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:

$$\mathbf{L(G) := \{ u \in T^* \mid S \underset{G}{\Rightarrow}^* u \}}$$

Ekvivalens nyelvtanok:

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

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 $L_1,\ldots, L_n$ nyelvek, akkor $\varphi(L_1,\ldots, L_n)$ is nyelv.

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

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 a korábbi normál formában adottak.

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

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

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

Unió

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

Ezért tekintsük az L_1 = L \ {ε} és L_2 = L' \ {ε} 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 $\mathbf{S_1}\to\varepsilon$ és $\mathbf{S_1}\to\mathbf{S_0}$ 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

$\mathbf{A} \rightarrow \mathbf{u}$ alakú szabályt felcserélünk egy $\mathbf{A} \rightarrow \mathbf{uS'}$ alakú szabályra, a többi szabályt változatlanul hagyjuk.

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

Lezárás

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

Legyen i = 3.

Definiáljuk a P_1 szabályhalmazt úgy, hogy minden

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

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

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

grammatika generálja az L* nyelvet.

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 $\rightarrow$ **ε** 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 $\mathbf{A \to a_1 \ldots a_k B}$ alakú szabályokat,
ahol $k \geq \mathbf{2}$ és $\forall i \in [1,k]$: $a_i \in T$ és $A \in N$, $B \in N$ vagy $B=\varepsilon$.

Helyettesítjük a következő szabályokkal:

$\mathbf{A \to a_1 Z_1}$,ahol $Z_1 \notin N$, azaz új nemterminális,

$\mathbf{Z_1 \to a_2 Z_2}$,ahol $Z_2 \notin (N \cup Z_1)$

…

$\mathbf{Z_{k-1} \to a_k B}$

Megjegyzés: Minden szabályra új nemterminálisokat vezetünk be.

Befejező szabályok átalakítása

Elhagyjuk az **A→a** alakú szabályokat, ahol a∈T és A∈N.

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

Vegyük fel P'-be az

A→aE és a **E→ε** 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

$H(A):=\{ B \in N \mid A \underset{G}{\Rightarrow}^{*} B\}$ halmazokat.

Ehhez definiáljuk a H_i ($i\geq1$) halmazokat:

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

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

$H_1(A) \subseteq H_2(A) \subseteq \ldots \subseteq H_k(A)= H_{k+1}(A)$ $\exists k$ 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’: $S \to aZ$
$Z \to bS$
$S \to B$
$B \to bB$
$B \to V$
$V \to aY$
$Y \to aE$
$E \to \varepsilon$

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

láncmentesítés után:

P’: $S \to aZ$
$Z \to bS$
$S \to bB \mid aY$
$B \to bB$
$B \to aY$
$V \to aY$
$Y \to aE$
$E \to \varepsilon$

Emlékeztető

Reguláris műveletek:

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

3-as nyelvcsalád leírásai

A 3-as nyelvcsalád nyelveit leírhatjuk

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

Bizonyítható, hogy

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

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

Reguláris nyelvek (rekurzív definíció)

- az elemi nyelvek: Ø, {ε}, {a} , ahol a ∈ U, azaz egy tetszőleges betű
- azon nyelvek, melyek az elemi nyelvekből az unió, a konkatenació 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). ($\mathcal{L}_{reg} \subseteq \mathcal{L}_3$)

Bizonyítás:

Elemi nyelvekhez adható 3-as típusú grammatika.

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

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

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

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:

Definició:

- az elemi regularis kifejezesek: Ø, ε, a , ahol a ∈ U
- ha R_1 es R_2 és R regularis kifejezesek akkor

  (R_1 | R_2);

  (R_1R_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
  $L_{(a|b)^*b} = \{\{a\} \cup \{b\}\}^*\{b\} = \{ub \mid u \in \{a,b\}^* \}$

- $L_{aa^*b^*} = \{a, aa, ab, aaa, aab, abb, \ldots\}$

  $aba \notin L_{aa^*b^*}$

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.

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 estén 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

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

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

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

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

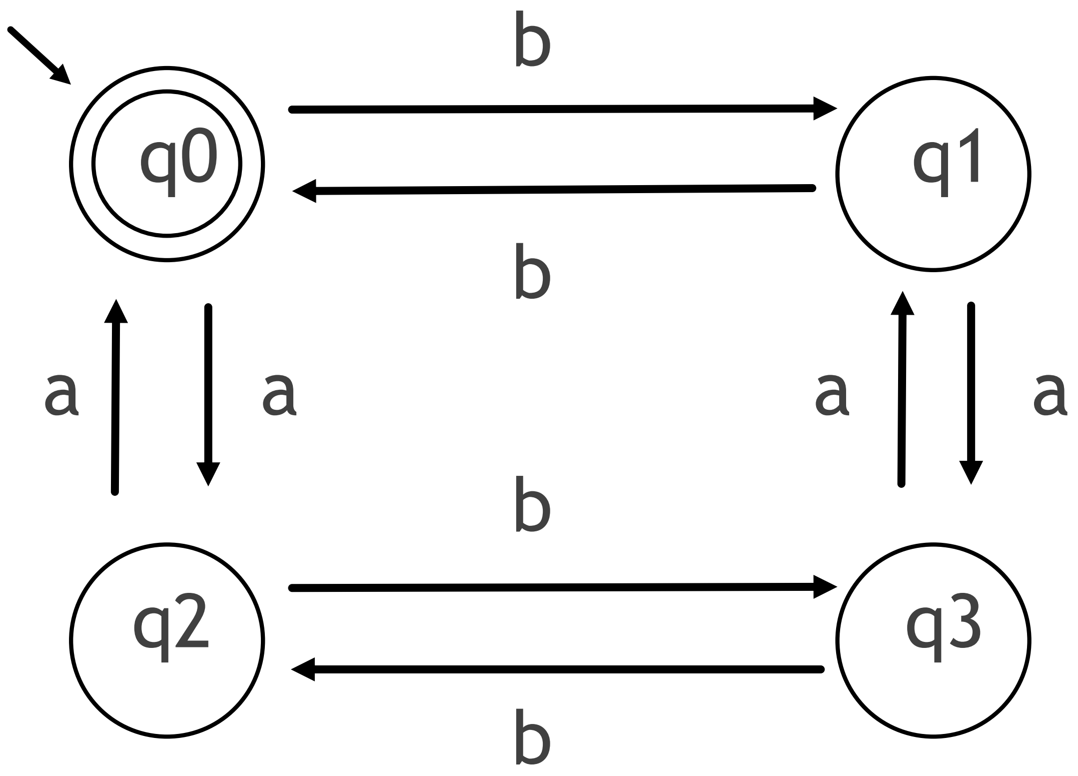
$L(A)=\{u \in T^*$ | u-ban páros sok ‚a' betű és páros sok ‚b' betű van$\}$

Példa – automata megadása táblázattal

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

Példa – automata megadása gráffal

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



Automata működése

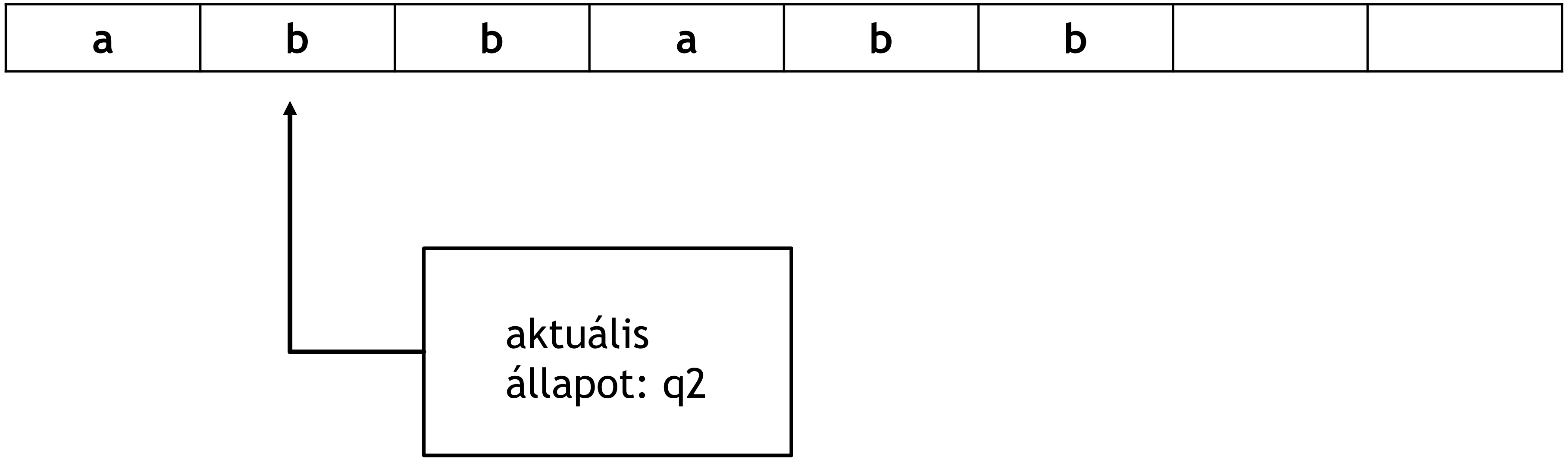
δ	a	b
q0	q2	q1
q1	q3	q0
q2	q0	q3
q3	q1	q2

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



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: q3

Automata működése

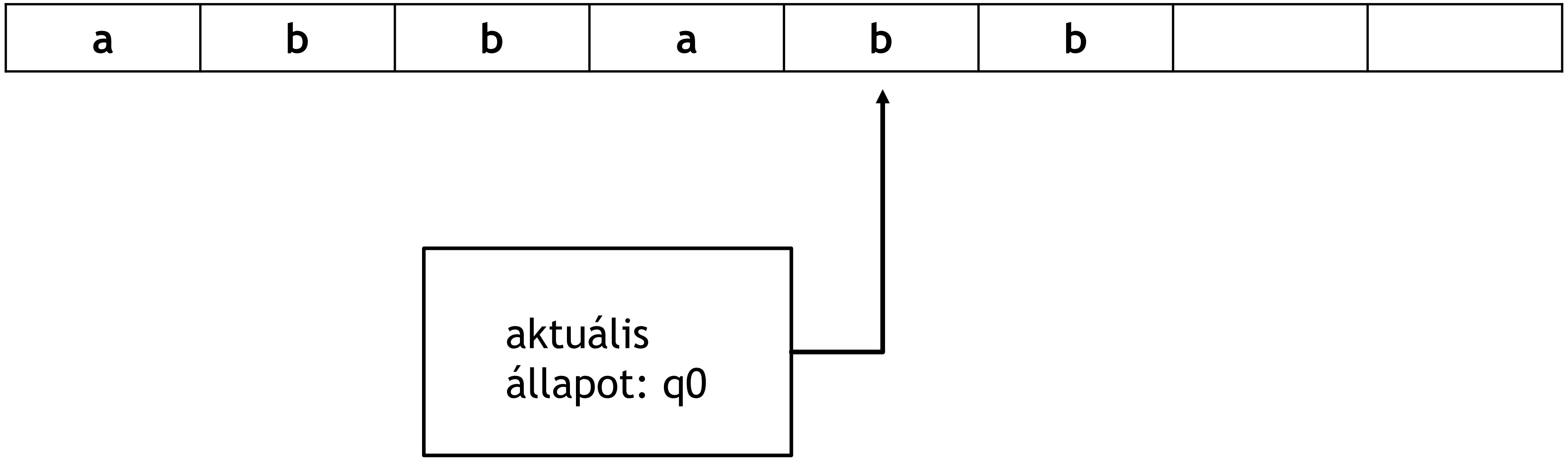
δ	a	b
q0	q2	q1
q1	q3	q0
q2	q0	q3
q3	q1	q2

a	**b**	**b**	**a**	**b**	**b**		

aktuális
állapot: q2

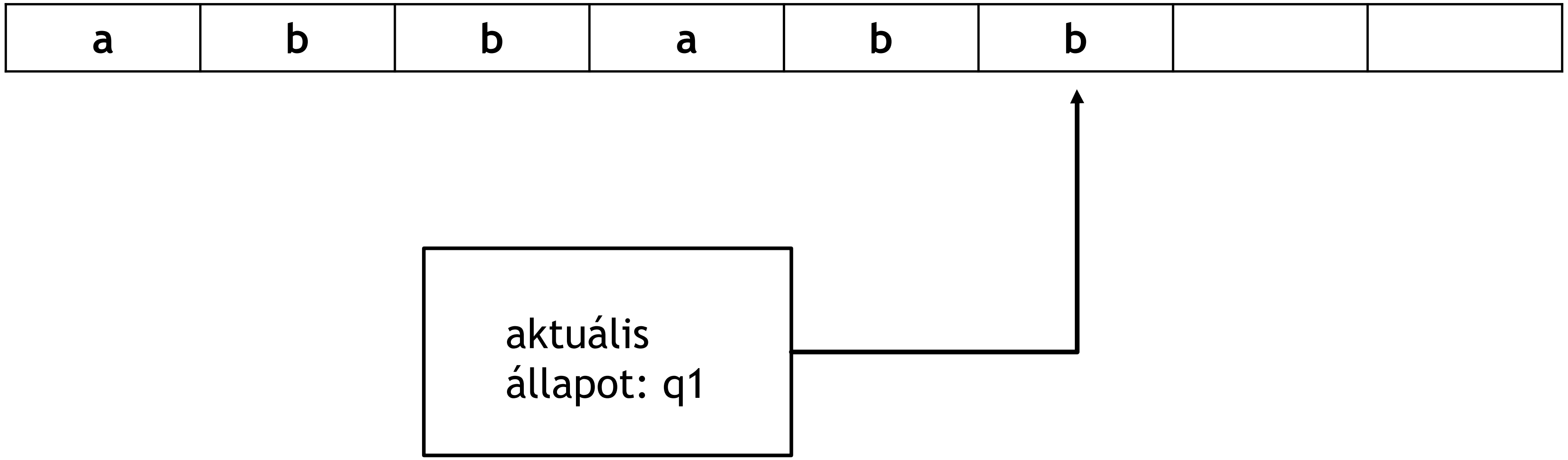
Automata működése

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



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

q0 elfogadó
állapot, tehát a
szó jó

Alternatív jelölés az állapot-átmenetre

$\delta(q, a) = p$ állapot átmenetet jelölhetjük egy

$qa \rightarrow p$ szabállyal.

Ha minden egyes (q, a) párra egyetlen $qa \rightarrow p$ szabály van, akkor a véges automata determinisztikus, egyébként nemdeterminisztikus.

Közvetlen redukció

Legyen $A = (Q, T, \delta, q_0, F)$ egy véges determinisztikus automata és legyenek $\mathbf{u, v \in QT^*}$.
(Konfiguráció: aktuális állapot, input hátralévő része.)

Azt mondjuk, hogy az A automata az u konfigurációt a v konfigurációra redukálja közvetlenül (jelölés: $u \Rightarrow_A v$), ha van olyan

$qa \rightarrow p$ szabály (azaz $\delta(q, a) = p$) és van olyan $w \in T^*$ szó, amelyre

$u = qaw$ és $v = pw$ teljesül.

Redukció

Definíció:

Az $A = (Q, T, \delta, q_0, F)$ véges automata az

$u \in QT^*$ konfigurációt a $v \in QT^*$ konfigurációra redukálja (jelölés: $u \underset{A}{\Rightarrow}^* v$),

ha vagy $u = v$, vagy van olyan

$z \in QT^*$, amelyre $u \underset{A}{\Rightarrow}^* z$ és $z \underset{A}{\Rightarrow} v$ teljesül.

Automata által elfogadott nyelv

Definíció:

Az $A = (Q, T, \delta, q_0, F)$ véges automata által elfogadott nyelv alatt az

$$L(A) := \{u \in T^* \mid \exists q_0 u \underset{A}{\Rightarrow}^* p \text{ és } p \in F\}$$

szavak halmazát értjük.

Megjegyzés: Ez azt jelenti, hogy van olyan működése az automatának, hogy a kezdőállapotból indulva végig olvasva az inputot elfogadóállapotba jut.

Köszönöm a figyelmet!