

Szavak és nyelvek

Ábécé: tetszőleges véges, nem üres halmaz

Betűk: tetszőleges ábécé elemei

Legyen Σ egy ábécé

Σ -feletti **szó:** Σ -beli betűk véges (akár 0 hosszú) sorozata

Egy w szó **hossza:** a w -ben szereplő betűk száma, jele: $|w|$

$|w|_a$: w -ben az a -k száma

Üres szó: a 0 hosszú szó, jele: ε

Σ^* : Az összes Σ -feletti szó halmaza

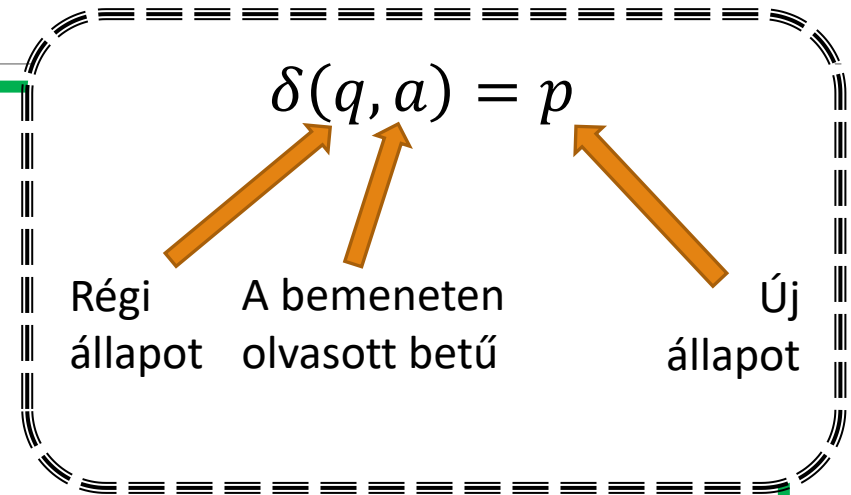
Σ -feletti **nyelv:** Σ^* egy részhalmaza

$\Sigma = \{0,1\}$, $\Sigma^* = \{\varepsilon, 0, 1, 00, 01, 10, 11, 000, \dots\}$
0 és 1 itt most betűk, nem számok

- $w = 01001$, $|w| = 5$, $|w|_0 = 3$
- $w = 000 = 0^3$, $|w| = 3$, $|w|_1 = 0$
- $\{0,1\}$ -feletti nyelvek:
 - Véges nyelvek:
 - $\{1, 01, 001\}$, $\{\varepsilon\}$, $\{\varepsilon, 1\}$, \emptyset
 - Végtelen nyelvek:
 - Σ^* , $\{0^n 1^m \mid n, m \geq 0\}$
 - $\{0^n 1^n \mid n \geq 0\}$
 - $\{w \in \Sigma^* \mid |w|_0 = |w|_1\}$
 - $\{0^p \mid p \text{ prímszám}\}$

Véges automaták

- **Véges automata:** $(Q, \Sigma, \delta, q_0, F)$
 - Q : **állapotok** véges, nemüres halmaza
 - Σ : **bemenő jelek (betűk)** ábécéje
 - q_0 : **kezdőállapot**
 - $F \subseteq Q$: **végállapotok**
 - $\delta: Q \times \Sigma \rightarrow Q$, az **átmeneti függvény**
- Ha $\delta(q, a) = p$, akkor $q \xrightarrow{a} p$ az M egy **átmenete**
- Az átmenetek alapján a δ egyértelműen leírható egy irányított címkézett gráffal, az **átmeneti diagrammal**:
 - A csúcsok az állapotok, és két csúcs közte egy betűvel címkézett éllel megfelel egy átmenetnek
- **M megadása** átmeneti diagrammal: megjelöljük a kezdő- és végállapotokat



Véges automaták

- Legyen $M = (Q, \Sigma, \delta, q_0, F)$ egy véges automata, $q \in Q$, $w = a_1 \dots a_n$ és $n \geq 0$


 Σ -beli betűk

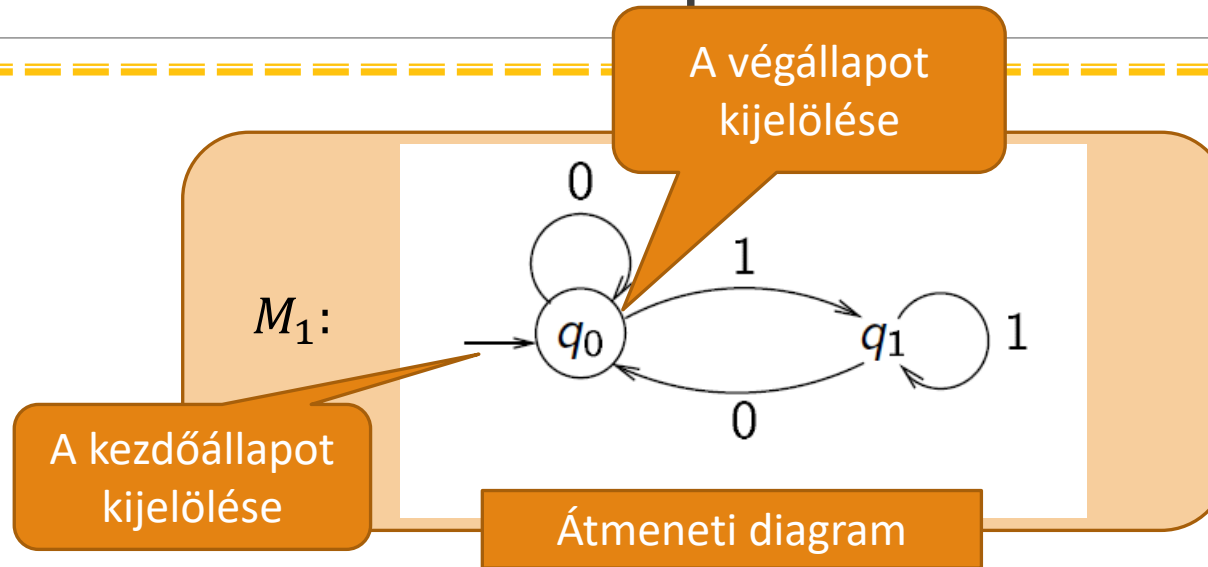
- Megy q -ból induló futása a w -n:** átmenetek egy olyan

$$q_1 \xrightarrow{a_1} q_2 \xrightarrow{a_2} q_3 \dots q_n \xrightarrow{a_n} q_{n+1} \text{ sorozata, ahol } q_1 = q$$

- Ha $n = 0$ (azaz ha $w = \varepsilon$), akkor $q_{n+1} = q_1$ és a futás 0 átmenetből áll,
- Ez a futás **sikeres**, ha $q_{n+1} \in F$
- M **elfogadja** w -t: M -nek van q_0 -ból induló sikeres futása a w -n
- Az M által **felismert nyelv**:

$$L(M) = \{w \in \Sigma^* \mid M \text{ elfogadja } w\text{-t}\}$$

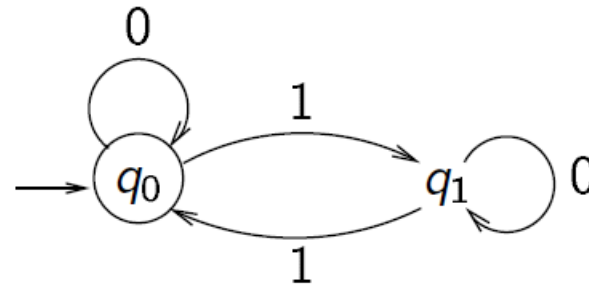
Véges automaták – 1. példa



- $Q = \{q_0, q_1\}$
- $\Sigma = \{0,1\}$
- Kezdőállapot: q_0
- $F = \{q_1\}$
- q_0 -ból induló sikeres futás a 0110 szón: $q_0 \xrightarrow{0} q_0 \xrightarrow{1} q_1 \xrightarrow{1} q_1 \xrightarrow{0} q_0$
- $L(M_1) = \{w \in \{0,1\}^* \mid w = \varepsilon \text{ vagy } w \text{ utolsó betűje } 0\}$

Véges automaták – 2. példa

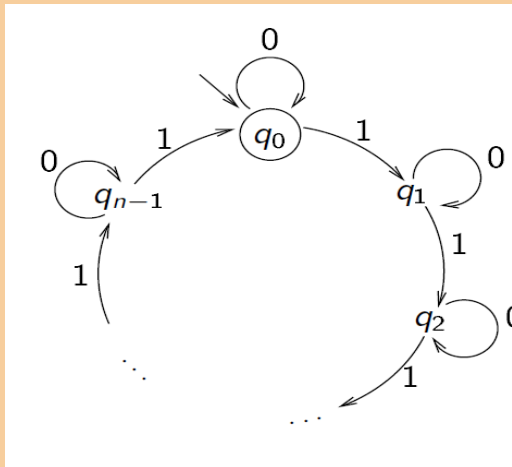
M_2 :



- $Q = \{q_0, q_1\}$
- $\Sigma = \{0,1\}$
- Kezdőállapot: q_0
- $F = \{q_0\}$
- q_0 -ból induló sikeres futás a 0110 szón: $q_0 \xrightarrow{0} q_0 \xrightarrow{1} q_1 \xrightarrow{0} q_1 \xrightarrow{1} q_0$
- $L(M_2) = \{w \in \{0,1\}^* \mid w\text{-ben az 1-esek száma páros}\}$

Véges automaták – 3. példa

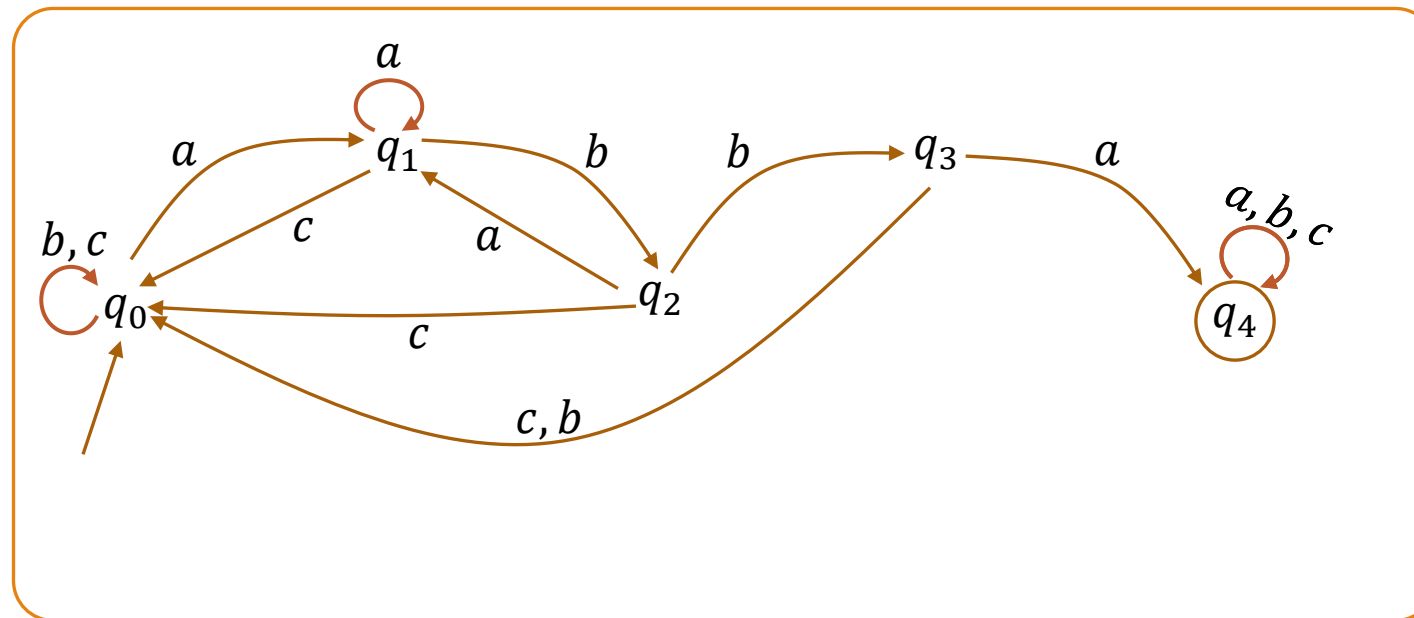
M_3 :



- $Q = \{q_0, q_1, \dots, q_{n-1}\}$
- $\Sigma = \{0,1\}$
- Kezdőállapot: q_0
- $F = \{q_0\}$
- $L(M_3) = \{w \in \{0,1\}^* \mid w\text{-ben az } 1\text{-esek száma } n\text{-nel osztható}\}$

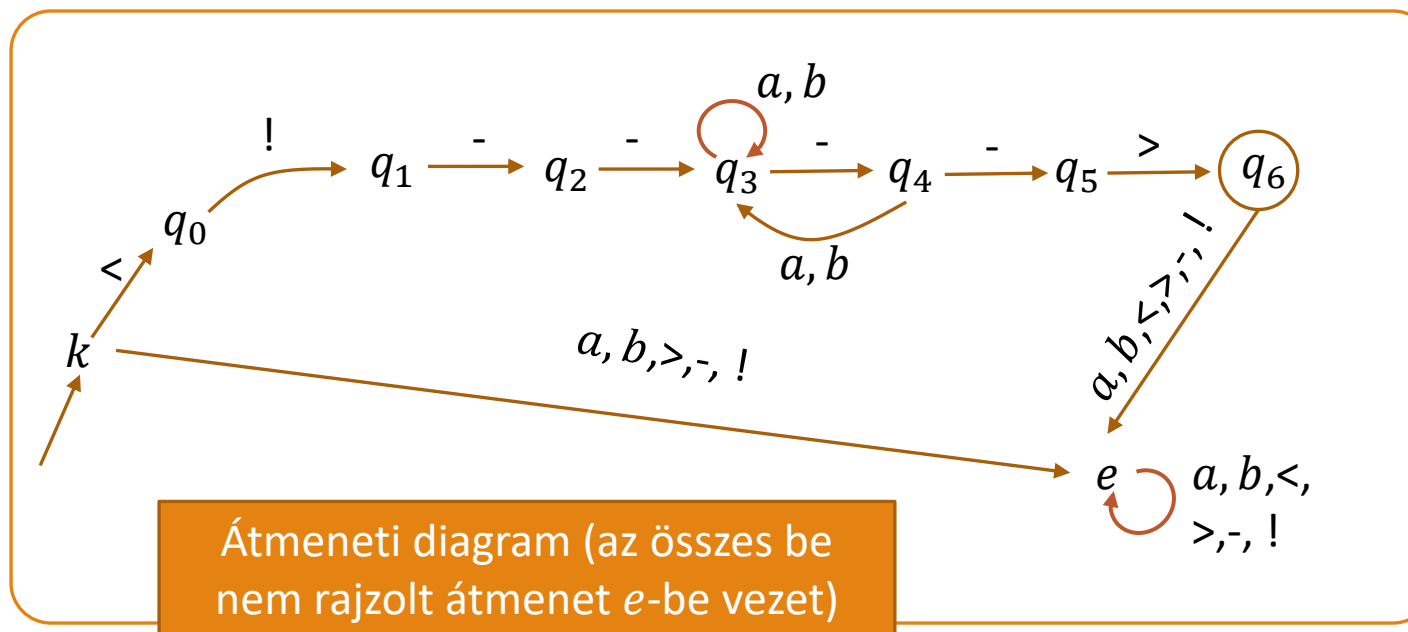
Véges automata mint mintaillesztő – Példa

- Ez a véges automata azt vizsgálja, hogy a bemenetként kapott szóban van-e *abba* részszó
 - Azaz olyan szavakra kerül végállapotba, melyekben az *abba* részszó
 - Az egyszerűség kedvéért feltesszük, hogy a bemenetben legfeljebb csak *a*, *b* és *c* betűk vannak



Véges automata mint lexikális elemző – Példa

- Ez a véges automata pedig azt vizsgálja, hogy a bemenetként kapott szó szintaxisa megfelel-e annak amit az XML-ben a kommentekre előírnak:
 - <!-- az eleje és --> a vége, és
 - a kommentben nem fordul elő két kötőjel egymás után
 - Az egyszerűség kedvéért feltesszük, hogy a kommentekben csak az *a* és *b* betűk, valamint a kötőjel (-) szerepelnek



Felismerhető nyelvek

Egy L nyelvet **felismerhetőnek** nevezünk ha van olyan M véges automata, ami felismeri L -et

Hogyan jellemezhetők a **felismerhető nyelvek**?

- A szavaik olyan (a nyelvtől függő) korlátozott számú egyszerű tulajdonsággal rendelkeznek, melyeket egy VA képes „megjegyezni”.

Legyen $\Sigma = \{0,1\}$, akkor a következő nyelvek felismerhetők:

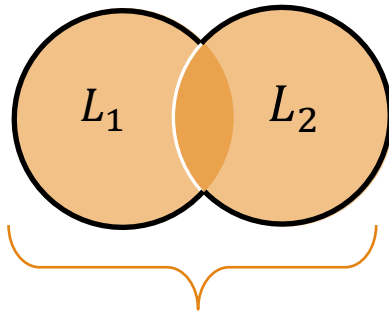
- A páros hosszú Σ -feletti szavak halmaza (a VA két tulajdonságot tart számon: páros vagy páratlan az épp elolvasott részszó)
- Az 1-esre végződő Σ -feletti szavak halmaza (a VA azt tartja számon, hogy az utoljára elolvasott betű 1-es vagy 0)
- Olyan Σ -feletti szavak halmaza, melyekben nem jelenik meg sem a 00 és sem az 11 részszóként (a VA azt tartja számon, hogy az utoljára elolvasott betű milyen)

Egy **nem felismerhető nyelv**: Olyan Σ -feletti szavak halmaza, melyekben ugyanannyi 1 és 0 van (azt kellene számon tartani, hogy a kétféle betű számának különbsége mennyi; ez nem megy VA-val)

Műveletek nyelveken

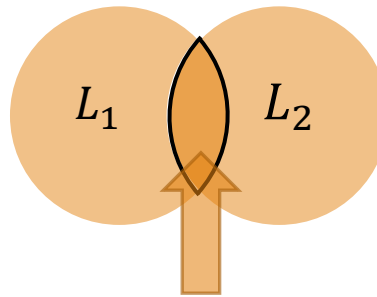
Halmazelméleti műveletek ($L_1, L_2 \subseteq \Sigma^*$):

Egyesítés



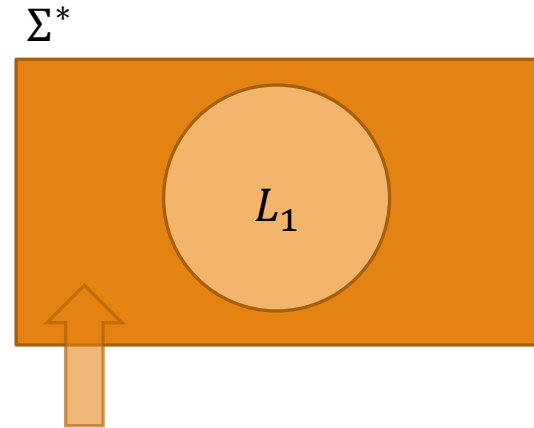
$$L_1 \cup L_2 = \{w \mid w \in L_1 \text{ vagy } w \in L_2\}$$

Metszet



$$L_1 \cap L_2 = \{w \mid w \in L_1 \text{ és } w \in L_2\}$$

Komplementer

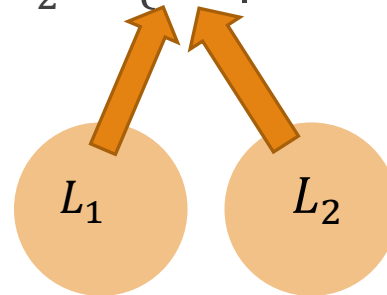


$$\overline{L_1} = \{w \in \Sigma^* \mid w \notin L_1\}$$

Műveletek nyelveken

Egyéb műveletek ($u, v \in \Sigma^*, L_1, L_2 \subseteq \Sigma^*$)

- u és v **konkatenációja**: $u \cdot v$ (u és v összefűzése, egymás után írása)
- $u \cdot v$ helyett általában uv -t fogunk írni
- Konkatenáció **kiterjesztése nyelvekre**: $L_1 \cdot L_2 = \{uv \mid u \in L_1, v \in L_2\}$



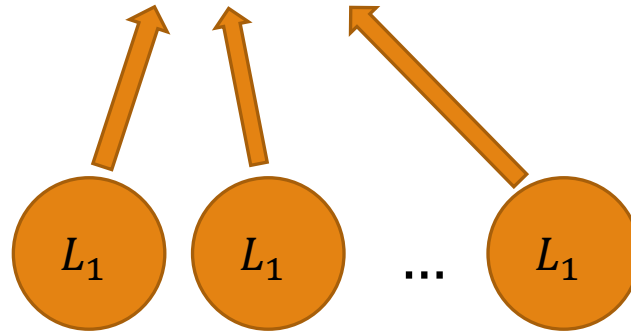
- $L_1 \cdot L_2$ helyett általában L_1L_2 -t fogunk írni

Pl.

$$\{01,10\} \cdot \{00,11\} = \{0100,0111,1000,1011\}$$

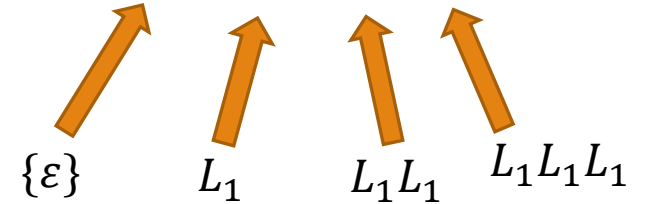
Műveletek nyelveken

- (Kleene) **iteráció**: $L_1^* = \{ u_1 u_2 \dots u_n \mid n \geq 0, u_1 u_2, \dots, u_n \in L_1 \}$



- Reguláris műveletek**: unió, konkatenáció, iteráció

$$L_1^* = L_1^0 \cup L_1^1 \cup L_1^2 \cup L_1^3 \dots$$



- $\{01\}^* = \{w \in \{0,1\}^* \mid w = \varepsilon \text{ vagy } w \text{ első betűje 0, utolsó 1 és } 00, 11 \text{ nem rész-szavak } w\text{-ben}\}$
- $\{1, \varepsilon\}\{01\}^*\{0, \varepsilon\} = \{w \mid w\text{-ben } 00 \text{ és } 11 \text{ nem rész-szavak}\}$

Néhány azonosság

Az egyesítés és a konkatenáció **asszociatív** (tetszőlegesen zárójelezhető), az egyesítés **kommutatív** (felcserélhető) is (a konkatenáció nyilván nem)

A konkatenáció **disztributív** az egyesítés felett:

$$\circ L_1 \cdot (L_2 \cup L_3) = (L_1 \cdot L_2) \cup (L_1 \cdot L_3) \text{ és } (L_1 \cup L_2) \cdot L_3 = (L_1 \cdot L_3) \cup (L_2 \cdot L_3)$$

További azonosságok:

- $\emptyset^* = \{\varepsilon\}$
- $L \cdot \{\varepsilon\} = \{\varepsilon\} \cdot L = L$
- $L \cdot \emptyset = \emptyset \cdot L = \emptyset$

Jelölés:

$$\circ L^+ = L \cdot L^* = L^* \cdot L = L_1^1 \cup L_1^2 \cup L_1^3 \dots$$

Tehát:

$$\circ L^* = L^+ \cup \{\varepsilon\}$$

Ezek pedig **NEM** azonosságok:

- $(L_1 \cup L_2)^* \neq L_1^* \cup L_2^*$
- (pl. $L_1 = \{a\}, L_2 = \{b\}$),
- $(L_1 \cap L_2)^* \neq L_1^* \cap L_2^*$
- (pl. $L_1 = \{a\}, L_2 = \{aa\}$)
- $L_1 \cup (L_2 \cdot L_3) \neq (L_1 \cup L_2) \cdot (L_1 \cup L_3)$
- (pl. $L_1 = \{a\}, L_2 = L_3 = \{\varepsilon\}$)

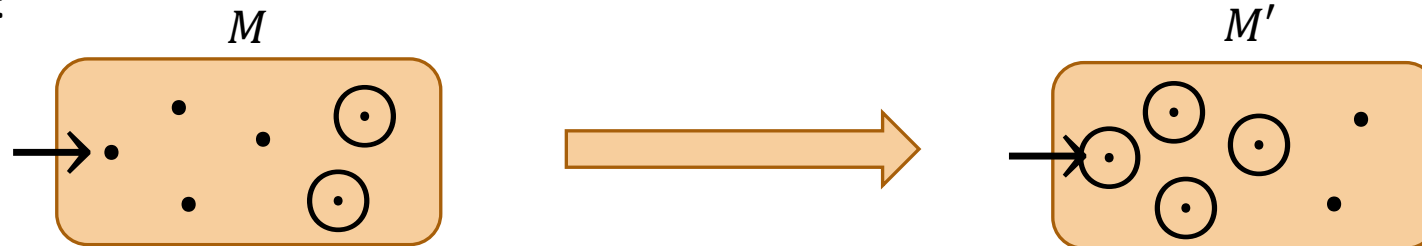
A felismerhető nyelvek zárttsági tulajdonságai

A felismerhető nyelvek zártak a halmazelméleti műveletekre

Bizonyítás

Komplementerképzésre való zárttság

- Legyen L egy felismerhető nyelv és $M = (Q, \Sigma, \delta, q_0, F)$ egy L -et felismerő véges automata
- Megadunk egy M' -t ami \bar{L} -et ismeri fel
- Ötlet:



- Legyen tehát $M' = (Q, \Sigma, \delta, q_0, Q - F)$
- Ekkor tetszőleges $w \in \Sigma^*$ szóra M -nek pontosan akkor **van** sikeres futása w -n, ha M' -nek **nincs** sikeres futása w -n

A felismerhető nyelvek zártsági tulajdonságai

Bizonyítás

Egyesítésre és metszetre való zártság

- Legyen L_i ($i = 1, 2$) felismerhető nyelv és $M_i = (Q_i, \Sigma, \delta_i, q_i, F_i)$ egy L_i -t felismerő véges automata
- Ötlet: megadunk egy M' -t, ami egyszerre szimulálja M_1 és M_2 számításait egy tetsz. $a_1 a_2 \dots a_n$ szón

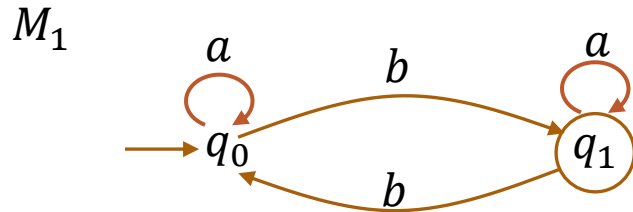
M_1 egy futása: $s_1 \xrightarrow{a_1} s_2 \xrightarrow{a_2} s_3 \dots s_n \xrightarrow{a_n} s_{n+1}, \quad s_1 = q_1$

M_2 egy futása: $r_1 \xrightarrow{a_1} r_2 \xrightarrow{a_2} r_3 \dots r_n \xrightarrow{a_n} r_{n+1}, \quad r_1 = q_2$

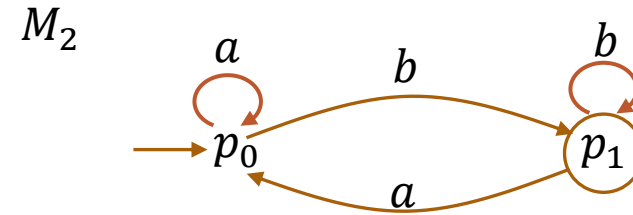
M' egy futása: $(s_1, r_1) \xrightarrow{a_1} (s_2, r_2) \xrightarrow{a_2} (s_3, r_3) \dots (s_n, r_n) \xrightarrow{a_n} (s_{n+1}, r_{n+1})$

- Legyen tehát $M' = (Q_1 \times Q_2, \Sigma, \delta', (q_1, q_2), F')$,
ahol $\delta'((s, r), a) = (\delta_1(s, a), \delta_2(r, a))$, minden $(s, r) \in Q_1 \times Q_2$ és $a \in \Sigma$ esetén
- Ha $F' = F_1 \times F_2$, akkor $L(M') = L_1 \cap L_2$
- Ha $F' = F_1 \times Q_2 \cup Q_1 \times F_2$, akkor $L(M') = L_1 \cup L_2$

A metszetre való zártság - Példa

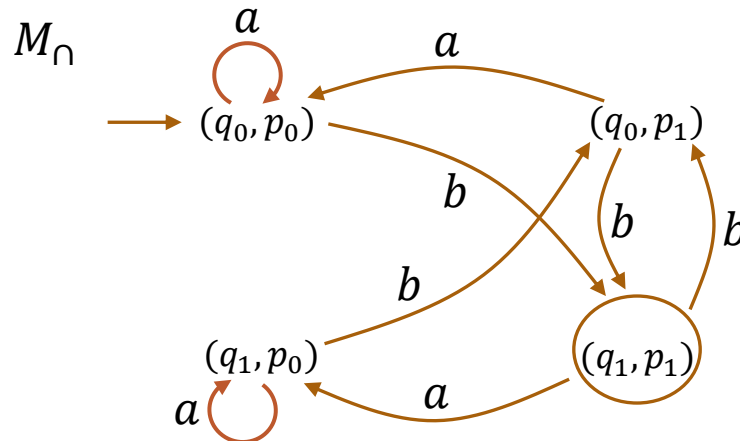


$$L(M_1) = \{u \in \{a, b\}^* \mid |u|_b \text{ páratlan}\}$$



$$L(M_2) = \{u \in \{a, b\}^* \mid u \text{ } b\text{-re végződik}\}$$

Adjunk véges automatát, ami az $L(M_1) \cap L(M_2) = \{u \in \{a, b\}^* \mid |u|_b \text{ páratlan és } u \text{ } b\text{-re végződik}\}$ nyelvet ismeri fel!



Reguláris nyelvek

Egy $L \subseteq \Sigma^*$ nyelvet **regulárisnak** nevezünk, ha előáll az

$$\emptyset \text{ és } \{a\} \quad (a \in \Sigma)$$

nyelvekből a három reguláris művelet:

- egyesítés,
- konkatenáció és
- Kleene-iterált

véges sokszori alkalmazásával

Például a $(\{1\} \cup \emptyset^*)(\{0\}\{1\})^*(\{0\} \cup \emptyset^*) = \{1, \varepsilon\}\{01\}^*\{0, \varepsilon\}$ nyelv reguláris

- Korábban már volt ez a nyelv: azon szavak halmaza melyekben nincs 00 vagy 11 részszo

Reguláris kifejezések

A reguláris nyelvek jelölésére alkalmasak a **reguláris kifejezések**

- **Különbség:** nem használjuk a $\{, \}$ zárójeleket
- az egyesítés (\cup) helyett plusz ($+$) jelet használunk

A műveletek sorrendjét itt is a $(,)$ zárójelekkel adjuk meg

- A felesleges zárójeleket elhagyjuk a $* > \cdot > +$ **prioritási** sorrend alapján
- A \cdot jelet általában elhagyjuk
- **Rövidítés:** a \emptyset^* helyett ε -t írunk

Ha R egy reguláris kifejezés, akkor az R által jelölt nyelvet $L(R)$ -rel jelöljük, és R felépítése szerinti indukcióval a következőképpen definiáljuk:

- Ha $R = a$, akkor $L(R) = \{a\}$
- Ha $R = \emptyset$, akkor $L(R) = \emptyset$
- Ha $R = R_1 R_2$, akkor $L(R) = L(R_1) L(R_2)$
- Ha $R = R_1 + R_2$, akkor $L(R) = L(R_1) \cup L(R_2)$
- Ha $R = R_1^*$, akkor $L(R) = L(R_1)^*$

Reguláris kifejezések - Példák

R reguláris kifejezés

$0(01 + 10)1 + \varepsilon$

$0(0 + 1)^*1$

$(00)^*$

$(1 + \varepsilon)(01)^*(0 + \varepsilon)$

$(0^*10^*1)^*0^*$

Az R által jelölt nyelv (azaz $L(R)$)

$\{\varepsilon, 0011, 0101\}$

$\{w \in \{0,1\}^* \mid w \text{ 0-val kezdődik és 1-gyel végződik}\}$

$\{w \in \{0,1\}^* \mid w \text{ páros hosszú és csak 0-t tartalmaz}\}$

$\{w \in \{0,1\}^* \mid w\text{-ben 00 és 11 nem részzavak}\}$

$\{w \in \{0,1\}^* \mid w \text{ páros sok 1-et tartalmaz}\}$

Egy nyelv akkor és csak akkor reguláris ha jelölhető reguláris kifejezéssel