Formális nyelvek és fordítóprogramok

http://www.ms.sapientia.ro/˜kasa/formalis.htm

Könyvészet (1.–7, 10. megtalálhatók a kari könyvtárban)

1. Csörnyei Zoltán, Kása Zoltán, Formális nyelvek és fordítóprogramok, Kolozsvári Egyetemi Kiadó, 2007.
   https://ms.sapientia.ro/ kasa/CsornyeiKasa.pdf

2. Csörnyei Zoltán, Fordítóprogramok, Typotex Kiadó, Budapest, 2006.

3. Csörnyei Zoltán, Fordítási algoritmusok, Erdélyi Tankönyvtanács, Kolozsvár, 2000.

4. Demetrovics János, Denev, J., Pavlov, R., A számítástudomány matematikai alapjai, Nemzeti Tankönyvkiadó, Budapest, 1999.

5. Fülöp Zoltán, Formális nyelvek és szintaktikus elemzésük, Polygon, Szeged, 1999.

6. Livovschi, L., Popovici, C. P., Georgescu, H., Ţăndăreanu, N., Bazele informaticii, Editura Didactica şi Pedagogică, Bucureşti, 1981.

7. Manna, Z., Programozáselmélet, Műszaki Könyvkiadó, Budapest, 1981.

8. Bach Iván, Formális nyelvek, Typotex, Budapest, 2005.
   http://mek.oszk.hu/05000/05099/

9. Révész György, Bevezetés a formális nyelvek elméletébe, Akadémiai Kiadó, Budapest, 1979.

10. Angol–magyar informatikai szótár,
    https://people.inf.elte.hu/plisaai/pdf/szotar.pdf

Grace HOPPER (1906–1992)

ellentengernagy, az Amerikai Egyesült Államok Haditengerészetének tisztje, matematikus, a számítástudomány egyik úttörője. Egyike volt a Harvard Mark I számítógép első programozóinak. Ő írta az első fordítóprogramot, a COBOL nyelv egyik megalkotója.

Alapfogalmak

ábécé: **szimbólumok (jelek) egy nem üres, véges halmaza**

Például $\Sigma = \{a, b, c, d, 0, 1\}$ **ábécé betűi:** $a, b, c, d, 0, 1.$

$a_1 a_2 \ldots a_n$ **szó**, $a_1, a_2, \ldots, a_n \in \Sigma$, $n \geq 0$, $n \geq 0$ **a szó hossza**

Ha $w = a_1 a_2 \ldots a_n$, **akkor** $|w| = n$.

$n = 0$, **üres szó**: ε **(vagy** λ**)**

$$\Sigma^* = \{a_1 a_2 \ldots a_n \mid n \geq 0, a_1, a_2, \ldots, a_n \in \Sigma\}$$

$\Sigma^+ = \Sigma^* \setminus \{\varepsilon\}$, $\quad \Sigma^n$

$$\Sigma^* = \Sigma^0 \cup \Sigma^1 \cup \ldots \cup \Sigma^n \cup \ldots \quad \text{és} \quad \Sigma^+ = \Sigma^1 \cup \Sigma^2 \cup \ldots \cup \Sigma^n \cup \ldots$$

Ha $u = a_1 a_2 \ldots a_m$ **és** $v = b_1 b_2 \ldots b_n$, **akkor** $u = v$, **ha**
$m = n$, **és**
$a_i = b_i$, $i = 1, 2, \ldots, n$.

Σ^***-ban:** **konkatenáció (illesztés, szorzat, összefűzés)**

$u = a_1 a_2 \ldots a_m$, $\quad v = b_1 b_2 \ldots b_n$: $\Longrightarrow$ $uv = a_1 a_2 \ldots a_m b_1 b_2 \ldots b_n$

$|uv| = |u| + |v|$

asszociatív, nem kommutatív, van egységeleme $\Longrightarrow$ Σ^* **monoid** **(egységelemes félcsoport)**

szavak hatványozása: **Ha** $u \in \Sigma^*$ **akkor** $u^0 = \varepsilon$
$u^n = u^{n-1} u$, $n \geq 1$

$u = a_1 a_2 \ldots a_n$ **szó tükörképe** $u^{-1} = a_n a_{n-1} \ldots a_1$ **(jelölések:** u^R, $\tilde{u}$**)**

$\left(u^{-1}\right)^{-1} = u$ **és** $(uv)^{-1} = v^{-1} u^{-1}$.

v **részszava** u-nak, ha $\exists p, q$ szavak úgy, hogy $u = pvq$.

Ha $pq \neq \varepsilon$, akkor v **valódi részszava** u-nak.

p az u **kezdőszelete** (prefixuma), q pedig **végszelete** (szufixuma)

$L \subseteq \Sigma^*$: Σ feletti **nyelv** (formális nyelv)

$\emptyset$ üres nyelv, $\{\varepsilon\}$ az üres szóból álló nyelv

Műveletek nyelvekkel

L, L_1, L_2 egy-egy Σ feletti nyelv

- **egyesítés:** $L_1 \cup L_2 = \{u \in \Sigma^* \mid u \in L_1 \text{ vagy } u \in L_2\}$
- **metszet:** $L_1 \cap L_2 = \{u \in \Sigma^* \mid u \in L_1 \text{ és } u \in L_2\}$
- **különbség:** $L_1 \setminus L_2 = \{u \in \Sigma^* \mid u \in L_1 \text{ és } u \notin L_2\}$
- **komplementum (v. komplementer nyelv):** $\overline{L} = \Sigma^* \setminus L$
  szokásos jelölése még $\complement L$ vagy $\complement(L)$
- **szorzat:** $L_1 L_2 = \{uv \mid u \in L_1, v \in L_2\}$
- **nyelv hatványa:** $L^0 = \{\varepsilon\}$, $\qquad L^n = L^{n-1}L,\ n \geq 1$
- **nyelv iteráltja:** $L^* = \bigcup_{i=0}^{\infty} L^i = L^0 \cup L \cup L^2 \cup \ldots \cup L^i \cup \ldots$

$$L^+ = L^* \setminus \{\varepsilon\}, \qquad L^+ = \bigcup_{i=1}^{\infty} L^i.$$

- **tükrözés:** $L^{-1} = \{u^{-1} \mid u \in L\}$

$$\left(L^{-1}\right)^{-1} = L, \qquad \left(L^{-1}\right)^n = (L^n)^{-1}$$

reguláris műveletek: egyesítés, szorzás, iteráció

Nyelvek megadása

- **Nyelvek megadása elemeik felsorolásával**

Nyelvek például:

$L_1 = \{\varepsilon, 0, 1\}$, $L_2 = \{a, aa, aaa, ab, ba, aba\}$,
$L_3 = \{\varepsilon, ab, aabb, aaabbb, aaaabbbb \ldots\}$

- **Nyelvek megadása tulajdonság segítségével**

$L_4 = \{a^n b^n \mid n = 0, 1, 2, \ldots\}$ (tehát $L_4 = L_3$)
$L_5 = \{uu^{-1} \mid u \in \Sigma^*\}$, $L_6 = \{u \in \{a, b\}^* \mid n_a(u) = n_b(u)\}$,
ahol $n_a(u)$ az u szóban levő a betűk számát jelöli.

- **Nyelvek megadása nyelvtannal**

generatív nyelvtan vagy nyelvtan (grammatika)

Generatív nyelvtan a $G = (N, T, P, S)$ rendezett négyes, ahol
- N, a változók (nemterminális jelek) ábécéje,
- T, a terminális jelek ábécéje, ahol $N \cap T = \emptyset$,
- $P \subseteq (N \cup T)^* N (N \cup T)^* \times (N \cup T)^*$ véges halmaz, vagyis P az (u, v) alakú helyettesítési szabályok vagy produkciók véges halmaza, ahol $u, v \in (N \cup T)^*$, és u tartalmaz legalább egy nemterminális jelet.
(u, v) jelölés helyett $u \to v$ is használható.
- $S \in N$ a nyelvtan kezdőszimbóluma.

Az $u \to v$, azaz (u, v) szabályban: u bal, v pedig jobb oldal.

Rövidítés:

$$u \to v_1,\ u \to v_2, \ldots, u \to v_r \quad \text{helyett} \quad u \to v_1 \mid v_2 \mid \ldots \mid v_r$$

$(N \cup T)^*$ **halmazban a közvetlen levezetés v. közvetlen deriváció reláció:**

$$u \Longrightarrow v, \qquad \textbf{ha} \quad u = p_1 p p_2, \quad v = p_1 q p_2 \quad \textbf{és} \quad (p,q) \in P.$$

Más jelek: $\vdash$ **vagy a** $\models$. $\qquad \Longrightarrow$ **helyett lehet** $\underset{G}{\Longrightarrow}$

A $\Longrightarrow$ **reláció reflexív és tranzitív lezártját** $\overset{*}{\Longrightarrow}$ **jelöli: levezetés vagy deriváció.**

$u \overset{*}{\Longrightarrow} v$, **ha** $\exists w_0, w_1, \ldots, w_n \in (N \cup T)^*,\ n \geq 0$,

$$u = w_0,, \quad w_0 \Longrightarrow w_1, \; w_1 \Longrightarrow w_2, \ldots, w_{n-1} \Longrightarrow w_n, \quad w_n = v$$

vagy

$$w_0 \Longrightarrow w_1 \Longrightarrow w_2 \Longrightarrow \ldots \Longrightarrow w_{n-1} \Longrightarrow w_n$$

$u \overset{+}{\Longrightarrow} v$, **ahol** $n \geq 1$ $\quad$ **(**$u \overset{*}{\Longrightarrow} v$ **esetében, ha** $n = 0$**, akkor** $u = v$**.)**

A $G = (N, T, P, S)$ **nyelvtan által generált nyelv:**

$$L(G) = \{u \in T^* \mid S \overset{*}{\Longrightarrow} u\}.$$

Példa. Legyen $G = (N, T, P, S)$**, ahol**
$N = \{S\}$,
$T = \{a, b\}$,
$P = \{S \to ab,\ S \to aSb\}$.

$L(G) = \{a^n b^n \mid n \geq 1\}$, **mivel**

$$S \underset{G}{\Longrightarrow} aSb \underset{G}{\Longrightarrow} a^2Sb^2 \underset{G}{\Longrightarrow} \ldots \underset{G}{\Longrightarrow} a^{n-1}Sb^{n-1} \underset{G}{\Longrightarrow} a^n b^n,$$

vagy $S \overset{*}{\underset{G}{\Longrightarrow}} a^n b^n$.

G_1 **és** G_2 **nyelvtanok ekvivalensek:** $G_1 \cong G_2$**, ha** $L(G_1) = L(G_2)$**.**

Példa. 1. Adott a következő két nyelvtan,
$G_1 = (\{S\}, \{a, b\}, \{S \to aSb,\ S \to \varepsilon\}, S)$ **és**
$G_2 = (\{S\}, \{a, b\}, \{S \to aSb,\ S \to ab\}, S)$**. Ekkor** $L(G_1) \backslash \{\varepsilon\} = L(G_2)$**.**
A két nyelvtan nem ekvivalens.

2. A következő két nyelvtan ekvivalens, mivel mindegyik az

$$\{a^n b^n c^n \mid n \geq 1\}$$

nyelvet generálja.

$G_1 = (N_1, T, P_1, S_1)$**, ahol**
$N_1 = \{S_1, X, Y\}$**,** $T = \{a, b, c\}$**,**
$P_1 = \{S_1 \to abc,\ S_1 \to aXbc,\ Xb \to bX,\ Xc \to Ybcc,\ bY \to Yb,$
$aY \to aaX,\ aY \to aa\}$**.**

$G_2 = (N_2, T, P_2, S_2)$**, ahol**
$N_2 = \{S_2, A, B, C\}$**,**
$P_2 = \{S_2 \to aS_2BC,\ S_2 \to aBC,\ CB \to BC,\ aB \to ab,\ bB \to bb,$
$bC \to bc,\ cC \to cc\}$**.**

Tétel. ***Létezik olyan formális nyelv, amelyet nem lehet nyelvtannal megadni.***

A bizonyításhoz kódoljuk a nyelvtanokat. Egy adott $G = (N, T, P, S)$ **nyelvtan esetében legyen**
$N = \{S_1, S_2, \ldots S_n\}$**,** $T = \{a_1, a_2, \ldots a_m\}$ **és** $S = S_1$**.**
A kódolás a következô:

$$S_i \textbf{ kódja } 10\underbrace{11\ldots11}_{i\textbf{-szer}}01, \qquad a_i \textbf{ kódja } 100\underbrace{11\ldots11}_{i\textbf{-szer}}001$$

A kódolásban a jeleket 000 választja el, a nyíl jele 0000, a szabályokat pedig 00000 jelsorozattal választjuk el.

Nyílván, elég csak a szabályokat kódolni. Példaként vegyük a következô nyelvtant:

$G = (\{S\}, \{a, b\}, \{S \to aSb,\ S \to ab\}, S)$.

S kódja 10101, a kódja 1001001, b kódja 10011001. A nyelvtan kódja:

$$\underbrace{10101}0000\underbrace{1001001}000\underbrace{10101}000\underbrace{10011001}00000\underbrace{10101}0000\underbrace{1001001}000\underbrace{10011001}$$

Ennek a kódolásnak az alapján a nyelvtanok sorba rendezhetôk, elôször a kód hossza szerint, majd az azonos hosszúságúak esetében ábécésorrendbe: $G_1, G_2, \ldots, G_k, \ldots$.

Tekintsük most a T ábécé fölötti nyelvek halmazát: $\mathcal{L}_T = \{L \mid L \subseteq T^*\}$. A T^* halmaz megszámlálhatóan végtelen, hisz szavai sorrendbe írhatók hosszuk szerint, azon belül pedig ábécésorrendbe. Legyen ez a sorrend $s_0, s_1, s_2, \ldots$, ahol természetesen $s_0 = \varepsilon$. Ekkor minden $L \subseteq T^*$ nyelvnek megfeleltethetünk egy $b_0, b_1, b_2, \ldots$ végtelen bináris sorozatot a következôképpen:

$$b_i = \begin{cases} 1, & \textbf{ha } s_i \in L \\ 0, & \textbf{ha } s_i \notin L \end{cases} \qquad i = 0, 1, 2, \ldots$$

Könnyű belátni, hogy az így keletkezett bináris sorozatok halmaza megszámlálhatatlanul végtelen, hisz minden sorozat elképzelhetô úgy, mint egy 1-nél kisebb pozitív valós szám kettes számrendszerbeli ábrázolása (a tizedesvesszôt az elsô számjegy elé képzeljük). Fordítva is igaz, hogy minden 1-nél kisebb pozitív szám kettes számrendszerbeli ábrázolása megfelel egy ilyen sorozatnak, ha a tizedesvesszô utáni részt vesszük. Mivel a $(0, 1)$ intervallum kontinuum számosságú, az lesz a végtelen bináris sorozatok halmaza is. Tehát $\mathcal{L}_T$ is kontinuum számosságú.

Mivel a formális nyelvek halmaza kontinuum számosságú, a nyelvtanok megszámlálhatóan végtelen halmaza és a nyelvek kontinuum számosságú halmaza között nem létezik kölcsönös és egyértelmű megfeleltetés. Tehát, létezik olyan formális nyelv, amelyet nem lehet nyelvtannal megadni.

Chomsky-féle nyelvosztályok

Noam Chomsky (Philadelphia, 1928. december 7. –) amerikai nyelvész, a MIT professzora, a generatív nyelvtan elméletének megalkotója, filozófus, politikai aktivista

Ha a helyettesítési szabályokra bizonyos megkötéseket teszünk, a következő négy nyelvtantípust különböztethetjük meg.

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

0-típusú (általános vagy mondatstruktúrájú), ha semmilyen megkötést nem teszünk a helyettesítési szabályaira.

1-típusú (környezetfüggő), ha minden szabálya $uXv \to uwv$ alakú, ahol $X \in N$, $u, v \in (N \cup T)^*$, $w \in (N \cup T)^+$. Ezenkívül megengedhető az $S \to \varepsilon$ szabály is, ha S nem szerepel szabály jobb oldalán.

2-típusú (környezetfüggetlen), ha minden szabálya $X \to w$ alakú, ahol $X \in N$, $w \in (N \cup T)^+$. Ezenkívül megengedhető az $S \to \varepsilon$ szabály is, ha S nem szerepel szabály jobb oldalán.

3-típusú (reguláris), ha szabályai $X \to aY$, $X \to a$ alakúak, ahol $a \in T$ és $X, Y \in N$. Ezenkívül megengedhető az $S \to \varepsilon$ szabály is, ha S nem szerepel szabály jobb oldalán.

Ha G i-típusú nyelvtan, akkor az $L(G)$ nyelvet szintén i-típusúnak (általánosnak, környezetfüggőnek, környezetfüggetlennek, regulárisnak) nevezzük.

Egy L nyelv i-típusú ($i = 0, 1, 2, 3$), ha létezik egy i-típusú G nyelvtan, amelyik az L nyelvet generálja, vagyis $L = L(G)$.

Jelölje $\mathcal{L}_i$ ($i = 0, 1, 2, 3$) az i-típusú nyelvek osztályát. Ekkor bizonyítható, hogy

$$\mathcal{L}_0 \supset \mathcal{L}_1 \supset \mathcal{L}_2 \supset \mathcal{L}_3.$$

A nyelvtantípusok fenti értelmezése alapján a bennfoglalás ($\supseteq$) nyilvánvaló, de a szigorú bennfoglalás ($\supset$) bizonyításra szorul.

Példák.

Környezetfüggő nyelvtan

$G_1 = (N_1, T_1, P_1, S_1)$, **ahol** $N_1 = \{S_1, A, B, C\}$, $\quad T_1 = \{a, 0, 1\}$

P_1 **elemei:**

$S_1 \to ACA$

$\underline{A}C \to \underline{A}ACA \mid \underline{A}Ba \mid \underline{A}aB$

$B \to AB \mid A$

$A \to 0 \mid 1$

Az $L(G_1)$ **nyelv olyan** uav **szavakból áll, ahol** $u, v \in \{0,1\}^*$ **és** $|u| \neq |v|$**.**

Példa.

$$S_1 \underset{G_1}{\Longrightarrow} ACA \overset{AC \to Aba}{\underset{G_1}{\Longrightarrow}} ABaA \overset{B \to A}{\underset{G_1}{\Longrightarrow}} AAaA \overset{A \to 0}{\underset{G_1}{\Longrightarrow}} 0AaA \overset{A \to 1}{\underset{G_1}{\Longrightarrow}} 01aA \overset{A \to 1}{\underset{G_1}{\Longrightarrow}} 01a1$$

Környezetfüggetlen nyelvtan

$G_2 = (N_2, T_2, P_2, K)$, **ahol** $N_2 = \{K, T, F\}$, $\quad T_2 = \{+, *, (,), a\}$

P_2 **elemei:**

$K \to K + T \mid T$

$T \to T * F \mid F$

$F \to (K) \mid a$

Az $L(G_2)$ **olyan algebrai kifejezésekből áll, amelyek a** T_2 **halmaz elemeiből képezhetők.**

Példa.

$$K \underset{G_2}{\Longrightarrow} K + T \underset{G_2}{\Longrightarrow} T + T \underset{G_2}{\Longrightarrow} F + T \underset{G_2}{\Longrightarrow} (K) + T \underset{G_2}{\Longrightarrow} (K) + F \underset{G_2}{\Longrightarrow}$$

$$(K){+}a \underset{G_2}{\Longrightarrow} (T){+}a \underset{G_2}{\Longrightarrow} (T{*}F){+}a \underset{G_2}{\Longrightarrow} (F{*}F){+}a \underset{G_2}{\Longrightarrow} (a{*}F){+}a \underset{G_2}{\Longrightarrow} (a{*}a){+}a$$

Reguláris nyelvtan

$G_3 = (N_3, T_3, P_3, S_3)$, **ahol** $N_3 = \{S_3, A, B\}$, $\quad T_3 = \{a, b\}$

P_3 **elemei:**

$S_3 \to aA$
$A \to aB \mid a$
$B \to aB \mid bB \mid a \mid b$

Az $L(G_3)$ **nyelv olyan,** a **és** b **betűkből képezhető szavakból áll, amelyek legalább két** a**-val kezdődnek.**

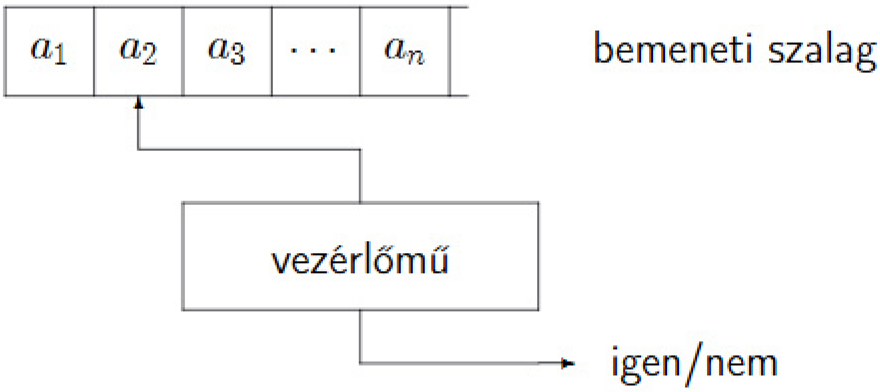
Példa.

$$S_3 \underset{G_3}{\Longrightarrow} aA \underset{G_2}{\Longrightarrow} aaB \underset{G_2}{\Longrightarrow} aabB \underset{G_2}{\Longrightarrow} aaba$$

nyelvosztályok	automaták
3. reguláris nyelvek	véges automaták
2. környezetfüggetlen nyelvek	veremautomaták
1. környezetfüggő nyelvek	korlátos Turing-automaták
0. mondatstruktúrájú nyelvek	Turing-automaták

Véges automaták és reguláris nyelvek

Véges automaták értelmezése



Nemdeterminisztikus véges automatának nevezzük az $\mathrm{A} = (Q, \Sigma, E, I, F)$ rendezett ötöst, ahol

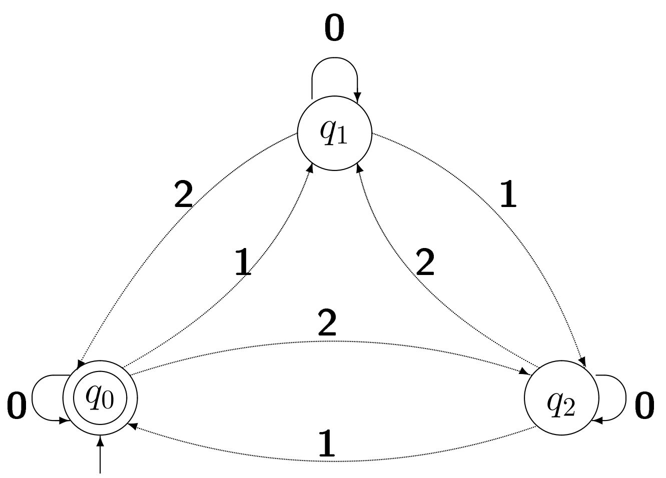
- Q egy véges, nem üres halmaz, az **állapotok** halmaza,
- Σ a **bemeneti ábécé**,
- E az **átmenetek** (vagy *élek*) halmaza, ahol $E \subseteq Q \times \Sigma \times Q$,
- $I \subseteq Q$ a **kezdőállapotok** halmaza,
- $F \subseteq Q$ a **végállapotok** halmaza.

nemdeterminisztikus véges automata — **irányított, címkézett gráf**, amelynek csúcsai az állapotok, és egy p csúcsából akkor vezet egy a betűvel megcímkézett él a q csúcsba, ha $(p, a, q) \in E$.

átmenetgráf

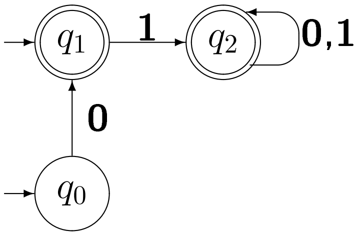
1. példa. $A = (Q, \Sigma, E, I, F)$, **ahol** $Q = \{q_0, q_1, q_2\}$, $\Sigma = \{0, 1, 2\}$,

$$E = \big\{(q_0, 0, q_0), (q_0, 1, q_1), (q_0, 2, q_2),\\ (q_1, 0, q_1), (q_1, 1, q_2), (q_1, 2, q_0),\\ (q_2, 0, q_2), (q_2, 1, q_0), (q_2, 2, q_1)\big\}, \quad I = \{q_0\}, \quad F = \{q_0\}.$$

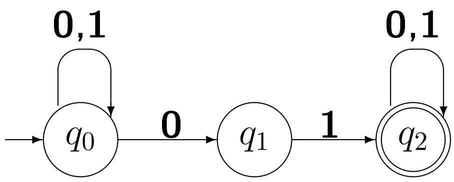


2. példa.

$I = \{q_0, q_1\}, F = \{q_1, q_2\}$ $\qquad$ $I = \{q_0\}, F = \{q_2\}$



A



B

Egy (p, a, q) **élnek** p **a kezdőpontja,** q **a végpontja,** a **pedig a címkéje.**

séta

$$(q_0, a_1, q_1), (q_1, a_2, q_2), \ldots, (q_{n-2}, a_{n-1}, q_{n-1}), (q_{n-1}, a_n, q_n)$$

élsorozat a nemdeterminisztikus véges automata egy sétája, amelynek címkéje az $a_1a_2\ldots a_n$ szó.

Ha $n = 0$, akkor $q_0 = q_n$ és $a_1a_2\ldots a_n = \varepsilon$ $\quad\Longrightarrow$ ***üres séta***.

a séta jelölése

$$q_0 \xrightarrow{a_1} q_1 \xrightarrow{a_2} q_2 \xrightarrow{a_3} \cdots \xrightarrow{a_{n-1}} q_{n-1} \xrightarrow{a_n} q_n,$$

vagy ha $w = a_1a_2\ldots a_n$, akkor röviden: $q_0 \xrightarrow{w} q_n$.

q_0 a séta kezdőpontja, q_n pedig a végpontja

Egy séta produktív, ha kezdőpontja kezdőállapot, végpontja pedig végállapot.

Azt mondjuk, hogy a nemdeterminisztikus véges automata felismer egy szót, ha az a szó egy produktív séta címkéje.

Az ε üres szót a nemdeterminisztikus véges automata akkor ismeri fel, ha van produktív üres séta, amely egyenértékű azzal, hogy van olyan kezdőállapot, amely egyben végállapot is.

Egy nemdeterminisztikus véges automata által felismert szavak halmazát a nemdeterminisztikus véges automata által felismert nyelvnek mondjuk.

Az A nemdeterminisztikus véges automata által felismert nyelv

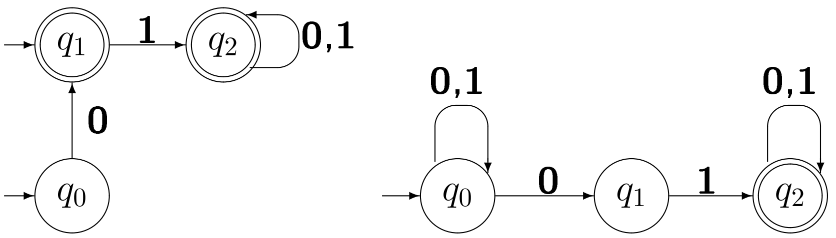
$$L(\mathrm{A}) = \left\{ w \in \Sigma^* \mid \exists p \in I,\ \exists q \in F,\ \exists p \xrightarrow{w} q \right\}.$$

Az A_1 és A_2 véges automaták ***ekvivalensek***, ha $L(\mathrm{A}_1) = L(\mathrm{A}_2)$.

Értelmezzük a következő átmenetfüggvényt:

$$\delta : Q \times \Sigma \to P(Q), \quad \delta(p,a) = \{q \in Q \mid (p,a,q) \in E\}$$

– egy p állapotnak és egy a betűnek megfelelteti azt az állapothalmazt, amelynek állapotaiba átmehet a véges automata, ha a p állapotban van és az olvasófej az a betűre mutat.



A B

δ	**0**	**1**
q_0	$\{q_1\}$	$\emptyset$
q_1	$\emptyset$	$\{q_2\}$
q_2	$\{q_2\}$	$\{q_2\}$

δ	**0**	**1**
q_0	$\{q_0, q_1\}$	$\{q_0\}$
q_1	$\emptyset$	$\{q_2\}$
q_2	$\{q_2\}$	$\{q_2\}$

A második automata megadható még:

$0, 1$
q_0, q_1, q_2
kezdő: q_0
vég: q_2
$q_0, 0, q_0$
$q_0, 0, q_1$
$q_0, 1, q_0$
$q_1, 1, q_2$
$q_2, 0, q_2$
$q_2, 1, q_2$

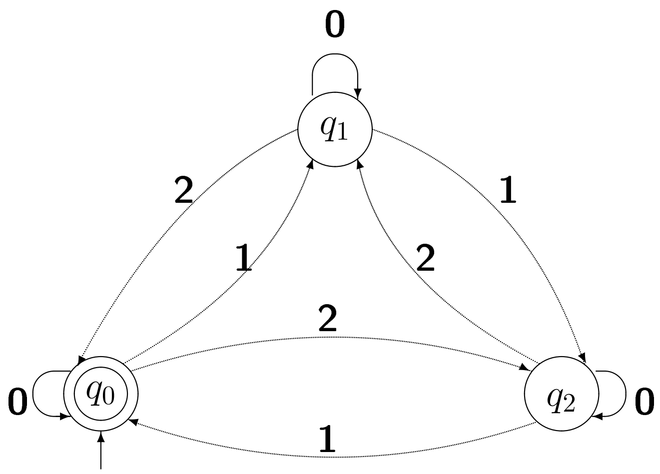
Jelöljük $|H|$-val a H halmaz elemeinek a számát.

Egy nemdeterminisztikus véges automata *determinisztikus*, ha

$$|I| = 1 \text{ és } |\delta(q, a)| \leq 1, \ \forall q \in Q, \ \forall a \in \Sigma.$$

A $|\delta(q, a)| \leq 1$ feltételt helyettesíthetjük a következővel:

$$(p, a, q) \in E, \ (p, a, r) \in E \implies q = r, \forall p, q, r \in Q, \forall a \in \Sigma.$$



δ	**0**	**1**	**2**
q_0	$\{q_0\}$	$\{q_1\}$	$\{q_2\}$
q_1	$\{q_1\}$	$\{q_2\}$	$\{q_0\}$
q_2	$\{q_2\}$	$\{q_0\}$	$\{q_1\}$

Ha a determinisztikus véges automata olyan, hogy $|\delta(q, a)| = 1$ minden $q \in Q$ állapotra és minden $a \in \Sigma$ betűre, akkor azt mondjuk, hogy teljes, determinisztikus véges automata.

Minden determinisztikus véges automata teljessé tehető egy új csapdaállapot bevezetésével.

$\mathrm{A} = (Q, \Sigma, E, \{q_0\}, F)$

A vele ekvivalens teljes, determinisztikus véges automata pedig

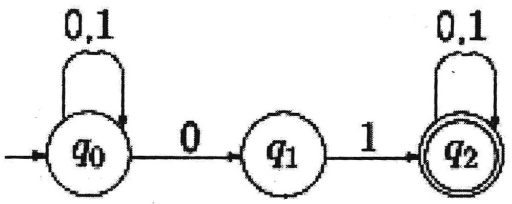
$\mathrm{A}' = (Q \cup \{s\}, \Sigma, E', \{q_0\}, F)$,
$E' = E \cup \big\{(p, a, s) \mid \delta(p, a) = \emptyset, p \in Q, a \in \Sigma\big\} \cup \big\{(s, a, s) \mid a \in \Sigma\big\}$.

Mivel ez az új állapot nem produktív, könnyű belátni, hogy $L(\mathrm{A}) = L(\mathrm{A}')$.

A következőkben, amennyiben csak véges automatát írunk, ezalatt mindig nemdeterminisztikus véges automatát értünk. Ha az automata determinisztikus, akkor ezt mindig kihangsúlyozzuk.

Példa:

nem teljes



teljes