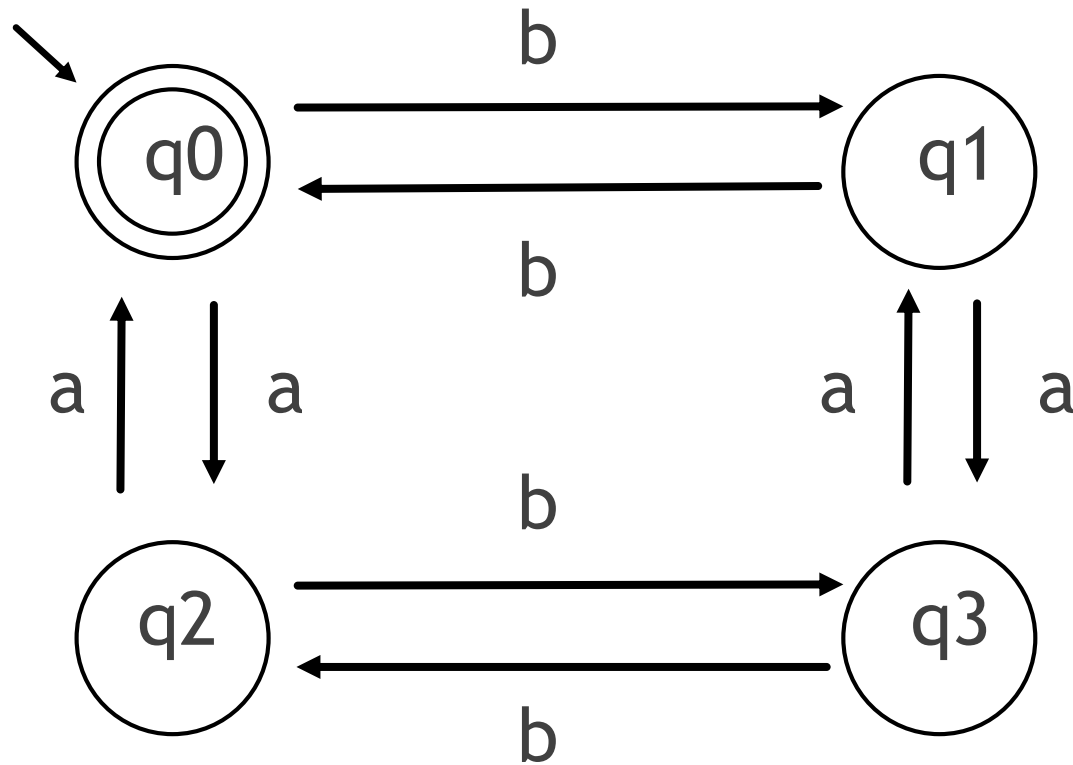
$\delta(q_3, a) = q_1$, $\delta(q_3, b) = q_2$.

$L(A) = \{u \in T^* \mid u\text{-ban páros sok } 'a' \text{ betű és páros sok } 'b' \text{ betű van}\}$

Példa - automata megadása táblázattal

δ	a	b
$\begin{matrix} \rightarrow \\ \leftarrow \end{matrix}$ q0	q2	q1
q1	q3	q0
q2	q0	q3
q3	q1	q2

Példa - automata megadása gráffal



δ	a	b
\Rightarrow q0	q2	q1
q1	q3	q0
q2	q0	q3
q3	q1	q2

Automata működése

δ	a	b
q0	q2	q1
q1	q3	q0
q2	q0	q3
q3	q1	q2

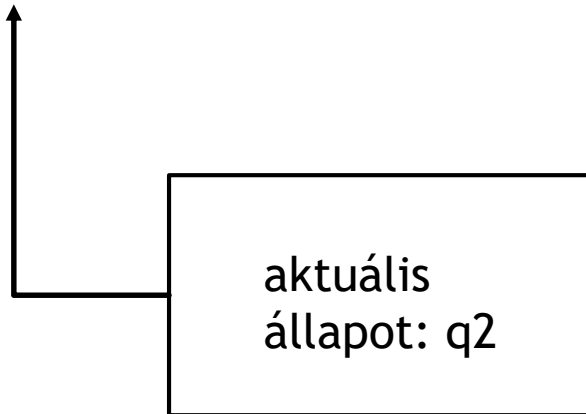
a	b	b	a	b	b		
---	---	---	---	---	---	--	--



aktuális
állapot: q0

Automata működése

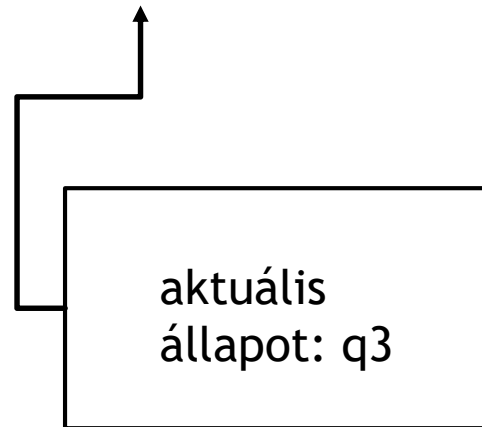
δ	a	b
q0	q2	q1
q1	q3	q0
q2	q0	q3
q3	q1	q2



Automata működése

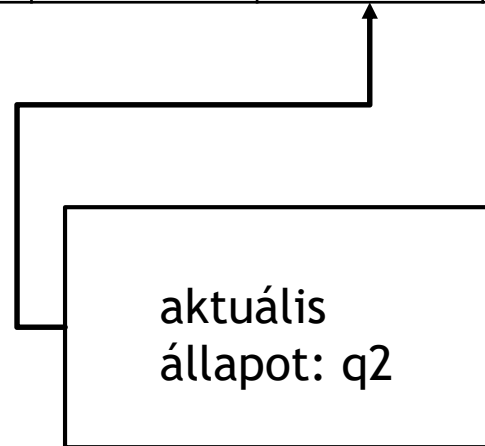
δ	a	b
q0	q2	q1
q1	q3	q0
q2	q0	q3
q3	q1	q2

a	b	b	a	b	b		
---	---	---	---	---	---	--	--



Automata működése

δ	a	b
q0	q2	q1
q1	q3	q0
q2	q0	q3
q3	q1	q2



Automata működése

δ	a	b
q0	q2	q1
q1	q3	q0
q2	q0	q3
q3	q1	q2

a	b	b	a	b	b		
---	---	---	---	---	---	--	--

aktuális
állapot: q0



Automata működése

δ	a	b
q0	q2	q1
q1	q3	q0
q2	q0	q3
q3	q1	q2



aktuális
állapot: q1



Automata működése

δ	a	b
$\begin{matrix} \rightarrow \\ \leftarrow \end{matrix} q_0$	q2	q1
q1	q3	q0
q2	q0	q3
q3	q1	q2

a	b	b	a	b	b		
---	---	---	---	---	---	--	--

aktuális
állapot: q0

q0 elfogadó
állapot, tehát a
szó jó

Köszönöm a figyelmet!