Számítási modellek

5: Reguláris kifejezések, véges automaták

Reguláris kifejezések

Felhasználások:

- szövegszerkesztőkben keresés csere
- kereső motorok
- szövegfeldolgozó programok (pl. sed és AWK)
- programozási nyelvek, lexikális elemzés
- genomelemzés (genom mint karakterlánc)
- spam/malware szűrő
- ...

Reguláris kifejezések

Legyenek V és $V' = \{\varepsilon, \cdot, +, *, (,)\}$ diszjunkt ábécék. A V ábécé feletti **reguláris kifejezés**eket rekurzív módon a következőképpen definiáljuk:

1. ε reguláris kifejezés V felett,
2. minden $a \in V$ reguláris kifejezés V felett,
3. Ha R reguláris kifejezés V felett, akkor R^* is reguláris kifejezés V felett,
4. Ha Q és R reguláris kifejezések V felett, akkor $(Q \cdot R)$ és $(Q + R)$ is reguláris kifejezések V felett.

* iteráció lezártja
· konkatenáció
\+ unió.

Reguláris kifejezések

Minden reguláris kifejezés egy **reguláris nyelvet reprezentál**, melyet az alábbi rekurzióval definiálunk:

1. az ε a {ε} nyelvet reprezentálja,
2. az a ($\in V$) az {a} nyelvet reprezentálja,
3. ha R az L nyelvet reprezentálja, akkor R^* az L^* nyelvet reprezentálja,
4. ha R az L nyelvet és R' az L' nyelvet reprezentálja, akkor
   ($R \cdot R'$) az LL' nyelvet reprezentálja,
   ($R + R'$) az $L \cup L'$ nyelvet reprezentálja.

Reguláris kifejezések

A zárójelek elhagyhatók, ha precedencia relációt definiálunk definiáljuk a műveleteken. A szokásos sorrend: $*, \cdot, +$.
Az alábbi reguláris kifejezések ekvivalensek:

- a^* ugyanaz, mint $(a)^*$ és a $\{a\}^*$ nyelvet reprezentálja,
- $(a + b)^*$ ugyanaz, mint $((a) + (b))^*$ és a $\{a, b\}^*$ nyelvet reprezentálja,
- $a^* \cdot b$ ugyanaz, mint $((a)^*) \cdot (b)$ és a $\{a\}^*b$ nyelvet reprezentálja,
- $b + ab^*$ ugyanaz, mint $(b) + ((a) \cdot (b)^*)$ és a $\{b\} \cup \{a\}\{b\}^*$ nyelvet reprezentálja,
- $(a + b) \cdot a^*$ ugyanaz, mint $((a) + (b)) \cdot ((a)^*)$ és a $\{a, b\}\{a\}^*$ nyelvet reprezentálja.

Reguláris kifejezések

Legyenek P, Q és R reguláris kifejezések. Ekkor fennáll:

- $P + (Q + R) = (P + Q) + R$
- $P \cdot (Q \cdot R) = (P \cdot Q) \cdot R$
- $P + Q = Q + P$
- $P \cdot (Q + R) = P \cdot Q + P \cdot R$
- $(P + Q) \cdot R = P \cdot R + Q \cdot R$
- $P^* = \varepsilon + P \cdot P^*$
- $\varepsilon \cdot P = P \cdot \varepsilon = P$
- $P^* = (\varepsilon + P)^*$

Reguláris kifejezések

Példa:

- A $(a + b)a^*$ és $aa^* + ba^*$ reguláris kifejezések által reprezentált nyelv ugyanaz:
  $\{ aa^n \mid n \geq 0\} \cup \{ ba^n \mid n \geq 0\}$.

- A $a + ba^*$ reguláris kifejezés által reprezentált nyelv:
  $\{ a, b, ba, ba^2, ba^3, \ldots\}$.

Reguláris kifejezések kifejezőereje

Tétel:

1) Minden reguláris kifejezés egy reguláris (3-típusú) nyelvet reprezentál.

2) Minden reguláris (3-típusú) nyelvhez megadható egy reguláris kifejezés, ami ezt a nyelvet reprezentálja.

Biz.:
1) következik abból, hogy a reguláris nyelvek osztálya $\mathcal{L}_3$ zárt a reguláris műveletekre.

Reguláris kifejezések kifejezőereje

Biz. (folyt.):

2) Megmutatjuk, hogy minden L reguláris nyelvhez, amelyet egy $G = (N, T, P, S)$ reguláris grammatika generál, konstruálható egy reguláris kifejezés, amely L nyelvet reprezentálja.

- Legyen $N = \{A_1, . . . , A_n\}$, $n \geq 1$, $S = A_1$.
- TF. minden szabály $A_i \rightarrow aA_j$ vagy $A_i \rightarrow \varepsilon$ alakú, ahol $a \in T$, $1 \leq i, j \leq n$.
- Azt mondjuk, hogy az A_m nemterminális **érintett** az $A_i \Rightarrow^* uA_j$ ($u \in T^*$) levezetésben, ha A_m előfordul valamely közbülső mondatformában A_i és uA_j között a levezetésben.

Reguláris kifejezések kifejezőereje

Biz. (folyt.):

- Az $A_i \Rightarrow^* uA_j$ levezetés ***k*-megszorított** ha $0 \le m \le k$ érvényes minden A_m nemterminálisra, ami előfordul a levezetésben.
- Legyen $E^k_{i,j} = \{u \in T^* \mid \exists\, A_i \Rightarrow^* uA_j$ k-megszorított levezetés$\}$.
- Megmutatjuk k szerinti indukcióval, hogy $E^k_{i,j}$ nyelvhez létezik reguláris kifejezés, ami $E^k_{i,j}$ nyelvet reprezentálja, $0 \le i,j,k \le n$.

Reguláris kifejezések kifejezőereje

Biz. (folyt.):

- k=0 (indukció kezdete):
  - $i \neq j$ esetén $E^0_{i,j}$ vagy üres, vagy T-beli szimbólumokból áll ($a \in E^0_{i,j}$ pontosan akkor, ha $A_i \rightarrow aA_j \in P$.)
  - $i = j$ esetén $E^0_{i,j}$ tartalmazza ε-t és T nulla vagy annál több elemét. Így $E^0_{i,j}$ reprezentálható reguláris kifejezéssel.

Reguláris kifejezések kifejezőereje

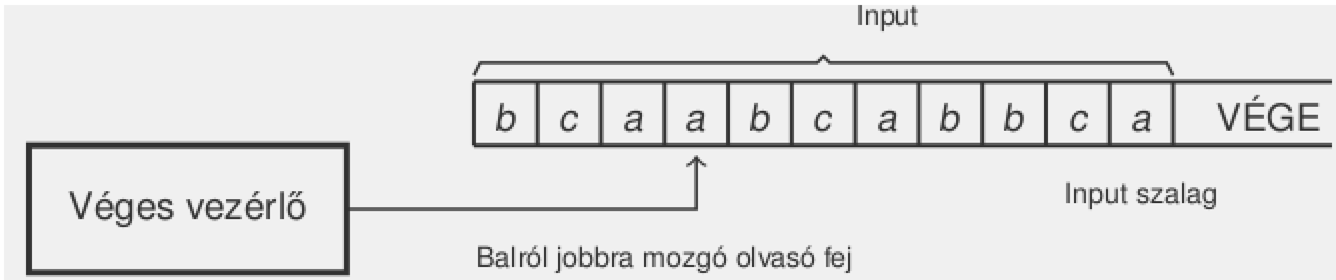
Biz. (folyt.):

- k-1 –>k (indukciós lépés):
  - TF: minden i, j-re $E^{k-1}_{i,j}$ reprezentálható egy reguláris kifejezéssel
  - Ekkor $E^{k}_{i,j} = E^{k-1}_{i,j} + E^{k-1}_{i,k} \cdot (E^{k-1}_{k,k})^* \cdot E^{k-1}_{k,j}$.
  - Ezért $E^{k}_{i,j}$ reprezentálható reguláris kifejezéssel.
  - Legyen I_ε azon i indexek halmaza, melyekre $A_i \to \varepsilon$.
  - Ekkor $L(G) = \cup_{i \in I_\varepsilon} E^{n}_{1,i}$ reprezentálható reguláris kifejezéssel. □

Véges Automata (VA)

- Formális nyelvek azonosítása felismerő eszközökkel is lehetséges. Ilyenek az automaták
- Egy automata alkalmas szavak feldolgozására és azonosítására.
- A grammatikák szintetizáló, míg az automaták analitikus megközelítést alkalmaznak.
- Az automata egy input szót vagy elfogad, vagy elutasít.

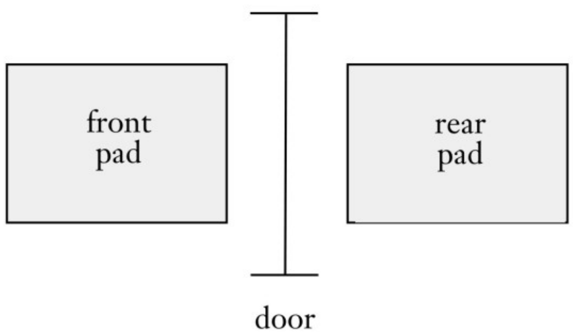
Véges Automata (VA)



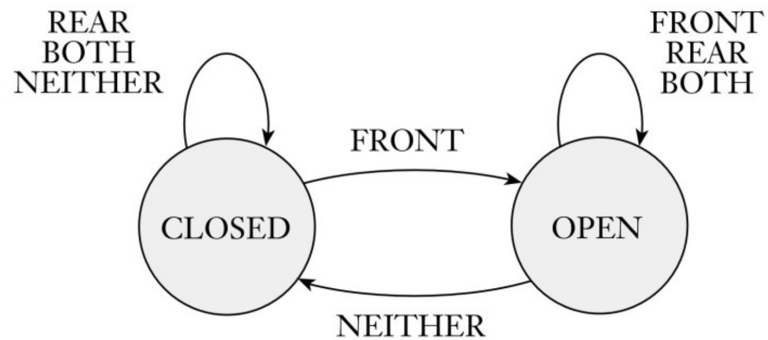
- A véges automata (VA) lépések sorozatát hajtja végre diszkrét időintervallumokban.
- A VA a kezdőállapotból indul.
- Az inputszó az inputszalagon helyezkedik el, az olvasófej az inputszó legbaloldalibb szimbólumán áll.
- A VA, miután elolvasott egy szimbólumot, az olvasófejet egy pozícióval jobbra mozgatja, majd állapotot vált az állapot-átmenet függvény szerint.
- Ha a VA elolvasta az inputot, megáll, elfogadja vagy elutasítja.

Véges Automata (VA)

- Példa: automata ajtó vezérlés



Állapot-átmenet diagramm:



Állapot-átmenet táblázat

state	input signal:	NEITHER	FRONT	REAR	BOTH
	CLOSED	CLOSED	OPEN	CLOSED	CLOSED
	OPEN	CLOSED	OPEN	OPEN	OPEN

Véges Automata (VA)

- Felhasználás:
  - Önműködő ajtó
  - Kávéautomata
  - Mintafelismerés
  - Matematikai rejtvények (pl. átkelés a folyón)
  - Mintafelismerés
  - Beszédfeldolgozás
  - Optikai karakterfelismerés
  - Részvény árak előrejelzése a tőzsdén
  - ...

Véges Automata (VA)

A véges automata egy 5-ös $A = (Q, T, \delta, q_0, F)$, ahol

- Q az **állapotok** egy véges, nemüres halmaza,
- T az **inputszimbólumok** véges ábécéje,
- $\delta : Q \times T \rightarrow Q$ (teljes) leképezés, az ún. **állapot-átmenet függvény**,
- $q_0 \in Q$ a **kezdőállapot**,
- $F \subseteq Q$ az **elfogadó állapotok halmaza**.

Véges Automata (VA)

Megjegyzés:

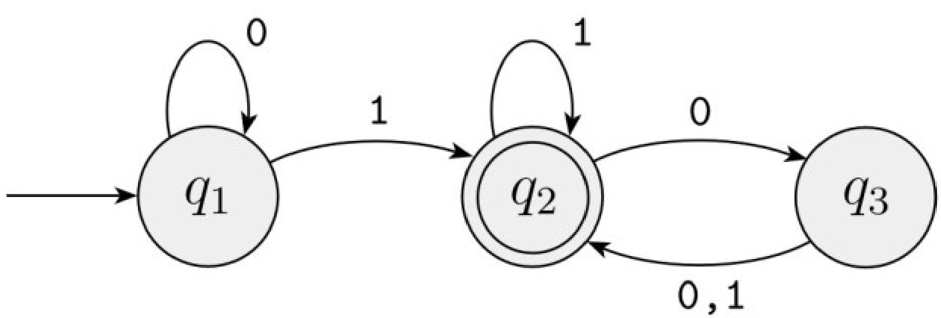
- A δ $Q \times T \rightarrow Q$ leképezés kiterjeszthető $\hat{\delta} : Q \times T^* \rightarrow Q$ leképezéssé:
  - $\hat{\delta}(q, \varepsilon) = q$,
  - $\hat{\delta}(q, xa) = \delta(\hat{\delta}(q, x), a)$ minden $x \in T^*$-ra és $a \in T$-re.

Véges Automata (VA)

Példa:

- Legyen $A = (Q, T, \delta, q_1, F)$ egy VA, ahol
  $Q = \{q_1, q_2, q_3\}$, $T = \{0, 1\}$, $F = \{q_2\}$,
  $\delta(q_1, 0) = q_1$, $\delta(q_1, 1) = q_2$, $\delta(q_2, 0) = q_3$, $\delta(q_2, 1) = q_2$,
  $\delta(q_3, 0) = \delta(q_3, 1) = q_2$.

  Állapot-átmenet diagramm:



  Állapot-átmenet táblázat:

δ	0	1
q_1	q_1	q_2
q_2	q_3	q_2
q_3	q_2	q_2

- Az elfogadott nyelv $L(A)=\{w \mid w$ legalább egy 1-est tartalmaz és az utolsó 1-est páros számú 0 követi$\}$

Véges Automata (VA)

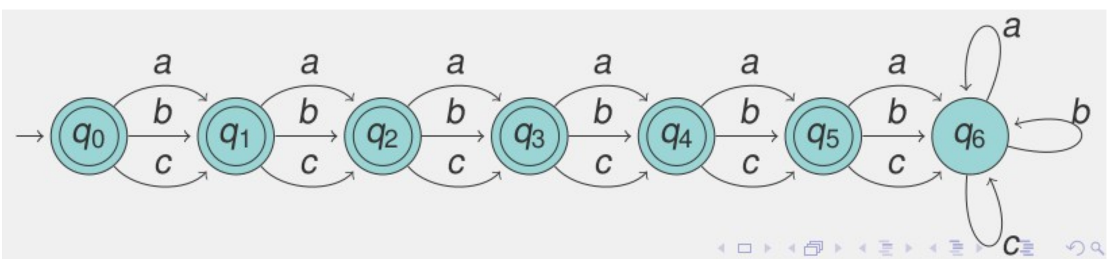
Példa:

- Legyen $T = \{a,b,c\}$.
  Adjunk egy VA-t, ami azokat a szavakat fogadja el, melyek hossza ≤ 5.

  Megoldás:

- Formálisan:
  $A=(\{q_0, \ldots, q_6\}, \{a, b, c\}, \delta, q_0, \{q_0, \ldots, q_5\})$,
  $\delta(q_i, t) = q_{i+1}$, for $i = 0, \ldots, 5$, $t \in \{a, b, c\}$,
  $\delta(q_6, t) = q_6$, for $t \in \{a, b, c\}$

- Állapot-átmenet diagramm:

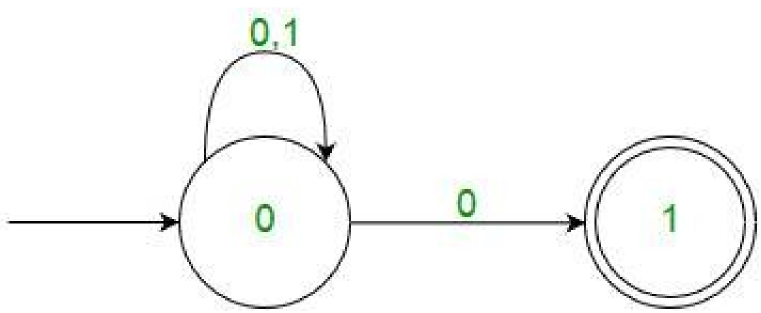


Állapot-átmenet táblázat:

	a	b	c
$\leftrightarrows q_0$	q_1	q_1	q_1
$\leftarrow q_1$	q_2	q_2	q_2
$\leftarrow q_2$	q_3	q_3	q_3
$\leftarrow q_3$	q_4	q_4	q_4
$\leftarrow q_4$	q_5	q_5	q_5
$\leftarrow q_5$	q_6	q_6	q_6
q_6	q_6	q_6	q_6

Determinisztikus és nemdeterminisztikus VA

- **Determinisztikus véges automata (DVA):** A δ függvény egyértékű, azaz $\forall\ (q, a) \in Q \times T$ pontosan egy s állapot van, amire $\delta(q, a) = s$.
- **Nemdeterminisztikus véges automata (NVA)**: A δ függvény többértékű, azaz $\delta : Q \times T \rightarrow 2^Q$.
  - Több kezdeti állapot megengedett, $Q_0 \subseteq Q$ a kezdeti állapotok halmaza.
  - Megengedett, hogy $\delta(q, a) = \emptyset$ valamely (q,a)-ra, azaz a gép elakad.
  - ε-átmenet megengedett, azaz állapot-átmenet új szimbólum olvasása nélkül.



NVA példa

Determinisztikus és nemdeterminisztikus VA

- A nemdeterminizmus új tulajdonságai:
  - Több út is lehetséges (több választás minden lépésnél).
  - Az ε-átmenet egy átmenet az input olvasása nélkül.
  - Elfogadja az inputot, ha valamely út elfogadó állapotba vezet.

Determinisztikus és nemdeterminisztikus VA

- Alternatív jelölés: Az állapot-átmeneteket $qa \to p$ alakú szabályok formájában is megadhatjuk $p \in \delta(q, a)$ esetén.
- Legyen M_δ az $A = (Q, T, \delta, Q_0, F)$ NVA δ állapot-átmenet függvénye által a fenti módon adott szabályok halmaza.
- Ha minden (q, a) párra M_δ pontosan egy $qa \to p$ szabályt tartalmaz, akkor a VA determinisztikus, egyébként nemdeterminisztikus.

VA – redukció

- Legyen $A = (Q, T, \delta, Q_0, F)$ egy VA, legyenek $u,v \in QT^*$ szavak. Az A az u szót **egy lépésben** (**közvetlenül**) a v szóra **redukálja** (jelölés: $u \Rightarrow_A v$, vagy röviden $u \Rightarrow v$), ha van olyan $qa \rightarrow p \in M_\delta$ szabály (azaz $\delta(q, a) = p$) és olyan $w \in T^*$ szó, hogy $u = qaw$ és $v = pw$ teljesül.
- Az $A = (Q, T, \delta, Q_0, F)$ az u szót v szóra **redukálja** (jelölés: $u \Rightarrow_A^* v$, vagy röviden: $u \Rightarrow^* v$, ha
  - vagy $u = v$,
  - vagy $\exists$ olyan szó $z \in QT^*$, amire $u \Rightarrow^* z$ és $z \Rightarrow$ v.
- A $\Rightarrow^* \subseteq QT^* \times QT^*$ reláció a $\Rightarrow$ reláció reflexív, tranzitív lezártja.

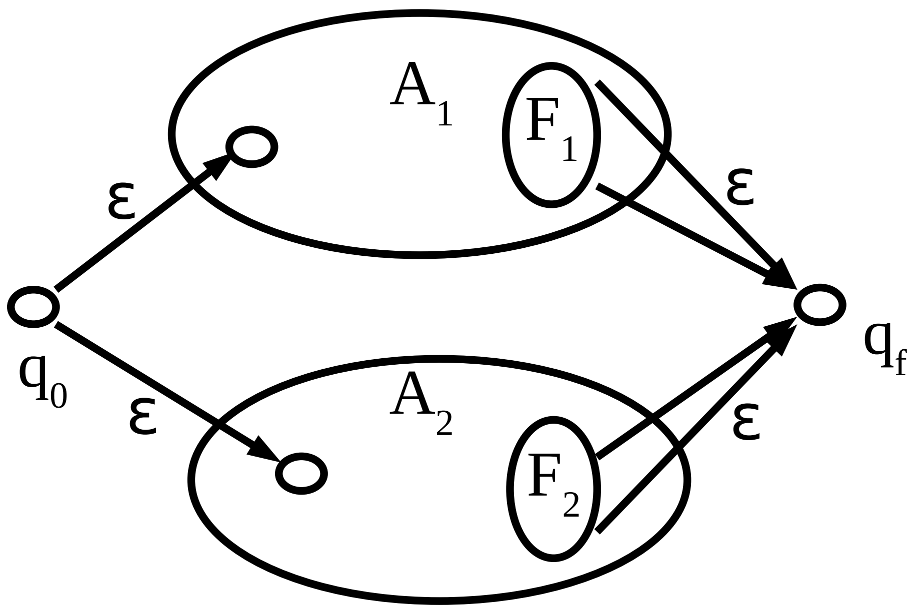
VA – elfogadott nyelv

- Az $A = (Q, T, \delta, q_0, F)$ NVA által **elfogadott/felismert nyelv**: $L(A) = \{u \in T^* \mid q_0u \Rightarrow^* p$ valamely $q_0 \in Q_0$-ra és $p \in F$-re$\}$.
- Egy DVA A esetén egy kezdőállapot van $Q_0=\{q_0\}$. DVA A által elfogadott nyelv: $L(A) = \{u \in T^* \mid q_0u \Rightarrow^* p$ valamely $p \in F$-re$\}$.

NVA L_1 U L_2 elfogadásához

Tétel: Ha L_1 és L_2 reguláris nyelvek, akkor L_1U L_2 is reguláris nyelv.

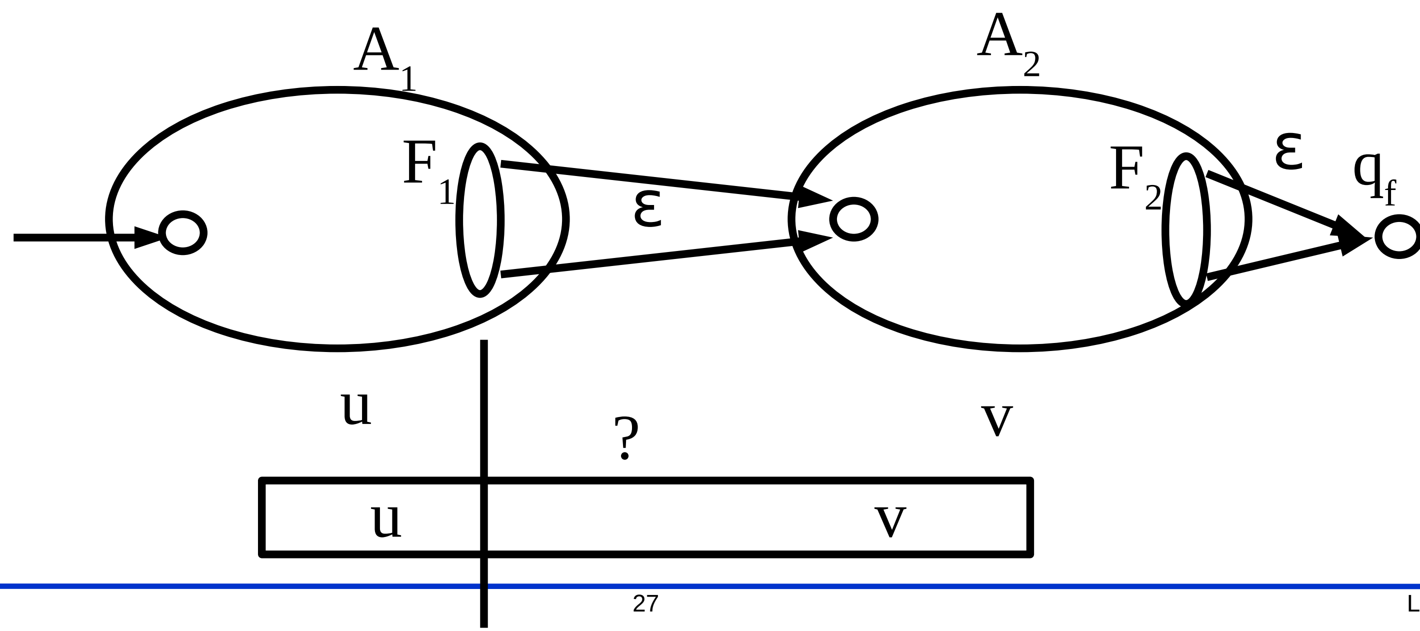
Biz. (vázlat): Legyen A_1 egy DVA, ami elfogadja L_1-t, A_2 egy DVA, ami elfogadja L_2-t. A következő NVA elfogadja L_1U L_2-t.



NFA L_1L_2 elfogadásához

Tétel: Ha L_1 és L_2 reguláris nyelvek, akkor L_1L_2 is reguláris nyelv.

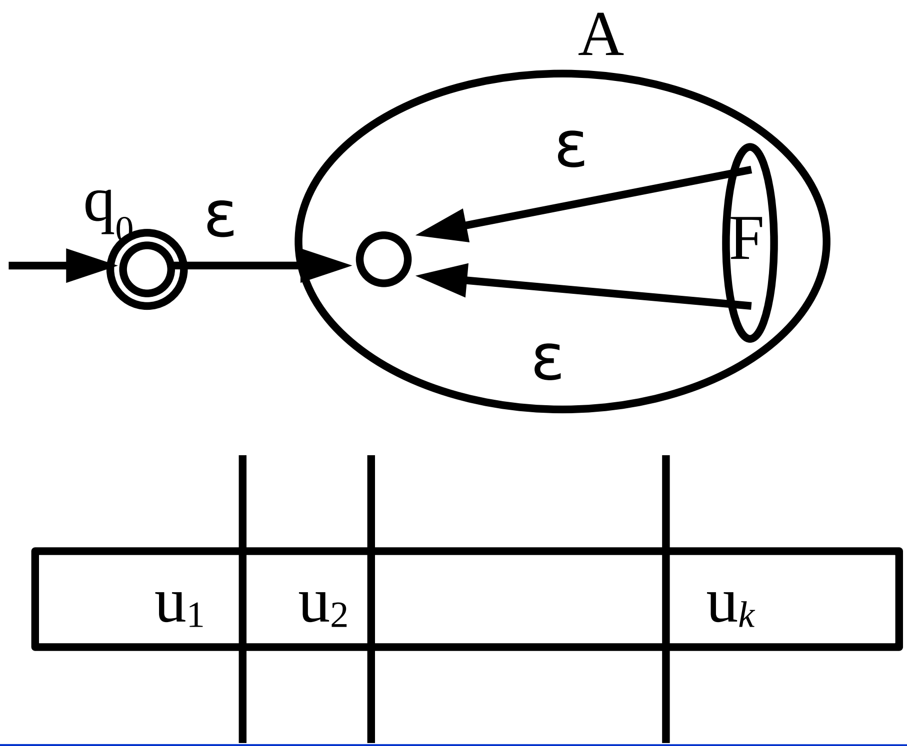
Biz. (vázlat): Legyen A_1 egy DVA, ami elfogadja L_1-t,
A_2 egy DVA, ami elfogadja L_2-t.
A következő NVA elfogadja L_1L_2-t.



NFA *L** elfogadásához

Tétel: Ha *L* egy reguláris nyelv, akkor *L** is reguláris nyelv.

Biz. (vázlat): Legyen *A* egy DVA, ami elfogadja *L*-t,
A következő NVA elfogadja *L**-t.



A NVA számítási ereje

- **Tétel**: Minden $A = (Q, T, \delta, Q_0, F)$ NVA-hoz konstruálható egy DVA $A' = (Q', T, \delta', q'_0, F')$, úgy hogy $L(A) = L(A')$ teljesül.

- Ötlet: A DVA nyomon követi a lehetséges állapotok részhalmazát az NVA-ban
- Megjegyzés: Legrosszabb esetben $|Q'| = 2^{|Q|}$.

A NVA számítási ereje

Biz.:

- Legyen $Q' = 2^Q$, azaz Q' a Q halmaz összes részhalmának halmaza ($|Q'| = 2^{|Q|}$).
- Legyen $\delta' : Q' \times T \rightarrow Q'$ következőképp definiált: $\delta'(q', a) = \bigcup_{q \in q'} \delta(q, a)$.
- Legyen $q'_0 = Q_0$ és $F' = \{q' \in Q' \mid q' \cap F \neq \emptyset\}$
- $L(A) \subseteq L(A')$ bizonyításához, bebizonyítjuk Lemma 1-t. $L(A') \subseteq L(A)$-hoz Lemma 2-t.
- Előbb egy példa.

NVA – DVA

Példa:

- Legyen $A = (Q,T,\delta,Q_0,F)$ egy NVA, ahol
  $Q = \{q_0, q_1, q_2\}$, $T = \{a, b\}$, $Q_0 = \{q_0\}$, $F = \{q_2\}$.
  δ következőképpen adott:
  $\delta(q_0, a) = \{q_0, q_1\}$, $\delta(q_0, b) = \{q_1\}$,
  $\delta(q_1, a) = \emptyset$, $\delta(q_1, b) = \{q_2\}$,
  $\delta(q_2, a) = \{q_0, q_1, q_2\}$, $\delta(q_2, b) = \{q_1\}$.
  Adjunk egy A' DVA-t , ami ekvivalens A-val.

  Megoldás:

- DVA: $A' = (Q',T,\delta',q'_0,F')$, ahol
  $Q' = \{\emptyset,\{q_0\},\{q_1\},\{q_2\},\{q_0, q_1\},\{q_0,q_2\},\{q_1,q_2\},\{q_0,q_1,q_2\}\}$,
  $q'_0 = \{q_0\}$,
  $F' = \{\{q_2\},\{q_0,q_2\},\{q_1,q_2\},\{q_0,q_1,q_2\}\}$,
  folyt. következő fólia

NVA – DVA

Példa (folyt.):

- δ:

$$\begin{array}{cc} \delta(q_0, a) = \{q_0, q_1\}, & \delta(q_0, b) = \{q_1\}, \\ \delta(q_1, a) = \emptyset, & \delta(q_1, b) = \{q_2\}, \\ \delta(q_2, a) = \{q_0, q_1, q_2\}, & \delta(q_2, b) = \{q_1\}. \end{array}$$

- δ':

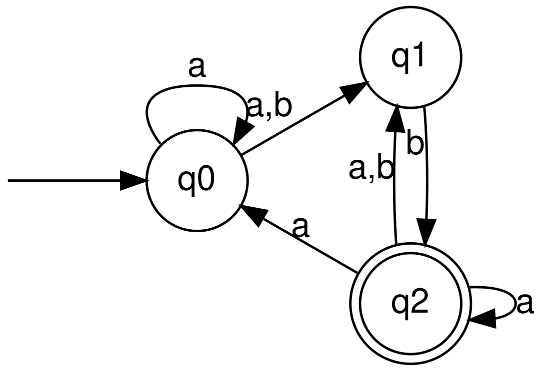
$$\begin{array}{cc} \delta'((\emptyset, a) = \emptyset, & \delta'((\emptyset, b) = \emptyset, \\ \delta'((\{q_0\}, a) = \{q_0, q_1\}, & \delta'((\{q_0\}, b) = \{q_1\}, \\ \delta'((\{q_1\}, a) = \emptyset, & \delta'((\{q_1\}, b) = \{q_2\}, \\ \delta'((\{q_2\}, a) = \{q_0, q_1, q_2\}, & \delta'((\{q_2\}, b) = \{q_1\}, \\ \delta'((\{q_0, q_1\}, a) = \{q_0, q_1\}, & \delta'((\{q_0, q_1\}, b) = \{q_1, q_2\}, \\ \delta'((\{q_0, q_2\}, a) = \{q_0, q_1, q_2\}, & \delta'((\{q_0, q_2\}, b) = \{q_1\}, \\ \delta'((\{q_1, q_2\}, a) = \{q_0, q_1, q_2\}, & \delta'((\{q_1, q_2\}, b) = \{q_1, q_2\}, \\ \delta'((\{q_0, q_1, q_2\}, a) = \{q_0, q_1, q_2\}, & \delta'((\{q_0, q_1, q_2\}, b) = \{q_1, q_2\}. \end{array}$$

NVA – DVA

Példa (folyt.):

NVA

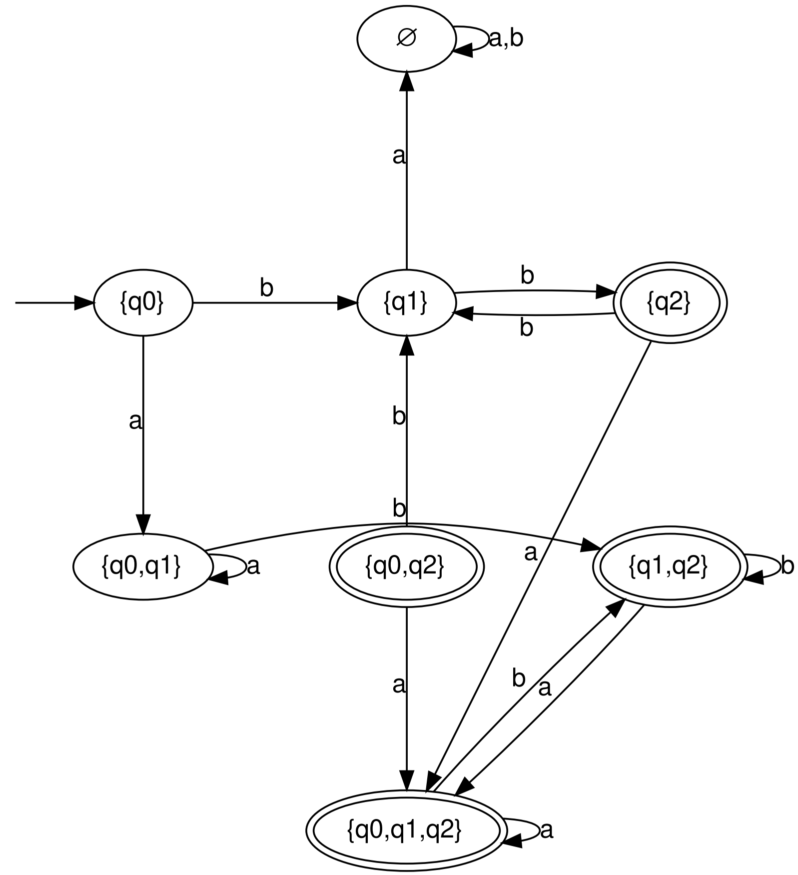
$$\begin{array}{cc} \delta(q_0,a)=\{q_0,q_1\}, & \delta(q_0,b)=\{q_1\}, \\ \delta(q_1,a)=\emptyset, & \delta(q_1,b)=\{q_2\}, \\ \delta(q_2,a)=\{q_0,q_1,q_2\}, & \delta(q_2,b)=\{q_1\}. \end{array}$$



$F = \{q_2\}$

DVA

$$\begin{array}{cc} \delta'((\emptyset,a)=\emptyset, & \delta'((\emptyset,b)=\emptyset, \\ \delta'((\{q_0\},a)=\{q_0,q_1\}, & \delta'((\{q_0\},b)=\{q_1\}, \\ \delta'((\{q_1\},a)=\emptyset, & \delta'((\{q_1\},b)=\{q_2\}, \\ \delta'((\{q_2\},a)=\{q_0,q_1,q_2\}, & \delta'((\{q_2\},b)=\{q_1\}, \\ \delta'((\{q_0,q_1\},a)=\{q_0,q_1\}, & \delta'((\{q_0,q_1\},b)=\{q_1,q_2\}, \\ \delta'((\{q_0,q_2\},a)=\{q_0,q_1,q_2\}, & \delta'((\{q_0,q_2\},b)=\{q_1\}, \\ \delta'((\{q_1,q_2\},a)=\{q_0,q_1,q_2\}, & \delta'((\{q_1,q_2\},b)=\{q_1,q_2\}, \\ \delta'((\{q_0,q_1,q_2\},a)=\{q_0,q_1,q_2\}, & \delta'((\{q_0,q_1,q_2\},b)=\{q_1,q_2\}. \end{array}$$



$F' = \{\{q_2\},\{q_0,q_2\},\{q_1,q_2\},\{q_0,q_1,q_2\}\}$

A NVA számítási ereje

Lemma 1:

- Minden $p,q \in Q$, $q' \in Q'$ és $u,v \in T^*$ esetén,
  ha $qu \Rightarrow^*_A pv$ és $q \in q'$,
  akkor $\exists\, p' \in Q'$, hogy
  $q'u \Rightarrow^*_{A'} p'v$ és $p \in p'$.

Biz.:

- Indukció a redukciós lépések száma n szerint
  $qu \Rightarrow^*_A pv$ redukcióban.
- $n=0$: az állítás triviálisan teljesül, $p'=q'$.

A NVA számítási ereje

Biz. (Lemma 1, folyt.):

- $n \to n+1$:
  TF., az állítás igaz minden $\leq n$ lépésből álló redukcióra.
- Legyen $qu \Rightarrow^*_A pv$ egy $n + 1$ lépésből álló redukció.
  Ekkor valamely $q_1 \in Q$ és $u_1 \in T^*$-re teljesül
  $qu \Rightarrow_A q_1u_1 \Rightarrow^*_A pv$.
- Tehát $\exists\ a \in T$, amire $u = au_1$ és $q_1 \in \delta(q, a)$.
- Mivel $q \in q'$-re $\delta(q, a) \subseteq \delta'(q', a)$,
  q'_1 megadható mint $q'_1 = \delta'(q', a)$.
- Következésképp, $q'u \Rightarrow_{A'} q'_1u_1$, ahol $q_1 \in q'_1$.
- Az indukciós feltétel alapján,
  $\exists\ p' \in Q'$, amire $q'_1u_1 \Rightarrow^*_{A'} p'v$ és $p \in p'$, ami bizonyítja az állítást. $\square$

A NVA számítási ereje

Biz. (Tétel, folyt.):

- Legyen $u \in L(A)$, azaz $q_0 u \Rightarrow^*_A p$, valamely $q_0 \in Q_0$-ra és $p \in F$-re.
- Lemma 1 alapján $\exists\, p'$, amire $q'_0 u \Rightarrow^*_{A'} p'$, ahol $p \in p'$.
- F' definíciója szerint, $p \in p'$-ből és $p \in F$-ből következik, hogy $p' \in F'$, ami bizonyítja, hogy $L(A) \subseteq L(A')$.
- $L(A') \subseteq L(A)$-hoz bizonyítjuk Lemma 2-t.

A NVA számítási ereje

Lemma 2:

- Minden $p', q' \in Q'$, $p \in Q$ és $u, v \in T^*$ esetén,
  - ha $q'u \Rightarrow^*_{A'} p'v$ és $p \in p'$,
  - akkor $\exists\, q \in Q$, amelyre $qu \Rightarrow^*_A pv$ és $q \in q'$.

Biz.:

- Indukció a redukció lépéseinek száma n alapján.
- $n = 0$: Az állítás triviálisan érvényes.

A NVA számítási ereje

Biz. (Lemma 2, folyt.):

- $n \to n+1$: TF. Az állítás igaz minden $\leq n$ lépésből álló redukcióra.
- Legyen $q'u \Rightarrow^*_{A'} p'v$ egy $n + 1$ lépésből álló redukció. Ekkor $q'u \Rightarrow^*_{A'} p'_1 v_1 \Rightarrow_{A'} p'v$, ahol $v_1 = av$, valamely $p'_1 \in Q'$-ra és $a \in T$-re.
- Ekkor $p \in p' = \delta'(p'_1, a) = \bigcup_{p_1 \in p'_1} \delta(p_1, a)$.
- Következésképpen, $\exists\, p_1 \in p'_1$, amelyre $p \in \delta(p_1, a)$.
- Erre a p_1-re teljesül, hogy $p_1 v_1 \Rightarrow_A pv$.
- Az indukciós feltétel alapján, $qu \Rightarrow^*_A p_1 v_1$, valamely $q \in q_0$-ra, amiből következik az állítás. $\square$

A NVA számítási ereje

Biz. (Tétel, folyt.):

- Legyen $q'_0u \Rightarrow^*_{A'} p'$ és $p' \in F$.
- F' definíciója alapján $\exists\, p \in p'$, amelyre $p \in F$.
- Ekkor, Lemma 2 alapján, valamely $q_0 \in q'_0$-ra teljesül, hogy $q_0u \Rightarrow^*_A p$.
- Ez bizonyítja a tétel állítását. $\square$

Következmények

Korollár 1:

- A reguláris nyelvek osztálya $\mathcal{L}_3$ zárt a komplemens operációra.

Biz.:

- Legyen L egy nyelv, amelyet egy $A = (Q,T,\delta,q_0,F)$ VA felismer/elfogad.

- Ekkor a $\overline{L} = T^* - L$ nyelv felismerherő az $\overline{A} = (Q,T,\delta,q_0,Q-F)$ VA-val.

Következmények

Korollár 2:

- A reguláris nyelvek osztálya $\mathcal{L}_3$ zárt a metszet operációra.

Biz.:

- Tudjuk, hogy $\mathcal{L}_3$ zárt az unió operációra.
- $L_1 \cap L_2 = \overline{\overline{L_1} \cup \overline{L_2}}$.
- Korollár 1 alapján következik az állítás.

VA – Myhill-Nerode Tétel

- Legyen L egy T ábécé feletti nyelv. Az ***L* nyelv által indukált $\boldsymbol{E_L}$ reláció** egy olyan bináris reláció T^*-on, amelyre teljesül, hogy

  $\forall\, u, v \in T^*$-ra uE_Lv akkor és csak akkor, ha

  $\nexists\, w \in T^*$, melyre az uw és vw szavak közül pontosan az egyik eleme L-nek (azaz $\forall\, w \in T^* : uw \in L \Leftrightarrow vw \in L$).

- E_L egy **ekvivalencia reláció** és **jobb-invariáns**. (Jobb-invariáns: ha uE_Lv, akkor uwE_Lvw is fennáll $\forall\, w \in T^*$ szóra.)

- Az **$\boldsymbol{E_L}$ indexe** az ekvivalencia osztályok száma.

Tétel (Myhill-Nerode): $L \subseteq T^*$ pontosan akkor ismerhető fel egy DVA-val, ha E_L véges indexű.

VA – Myhill-Nerode Tétel

Tétel (Myhill-Nerode): $L \subseteq T^*$ pontosan akkor ismerhető fel egy DVA-val, ha E_L véges indexű.

- Ez index egyenlő egyenlő az L-t felismerő minimális DVA állapotainak számával.

Minimális állapotszámú DVA

- Egy *A* DVA minimális állapotszámú (**minimális DVA**), ha nincs olyan *A’* DVA, amely ugyanazt a nyelvet ismeri fel, mint *A*, de *A’* állapotainak száma kisebb, mint *A* állapotainak száma.

Tétel: Az *L* reguláris nyelvet elfogadó minimális állapotszámú DVA az izomorfizmus erejéig egyértelmű.

Minimális állapotszámú DVA

Tétel: Az L reguláris nyelvet elfogadó minimális állapotszámú DVA az izomorfizmus erejéig egyértelmű.

- Legyen $A = (Q, T, \delta, q_0, F)$ egy DVA. Definiáljuk az $R \subseteq Q \times Q$ relációt a következőképp:
  pRq ha $\forall\ x \in T^*$ input szóra teljesül, hogy
  $px \Rightarrow^*_A r$ akkor és csak akkor, ha $qx \Rightarrow^*_A r'$ valamely $r, r' \in F$ állapotokra.
  ($r = r'$ lehetséges).
- p és q állapotok **megkülönböztethetők**, ha
  $\exists\ x \in T^*$, amelyre vagy $px \Rightarrow^*_A r$, $r \in F$, vagy $qx \Rightarrow^*_A r'$, $r' \in F$,
  de mindkét redukció nem lehetséges.
  Különben, p és q **nem megkülönböztethető**.
- Ha p és q nem megkülönböztethető, akkor $\delta(p, a) = s$ és $\delta(q, a) = t$ sem megkülönböztethető egyetlen $a \in T$-re sem.
- Ha $\delta(p, a) = s$ és $\delta(q, a) = t$ megkülönböztethető $x \in T^*$-re, akkor p és q megkülönböztethetők ax-re.

Minimális állapotszámú DVA

- Legyen $A = (Q, T, \delta, q_0, F)$ egy DVA. A q állapot **elérhető** a kezdőállapotból, ha létezik egy redukció $q_0x \Rightarrow^* q$, ahol x egy szó T felett.
- A DVA $A = (Q, T, \delta, q_0, F)$ **összefüggő**, ha minden állapota elérhető a kezdő állapotból.
- Legyen ***H*** a q_0-ból **elérhető állapotok halmaza**. H-t a következőképpen határozzuk meg:
  Legyen $H_0 = \{q_0\}$, $H_{i+1} = H_i \cup \{r \mid \delta(q, a) = r , q \in H_i , a \in T\}$, $i = 1, 2, \ldots.$

  Ekkor $\exists\, k \geq 0 : H_k = H_l$, minden $l \geq k$. Legyen $H = H_k$.
- Definiáljuk az $A' = (Q', T, \delta', q_0, F')$ DVA-t következőképpen:
  $Q' = H$, $F' = F \cap H$ és $\delta' : H \times T \rightarrow H$ úgy hogy $\delta'(q, a) = \delta(q, a)$,
  ha $q \in H$.
- Könnyen látható, hogy A' is összefüggő és ugyanazt a nyelvet fogadja el, mint A. A' a **legnagyobb összefüggő részautomatája** A-nak.

Minimális állapotszámú DVA

Computing_Reachable_States
(from: https://en.wikipedia.org/wiki/DFA_minimization)

- let reachable_states := {q0}
- let new_states := {q0}
- do {
-     temp := the empty set
-     for each q in new_states do
-         for each c in T do
-             temp := temp ∪ {p such that p = δ(q,c)}
-     new_states := temp \ reachable_states
-     reachable_states := reachable_states ∪ new_states
- } while (new_states ≠ the empty set)
- unreachable_states := Q \ reachable_states

Minimális állapotszámú DVA

- Minimális DVA kiszámítása
  (Hopcroft partíció finomítás algoritmusa):
  - Meghatározzuk, hogy a DVA összefüggő, vagy nem.
    - Ha nem összefüggő akkor a legnagyobb összefüggő részautomatáját tekintjük. A továbbiakban feltesszük, hogy a DVA összefüggő.
  - Ezután partícionáljuk az állapotok halmazát (megkülönböztethetőség szerint ekvivalenciaosztályokra bontjuk)
    (1.-3. lépés)

Minimális állapotszámú DVA

- **1. lépés**:
  - Az állapotok halmazát két partícióra osztjuk: F-re és $Q - F$-re. (Az F-beli állapotok megkülönböztethetők a $Q - F$ -beli állapotoktól az üres szóval).
  - Ismételjük a partíciók további partíciókra való bontását (2. lépés) mindaddig, amíg a partíciók száma változatlan marad.
- **2. lépés**:
  - Tekintsük egy tetszőleges P partíciót. Vegyük az a szimbólumot és tekintsük a $\delta(p, a)$ állapotot minden $p \in P$ állapotra.
    Ha az így kapott állapotok különböző partíciókhoz tartoznak, akkor az eredeti partíciót annyi új partícióra osztjuk, ahány ilyen módon keletkezik.
  - Ismételjük ezt minden input szimbólumra és minden partícióra, amíg új partíció már nem keletkezik.

Minimális állapotszámú DVA

- **3. lépés**:
  - Meghatározzuk a minimális DVA-t:
  - Minden B_i partícióra tekintünk egy b_i reprezentáns állapotot.
  - Konstruálunk egy $A' = (Q', T, \delta', q_0, F')$ DVA-t, ahol
    - Q' a partíciók reprezentánsainak a halmaza,
    - q'_0 annak a partíciónak a reprezentánsa, amely tartalmazza q_0-t,
    - $\delta'(b_i, a) = b_j$, ha $\exists\ q_i \in B_i$ és $q_j \in B_j$, amelyre $\delta(q_i, a) = q_j$.
    - $F' = \{b_f\}$ annak a partíciónak a reprezentánsa, amely tartalmazza F elemeit.

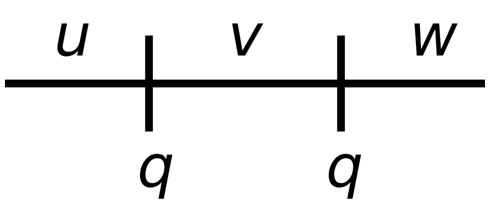
Pumping lemma reguláris nyelvekre

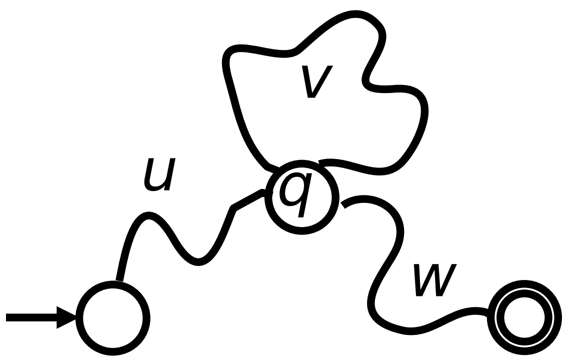
- Szükséges feltétel, hogy egy nyelv reguláris legyen (azaz egy VA-val felismerhető).

- **Tétel** (pumping lemma reguláris nyelvekre): Minden L reguláris nyelvre létezik egy n természetes szám, úgy hogy minden $z \in L$ szóra, amely hossza $|z|>n$, érvényes, hogy z felírható mint $z=uvw$, amire a következő feltételek teljesülnek:
  1. $|uv| \leq n$,
  2. $|v|>0$,
  3. $uv^iw \in L$, minden $i \geq 0$-ra.

Pumping lemma reguláris nyelvekre

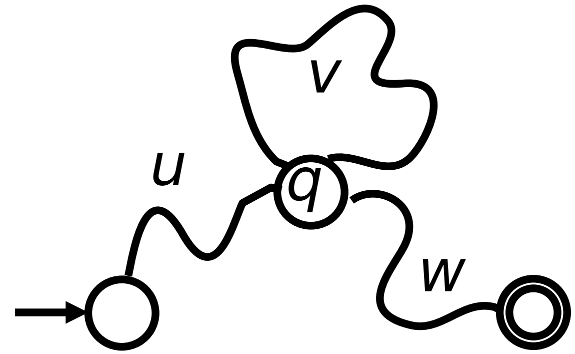
- **Biz.**:
  - Legyen L egy reguláris nyelv és $A=(Q, T, \delta, q_0, F)$ egy minimális DVA, ú.h. $L(A)=L$.
  - Legyen $n=|Q|+1$. Legyen $z \in L$ egy tetszőleges szó, amelyre $|z|>n$.
  - Tekintsük A-t z inputtal. Kell lennie egy q állapotnak, amelyet A legalább kétszer meglátogat z feldolgozása során. Már az első n állapot átmenet során is léteznie kell egy ilyen q állapotnak.
  - Legyen u a z prefixe, amelyet A feldolgoz q első előfordulásáig, v pedig z részszava, amelyet q első és második előfordulása között dolgoz fel. Ekkor $|uv| \leq n$.

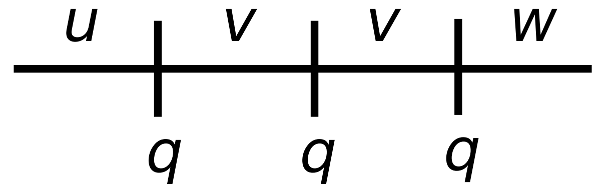




Pumping lemma reguláris nyelvekre

- **Biz.** (folyt.):
  - Mivel legalább egy állapotátmenet történt A-ban q két előfordulása között, azaz A legalább egy szimbólumot olvasott, $|v|>0$.
  - Ha A a q állapotból indul és a w szót olvassa, akkor elfogadó végállapotba kerül. Ennek megfelelően az uw-t elfogadja A.
  - Hasonlóképpen, A elfogad minden uv^iw, $i>0$ alakú szót is, mivel u olvasása után A q állapotba kerül, q-ból indulva v^i olvasása után A ismét q állapotba kerül, végül w olvasása után A egy elfogadó végállapotba jut. Ezzel teljes a bizonyítás. □

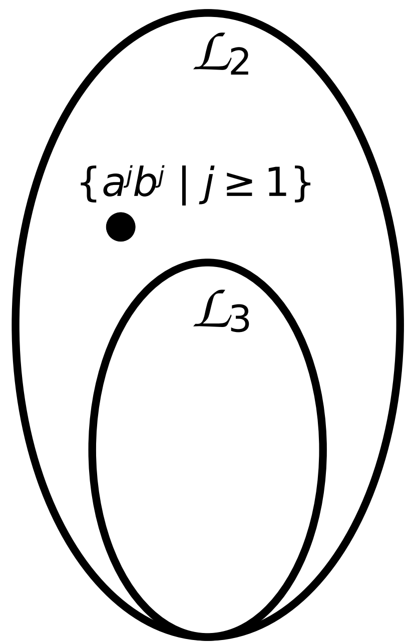




Pumping lemma alkalmazása

Állítás: A $L=\{a^j b^j \mid j \geq 1\}$ nyelv nem reguláris.

Biz.: TF. hogy G egy L-t generáló reguláris grammatika.
Ekkor a reguláris pumping lemma szerint $\exists\, n \geq 0$, úgy hogy minden $z \in L$ szóra, ahol $|z|>n$, érvényes, hogy z felírható $z=uvw$-ként, $|uv| \leq n$, $|v|>0$, továbbá $uv^iw \in L$, minden $i \geq 0$-ra.
Tekintsünk egy $a^m b^m$ szót, ahol $m>n$.
Mivel $|uv| \leq n$, Az uv csak a szimbólumot tartalmazhat. Mivel $|v|>0$, $i \geq 2$ esetén uv^iw több a-t tartalmaz, mint b-t.
Ezért $uv^iw \notin L$. $\square$



Egy környezet-független grammatika, ami L-t generálja: $S \rightarrow ab$, $S \rightarrow aSb$.

Reguláris grammatika átalakítása ekvivalens VA-ra

1) Konstruáljunk egy *G*-vel ekv. ε-mentes *G'* reguláris grammatikát (l. köv. fólia);

2) Hozzunk létre egy *M* VA-t, amely *G'* minden nemterminálisához tartalmaz egy állapotot. Legyen a kezdőállapot az az állapot, amely a *G'* kezdőszimbólumát *S'*-t reprezentálja;

3) Adjunk hozzá egy új *F* állapotot, ami végállapot;

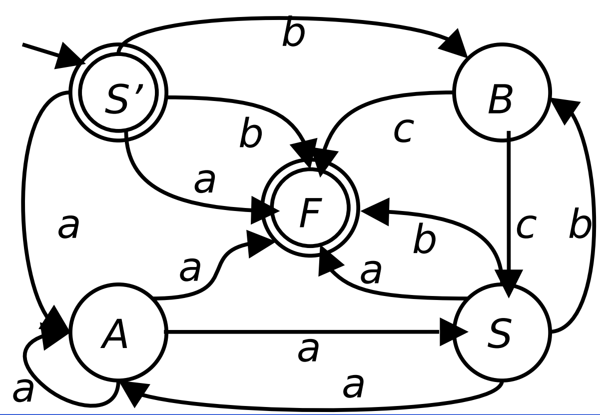
4) Ha $S' \to \varepsilon$ szabályt tartalmazza *G'*,
   - legyen az set *S'*-t reprezentáló állapot végállapot;

5) Minden $A \to aB$ szabályra *G'*-ben,
   - adjunk hozzá *M*-hez egy átmenetet *A*-ból *B*-be *a* cimkével;

6) Minden $A \to a$ szabályra *G'*-ben,
   - Adjunk hozzá *M*-hez egy átmenetet *A*-ból *F*-be *a* cimkével.

Példa:

- *G*:
  - $S \to a|aA|bB|\varepsilon$
  - $A \to aA|aS$
  - $B \to cS|\varepsilon$
- *G'*:
  - $S' \to a|b|aA|bB|\varepsilon$
  - $S \to a|b|aA|bB$
  - $A \to a|aA|aS$
  - $B \to c|cS$



Reguláris grammatika ε-mentessé alakítása

Egy *G* reguláris grammatika ε-mentes nincs ε-szabály kivéve $S \to \varepsilon$, ahol *S* a kezdőszimbólum és *S* nem jelenik meg a szabályok jobb oldalán.

G reguláris grammatika ε-mentessé alakítása:

1) Másoljuk az összes nem ε-szabályt *G*-ből to *G'*-be. Legyen *S* a kezdőszimbólum *G'*-ben;

2) Minden *N* nemterminálisra, ami ε-ná válhat (While $\exists N : N \to \varepsilon$ do),
   - másoljunk minden szabályt, amelyben *N* megjelenik a jobb oldalon *N*-nel és anélkül is;

3) Ha $S \to \varepsilon$ az eredeti szabályok között van,
   - adjunk hozzá egy új *S'* kezdőszimbólumot *G'*-ben,
   - adjuk hozzá a $S' \to \varepsilon$ szabályt
   - másoljuk az összes szabályt, ahol *S* van a bal oldalon, azokhoz, ahol *S'* van a bal oldalon.

Példa:

- *G*:
  - $S \to aA \mid bB \mid \varepsilon$
  - $A \to aA \mid a \mid \varepsilon$
  - $B \to bB \mid b \mid \varepsilon$
- 1)
  - $S \to aA \mid bB$
  - $A \to aA \mid a$
  - $B \to bB \mid b$
- 2)
  - $S \to aA \mid bB \mid a \mid b$
  - $A \to aA \mid a$
  - $B \to bB \mid b$
- 3)
  - $S' \to aA \mid bB \mid a \mid b \mid \varepsilon$
  - $S \to aA \mid bB \mid a \mid b$
  - $A \to aA \mid a$
  - $B \to bB \mid b$

VA átalakítása reguláris grammatikává

VA A átalakítása G reguláris grammatikává:

1) Legyen T a terminálisok ábécéje G-ben – ugyanaz nint A-ban.

2) A nemterminálisok halmaza G-ben legyen az állapotok halmaza Q az A-ban .

3) A kezdőszimbólum S a G-ben legyen az A kezdőállapota.

4) Kezdetben G-ben a szabályok halmaza legyen $\emptyset$. Minden $(q,a)\to q'$ átmenethez az A-ban,
   - a) adjuk hozzá a $q\to aq'$ szabályt G-hez;
   - b) ha q' végállapot, adjuk hozzá a $q\to a$ szabályt is.

5) Ha S egy végállapot A-ban, adjuk hozzá G-hez a $S\to\varepsilon$ szabályt.

6) Ha G nem ε-mentes, alakítsuk azzá. (l. előző fólia)

Példa:

- $A=(Q,T,\delta,S,\{S,C\})$ with δ:
  - $(S,a)\to A$,
  - $(S,b)\to B$,
  - $(A,a)\to B$,
  - $(A,a)\to C$,
  - $(B,b)\to A$,
  - $(B,b)\to C$,
  - $(C,c)\to C$.
- 4)
  - $S\to aA \mid bB$,
  - $A\to aB \mid aC \mid a$,
  - $B\to bA \mid bC \mid b$,
  - $C\to cC \mid c$.
- 5)
  - $S\to aA \mid bB \mid \varepsilon$
  - $A\to aB \mid aC \mid a$
  - $B\to bA \mid bC \mid b$
  - $C\to cC \mid c$.

VA átalakítása reguláris kifejezéssé

Ötlet: TF. az *A* VA állapotai: 1,…,n, kezdő állapot: 1.
Kiszámítjuk a $T(i,j,k)$ reguláris kifejezéseket, amelyek leírják az összes sztringet, amely az *i* állapotból a *j*-be visz az {1,2,…,k} állapotokon keresztül. Az $L(A)$ nyelv az összes olyan sztring uniója, amely az 1-es állapotból egy $f \in F$ végállapotba visz bármelyik állapotokon keresztül:
$L(A) = \text{U}_{f\in F}\, T(1,f,n)$.

$T(i,j,k)$ kiszámítása:

1) $k=0$:

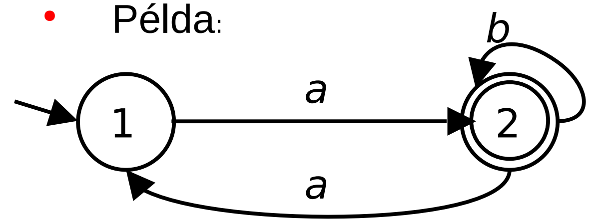
a) Ha $i=j$: $T(i,i,0) = \varepsilon+a+\ldots+z$ ahol $a,\ldots,z$ az *i* állapotból önmagába vezető átmenetek címkéi.
Ha nincsenek ilyen átmenetek, $T(i,i,0) = \varepsilon$.

b) Ha $i\neq j$: $T(i,j,0) = a+\ldots+z$ ahol $a,\ldots,z$ az *i* állapotból *j*-be vezető átmenetek címkéi.
Ha nincsenek ilyen átmenetek, $T(i,j,0) = \emptyset$.

2) $k > 0$:

$$T(i,j,k) = T(i,j,k\text{-}1) + T(i,k,k\text{-}1)(T(k,k,k\text{-}1)^*T(k,j,k\text{-}1)$$

- Példa:



- $T(1,1,0) = \varepsilon$
  $T(2,2,0) = \varepsilon+b$
  $T(1,2,0) = a$
  $T(2,1,0) = a$

- $T(1,1,1) = \varepsilon+\varepsilon(\varepsilon)^*\varepsilon = \varepsilon$
  $T(2,2,1) = \varepsilon+b+a(\varepsilon)^*a =$
  $\varepsilon+b+aa$
  $T(1,2,1) = a+\varepsilon(\varepsilon)^*a = a$
  $T(2,1,1) = a+a(\varepsilon)^*\varepsilon = a$

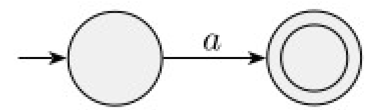
- $T(1,1,2) = \ldots$
  $T(2,2,2) = \ldots$
  $\underline{T(1,2,2)} =$
  $a+a(\varepsilon+b+aa)^*(\varepsilon+b+aa) =$
  $a+a(\varepsilon+b+aa)^+ =$
  $a+a(b+aa)^* =$
  $\underline{a(b+aa)^*}$
  $T(2,1,2) = \ldots$

Reguláris kifejezés átalakítása NVA-ra

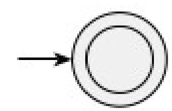
R Reguláris kifejezés átalakítása *N* NVA-ra:

NVA ami elfogadja $L(R)$-t

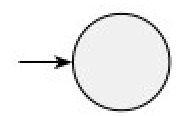
1. Ha $R = a$, $a \in T$, akkor $L(R) = \{a\}$



2. Ha $R = \varepsilon$, akkor $L(R) = \{\varepsilon\}$



3. Ha $R = \emptyset$, akkor $L(R) = \emptyset$



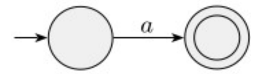
4. $R = R_1 \cup R_2$

5. $R = R_1 \cdot R_2$

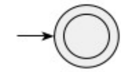
6. $R = R_1{}^*$

Reguláris kifejezés átalakítása NVA-ra

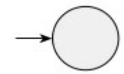
1. Ha $R = a$, $a \in T$, akkor $L(R) = \{a\}$



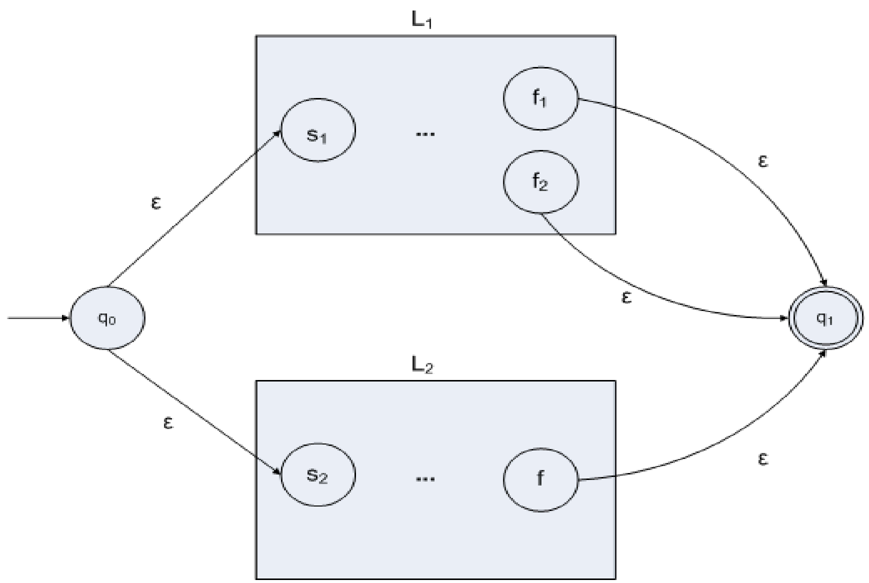
2. Ha $R = \varepsilon$, akkor $L(R) = \{\varepsilon\}$



3. Ha $R = \emptyset$, akkor $L(R) = \emptyset$



4. $R = R_1 \cup R_2$

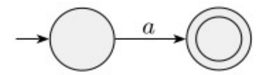


5. $R = R_1 \cdot R_2$

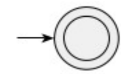
6. $R = R_1*$

Reguláris kifejezés átalakítása NVA-ra

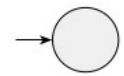
1. Ha $R = a$, $a \in T$, akkor $L(R) = \{a\}$



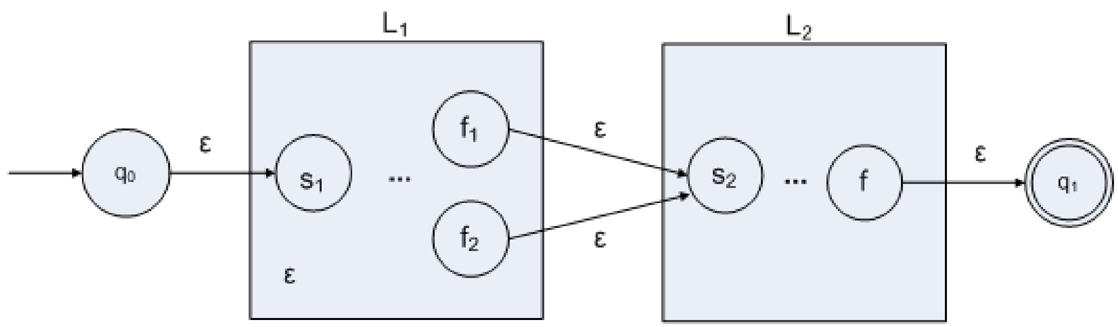
2. Ha $R = \varepsilon$, akkor $L(R) = \{\varepsilon\}$



3. Ha $R = \emptyset$, akkor $L(R) = \emptyset$



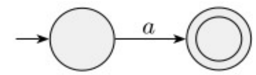
4. $R = R_1 \cup R_2$
5. $R = R_1 \cdot R_2$



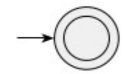
6. $R = R_1{}^*$

Reguláris kifejezés átalakítása NVA-ra

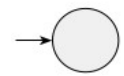
1. Ha $R = a$, $a \in T$, akkor $L(R) = \{a\}$



2. Ha $R = \varepsilon$, akkor $L(R) = \{\varepsilon\}$



3. Ha $R = \emptyset$, akkor $L(R) = \emptyset$



4. $R = R_1 \cup R_2$
5. $R = R_1 \cdot R_2$
6. $R = R_1{}^*$