Összefoglaló

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: $u \to v_1,\ u \to v_2, \ldots, u \to v_r$ **helyett** $u \to v_1 \mid v_2 \mid \ldots \mid v_r$

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

Levezetés (deriváció)

$S \Longrightarrow aSb \Longrightarrow a^2Sb^2 \Longrightarrow \ldots \Longrightarrow a^{n-1}Sb^{n-1} \Longrightarrow a^nb^n$,
vagy $S \stackrel{*}{\Longrightarrow} a^nb^n$.

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

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

A fenti példában: $L(G) = \{a^nb^n \mid n \geq 1\}$.

Chomsky-féle hierarchia

$G = (N, T, P, S)$ **nyelvtan:**

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

• 1-es típusú (környezetfüggő), ha minden szabálya $\alpha A\gamma \to \alpha\beta\gamma$ **alakú, ahol** $A \in N$, $\alpha, \gamma \in (N \cup T)^*$, $\beta \in (N \cup T)^+$. **Ezenkívül megengedhető az** $S \to \varepsilon$ **szabály is, ha** S **nem szerepel egyetlen szabály jobb oldalán sem.**

• 2-es típusú (környezetfüggetlen), ha minden szabálya $A \to \beta$ **alakú, ahol** $A \in N, \beta \in (N \cup T)^+$. **Ezenkívül megengedhető az** $S \to \varepsilon$ **szabály is, ha** S **nem szerepel egyetlen szabály jobb oldalán sem.**

• 3-as típusú (reguláris), ha szabályai $A \to aB$ **vagy** $A \to a$ **alakúak, ahol** $a \in T$ **és** $A, B \in N$. **Ezenkívül megengedhető az** $S \to \varepsilon$ **szabály is, ha** S **nem szerepel egyetlen szabály jobb oldalán sem.**

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

Kiterjesztett nyelvtanok

1-es típusú kiterjesztett nyelvtan. Minden szabály $\alpha \to \beta$ **alakú, ahol** $|\alpha| \le |\beta|$, **kivéve esetleg az** $S \to \varepsilon$ **szabályt.**

2-es típusú kiterjesztett nyelvtan. Minden szabály $A \to \beta$ **alakú, ahol** $A \in N, \beta \in (N \cup T)^*$.

3-as típusú kiterjesztett nyelvtan. Minden szabály $A \to uB$ **vagy** $A \to u$ **alakú, ahol** $A, B \in N$, $u \in T^*$.

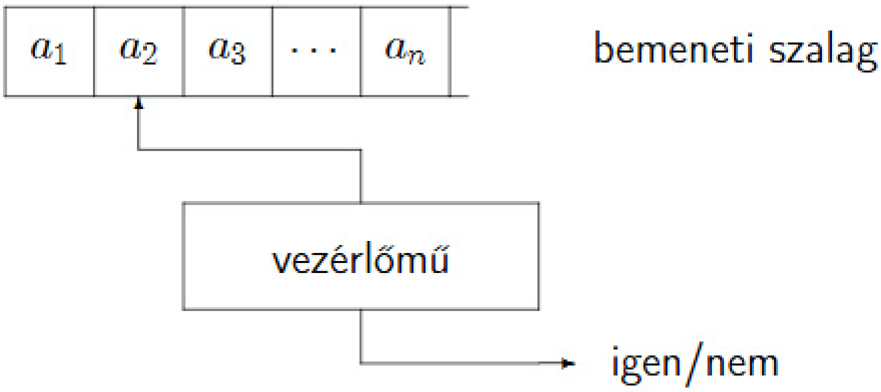
Tétel. ***Tetszőleges kiterjesztett nyelvtanhoz megadható egy vele ekvivalens, ugyanolyan típusú nyelvtan.***

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

nyelvtan	kiterjesztett nyelvtan
0-s típusú (általános vagy mondatszerkezetű), ha semmilyen megkötést nem teszünk a helyettesítési szabályaira	
1-es típusú (környezetfüggő), ha minden szabálya $\alpha A\gamma \to \alpha\beta\gamma$ alakú, ahol $A \in N$, $\alpha, \gamma \in (N \cup T)^*$, $\beta \in (N \cup T)^+$. Lehet még az $S \to \varepsilon$ szabály is, ha S nem szerepel egyetlen szabály jobb oldalán sem	1-es típusú kiterjesztett nyelvtan ha minden szabálya $\alpha \to \beta$ alakú, ahol $\|\alpha\| \leq \|\beta\|$, kivéve esetleg az $S \to \varepsilon$ szabályt.
2-es típusú (környezetfüggetlen), ha minden szabálya $A \to \beta$ alakú, ahol $A \in N$, $\beta \in (N \cup T)^+$. Lehet még az $S \to \varepsilon$ szabály is, ha S nem szerepel egyetlen szabály jobb oldalán sem	2-es típusú kiterjesztett nyelvtan, ha minden szabálya $A \to \beta$ alakú, ahol $A \in N, \beta \in (N \cup T)^*$.
3-as típusú (reguláris), ha minden szabálya $A \to aB$ vagy $A \to a$ alakú, ahol $a \in T$ és $A, B \in N$. Lehet még az $S \to \varepsilon$ szabály is, ha S nem szerepel egyetlen szabály jobb oldalán sem.	3-as típusú kiterjesztett nyelvtan, ha minden szabálya $A \to uB$ vagy $A \to u$ alakú, ahol $A, B \in N, u \in T^*$.

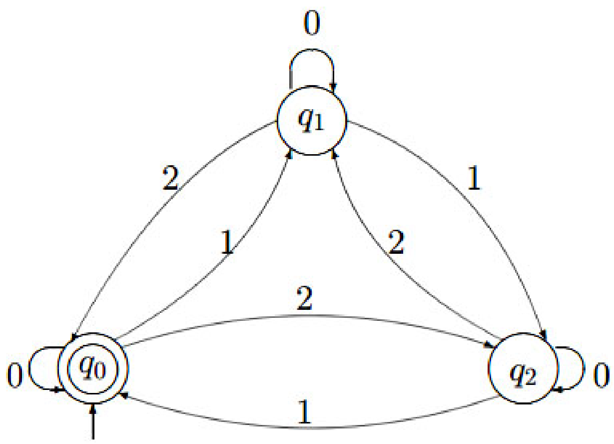
nyelvosztályok	automaták
3. reguláris nyelvek	véges automaták
2. környezetfüggetlen nyelvek	veremautomaták
1. környezetfüggő nyelvek	korlátos Turing-automaták
0. mondatszerkezetű nyelvek	Turing-automaták

Véges automaták



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.**



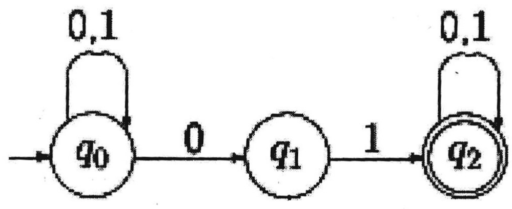
Egy nemdeterminisztikus véges automata *determinisztikus,* ha

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

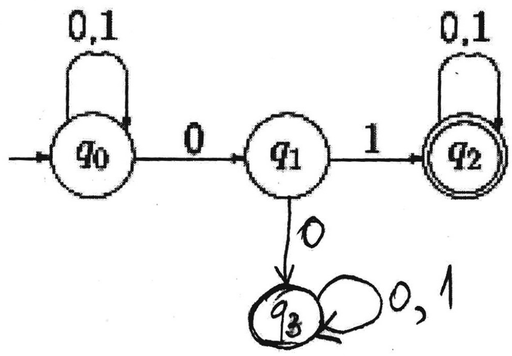
A $|\delta(q,a)| \le 1$ **feltételt helyettesíthetjük a következővel:**

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

nem teljes determinisztikus véges automata



teljes determinisztikus véges automata

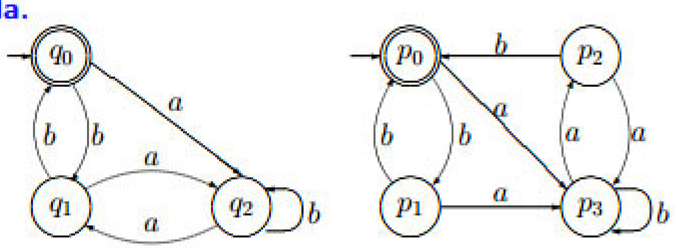


Determinisztikus véges automaták ekvivalenciája

teljes, determinisztikus véges automatákkal dolgozunk

$\delta(q, a)$ **halmaz mindig egyetlen elemet tartalmaz**

1. példa.

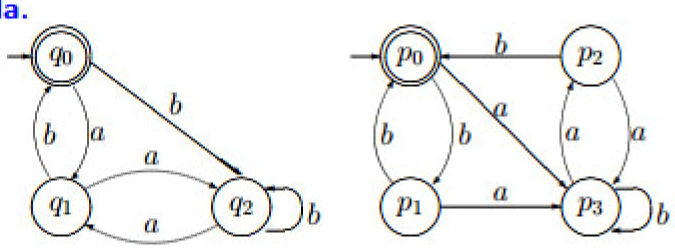


	a	b
(q_0, p_0)	(q_2, p_3)	(q_1, p_1)
(q_2, p_3)	(q_1, p_2)	(q_2, p_3)
(q_1, p_1)	(q_2, p_3)	(q_0, p_0)
(q_1, p_2)	(q_2, p_3)	(q_0, p_0)

ekvivalensek

---

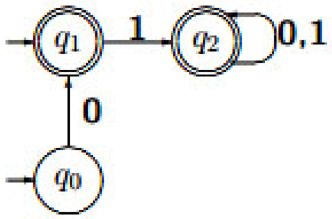
2. példa.



	a	b
(q_0, p_0)	(q_1, p_3)	(q_2, p_1)
(q_1, p_3)	(q_2, p_2)	(q_0, p_3)
(q_2, p_1)		
(q_2, p_2)		

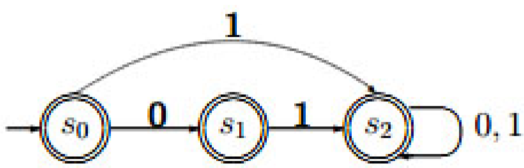
nem ekvivalensek

Nemdeteminisztikus automata $\longrightarrow$ determinisztikus automata



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

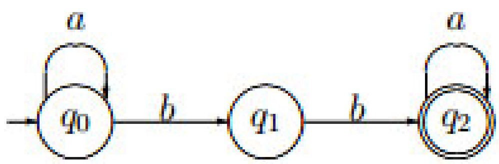
δ'	**0**	**1**
$s_0 = \{q_0, q_1\}$	$\{s_1\}$	$\{s_2\}$
$s_1 = \{q_1\}$	$\emptyset$	$\{s_2\}$
$s_2 = \{q_2\}$	$\{s_2\}$	$\{s_2\}$



Véges automaták és reguláris nyelvek kapcsolata

Példa. $A = (\{q_0, q_1, q_2\}, \{a, b\}, E, \{q_0\}, \{q_2\})$ determinisztikus véges automata, ahol $E = \{(q_0, a, q_0), (q_0, b, q_1), (q_1, b, q_2), (q_2, a, q_2)\}$.

δ	a	b
q_0	$\{q_0\}$	$\{q_1\}$
q_1	$\emptyset$	$\{q_2\}$
q_2	$\{q_2\}$	$\emptyset$



A tétel alapján: $G = (\{q_0, q_1, q_2\}, \{a, b\}, P, q_0)$

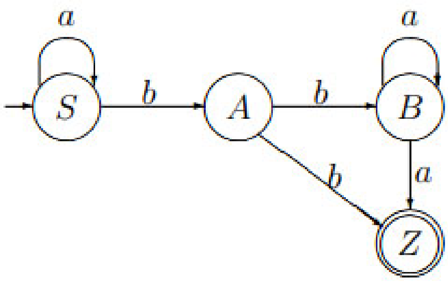
P szabályai a következők:
$q_0 \to aq_0 \mid bq_1$, $\quad q_1 \to bq_2 \mid b$, $\quad q_2 \to aq_2 \mid a$.

vagy átnevezve: $q_0 : S$; $\quad q_1 : A$; $\quad q_2 : B$

$S \to aS \mid bA$, $\quad A \to bB \mid b$, $\quad B \to aB \mid a$.

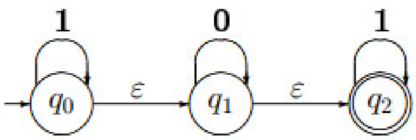
Példa. Adott a $G = (\{S, A, B\}, \{a, b\}, \{S \to aS, S \to bA, A \to bB, A \to b, B \to aB, B \to a\}, S)$ reguláris nyelvtan. A hozzá rendelt véges automata $A = (\{S, A, B, Z\}, \{a, b\}, E, \{S\}, \{Z\})$, ahol $E = \{(S, a, S), (S, b, A), (A, b, B), (A, b, Z), (B, a, B), (B, a, Z)\}$. Ennek átmenettáblázata a következő:

δ	a	b
S	$\{S\}$	$\{A\}$
A	$\emptyset$	$\{B, Z\}$
B	$\{B, Z\}$	$\emptyset$
E	$\emptyset$	$\emptyset$



Ez a véges automata egyszerűsíthető. A B és Z állapotok összevonhatók egyetlen végállapottá.

ε-lépéses automaták



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

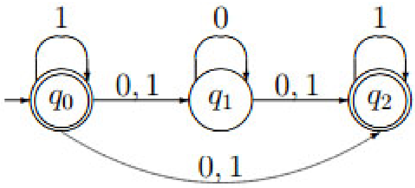
Alkalmazzuk az EPSZILON-MENTESÍTÉS algoritmust.

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

$\overline{\delta}$	0	1
q_0	$\{q_1, q_2\}$	$\{q_0, q_1, q_2\}$
q_1	$\{q_1, q_2\}$	$\{q_2\}$
q_2	$\emptyset$	$\{q_2\}$

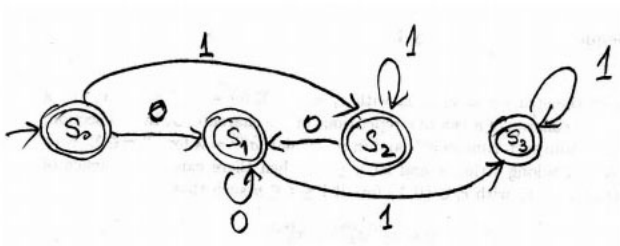
Tehát a $\overline{\mathrm{A}}$ véges automata átmenettáblázata és átmenetdiagramja:

$\overline{\delta}$	0	1
q_0	$\{q_1, q_2\}$	$\{q_0, q_1, q_2\}$
q_1	$\{q_1, q_2\}$	$\{q_2\}$
q_2	$\emptyset$	$\{q_2\}$

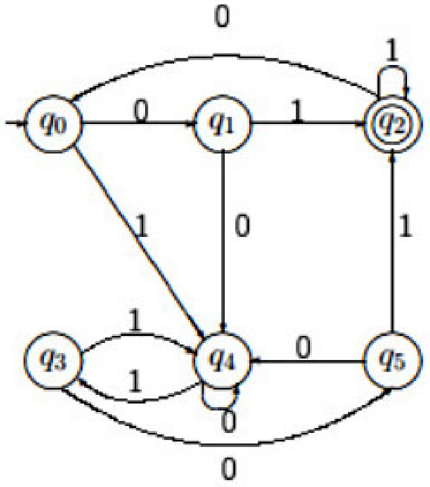


Ekvivalens determinisztikus automata hozzárendelése:

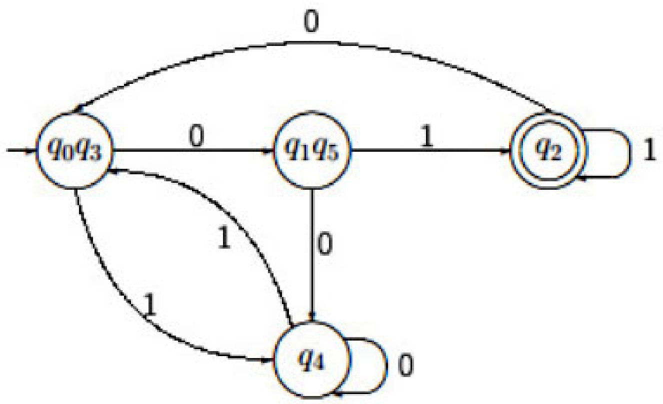
	0	1
$S_0 = \{q_0\}$	$\{S_1\}$	$\{S_2\}$
$S_1 = \{q_1, q_2\}$	$\{S_1\}$	$\{S_3\}$
$S_2 = \{q_0, q_1, q_2\}$	$\{S_1\}$	$\{S_2\}$
$S_3 = \{q_2\}$	$\emptyset$	$\{S_3\}$



Automaták minimalizálása

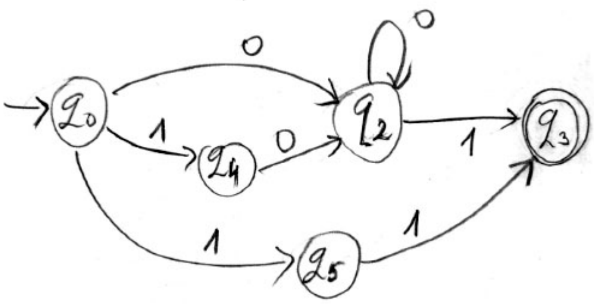


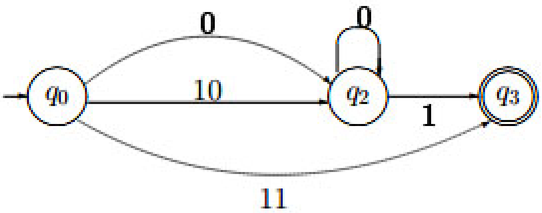
	q_1	q_2	q_3	q_4	q_5
q_4					*
q_3				*	*
q_2			*	*	*
q_1		*	*	*	
q_0	*	*		*	*

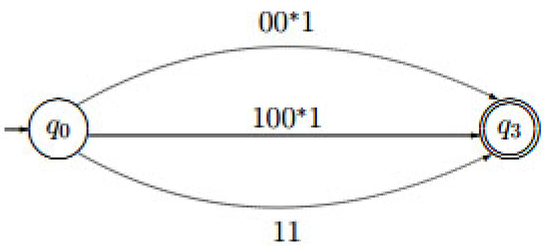


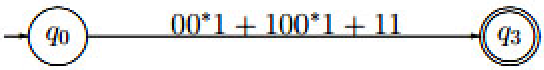
Reguláris kifejezések

Automata ⟶ reguláris kifejezés

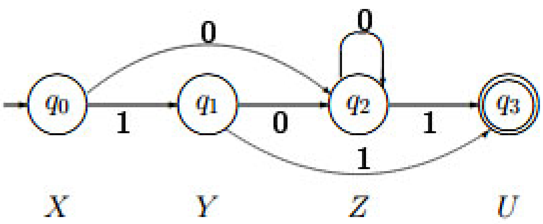








Automata ⟶ reguláris kifejezés (egyenletekkel)



$X = \varepsilon$
$Y = X1$
$Z = X0 + Y0 + Z0$
$U = Y1 + Z1.$

Ha egy egyenlet $X = X\alpha + \beta$ alakú, ahol α, β tetszőleges szavak, amelyek nem tartalmazzák az X változót, akkor könnyű ellenőrizni, egyszerű behelyettesítéssel, hogy $X = \beta\alpha^*$ megoldása az egyenletnek.

Példánkban

$X = \varepsilon$
$Y = X1$
$Z = X0 + Y0 + Z0$
$U = Y1 + Z1.$

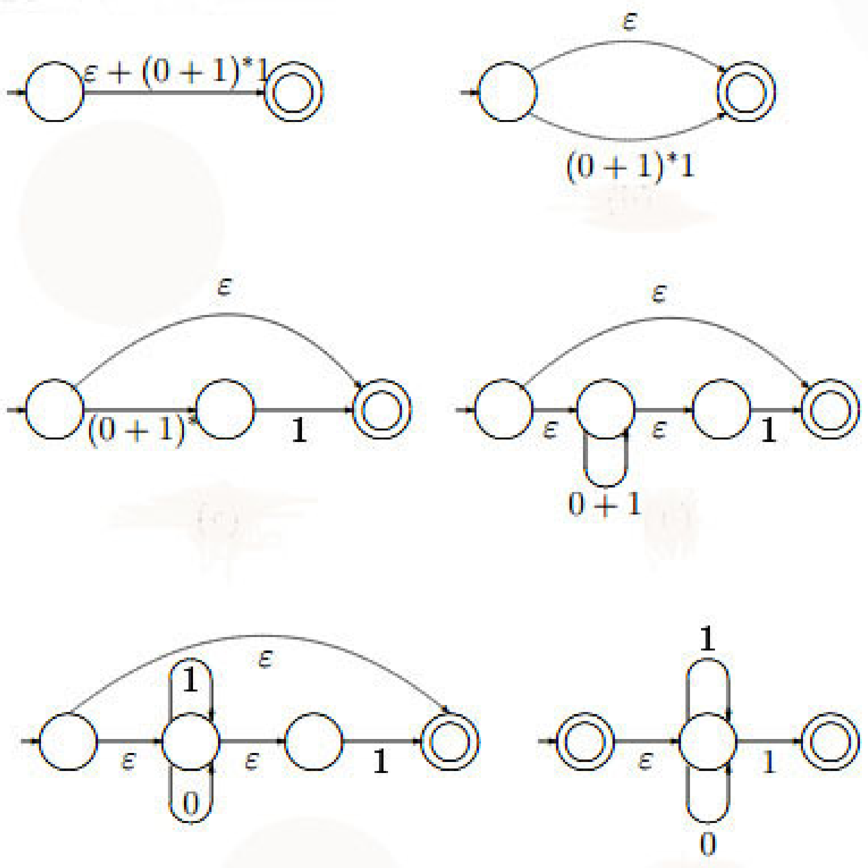
a fenti egyenletrendszer első egyenletének segítségével azt kapjuk, hogy $Y = 1$.

Innen $Z = 0 + 10 + Z0$, **azaz** $Z = Z0 + (0 + 10)$, **és ezt megoldva azt kapjuk, hogy** $Z = (0 + 10)0^*$.

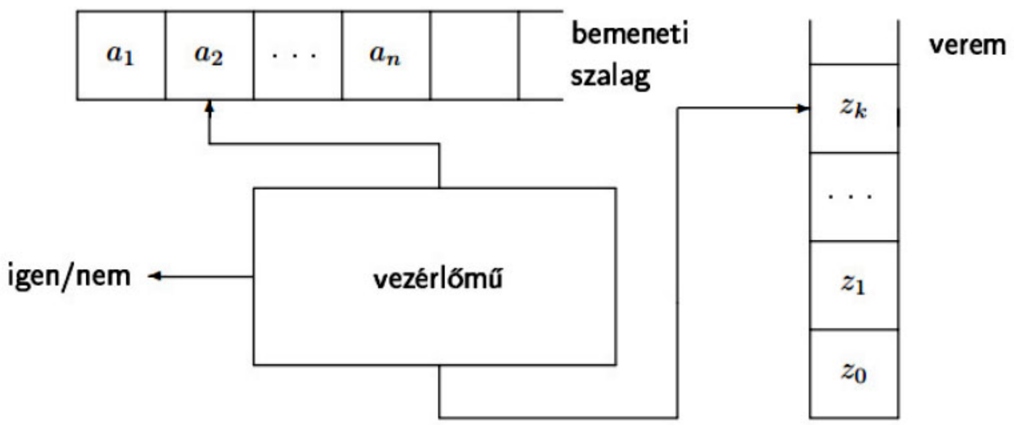
Innen pedig U **egyszerűen megkapható:** $U = 11 + (0 + 10)0^*1$.

Reguláris kifejezés ⟶ automata

Induljunk el az $\varepsilon + (0+1)^*1$ reguláris kifejezésből.



Veremutomaták és környezetfüggetlen nyelvek



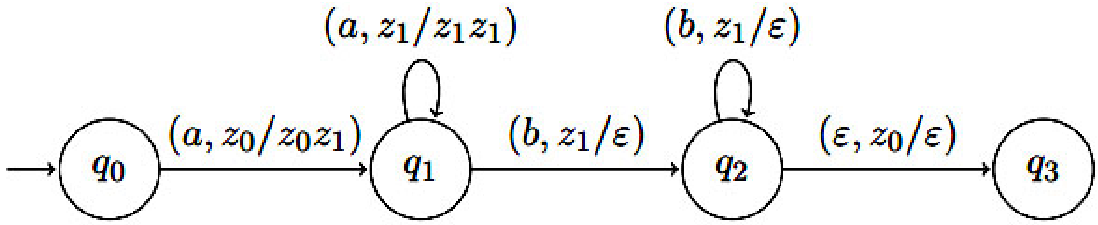
Nemdeterminisztikus veremautomatának nevezzük a

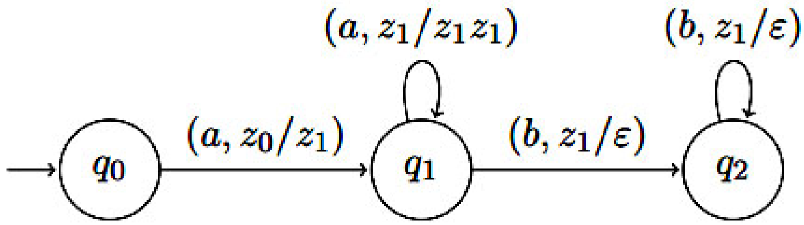
$$V = (Q, \Sigma, W, E, q_0, z_0, F)$$

rendezett hetest, ahol

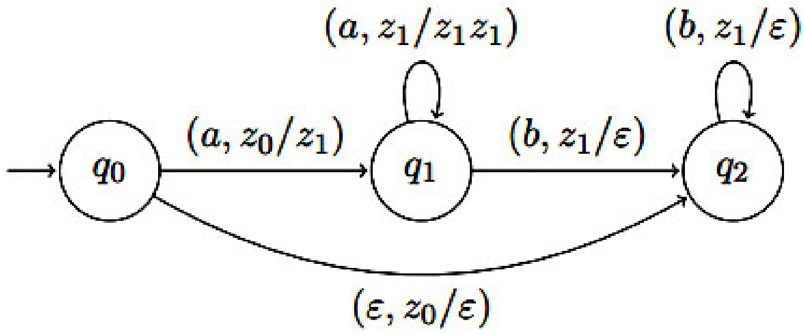
- Q az állapotok véges, nem üres halmaza,
- Σ a bemeneti ábécé,
- W a veremábécé,
- $E \subseteq Q \times (\Sigma \cup \{\varepsilon\}) \times W \times W^* \times Q$ az átmenetek vagy élek halmaza,
- $q_0 \in Q$ a kezdőállapot,
- $z_0 \in W$ a veremmemória kezdőjele,
- $F \subseteq Q$ a végállapotok halmaza.

$L_1 = \{a^n b^n \mid n \geq 1\}$

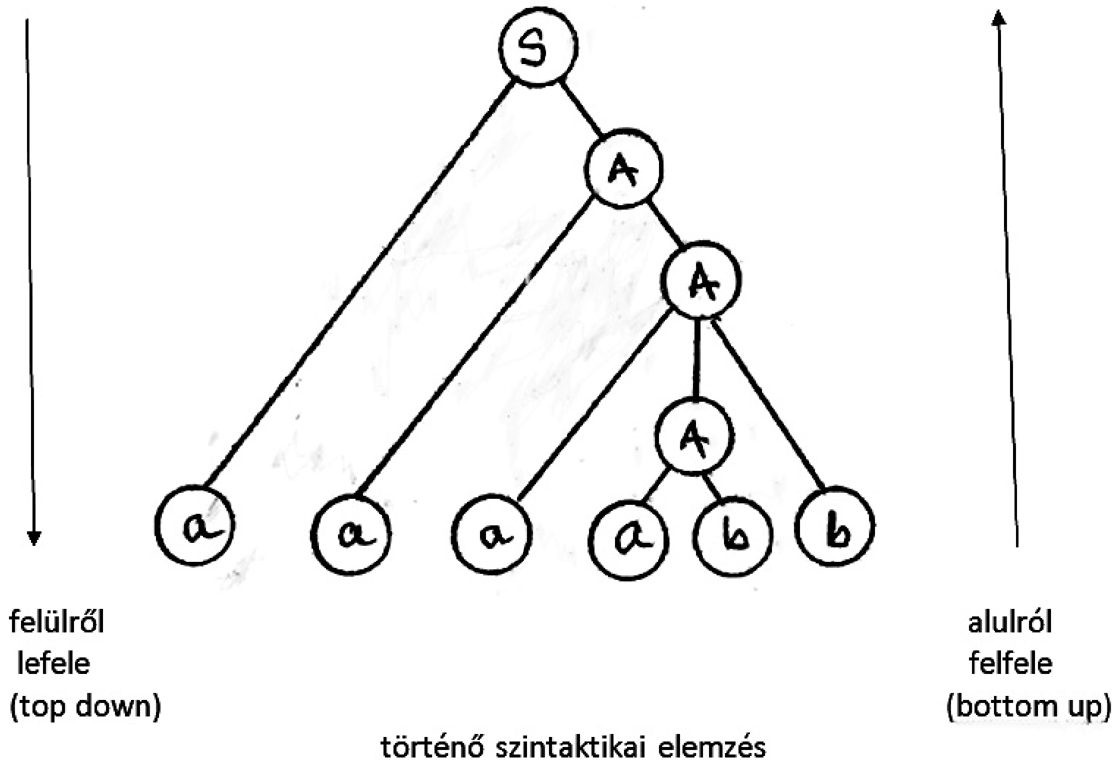




$L_2 = \{a^n b^n \mid n \geq 0\}$



Szintaktikai elemzés



Felülről lefele haladó (top down) elemzések:

- **LL**(1)**-elemzés**
- **rekurzív leszállás módszere**

LL(1)-elemzés

$$S \rightarrow aS \mid A$$
$$A \rightarrow bAc \mid d \mid \epsilon$$

$$FOLLOW_1(S) = \{\#\}$$
$$FOLLOW_1(A) = \{c, \#\}$$

$$FIRST_1(aS\ FOLLOW_1(S)) = \{a\}$$
$$FIRST_1(A\ FOLLOW_1(S)) = \{b, d, \#\}$$

$$FIRST_1(bAc\ FOLLOW_1(A)) = \{b\}$$
$$FIRST_1(d\ FOLLOW_1(A)) = \{d\}$$
$$FIRST_1(\epsilon\ FOLLOW_1(A)) = \{c, \#\}$$

	a	b	c	d	$\#$
S	$(aS, 1)$	$(A, 2)$		$(A, 2)$	$(A, 2)$
A		$(bAc, 3)$	$(\varepsilon, 5)$	$(d, 4)$	$(\varepsilon, 5)$
a	pop				
b		pop			
c			pop		
d				pop	
$\#$					$elfogad$

Az elemzés az *abc* szóra a következő:

$$
\begin{array}{llrrrll}
(abc\#, S\#, \varepsilon) & \xrightarrow{(aS,1)} & (& abc\#, & aS\#, & 1 &) \\
& \xrightarrow{pop} & (& bc\#, & S\#, & 1 &) \\
& \xrightarrow{(A,2)} & (& bc\#, & A\#, & 12 &) \\
& \xrightarrow{(bAc,3)} & (& bc\#, & bAc\#, & 123 &) \\
& \xrightarrow{pop} & (& c\#, & Ac\#, & 123 &) \\
& \xrightarrow{(\varepsilon,5)} & (& c\#, & c\#, & 1235 &) \\
& \xrightarrow{pop} & (& \#, & \#, & 1235 &) \\
& \xrightarrow{elfogad} & \textbf{OK} & & & &
\end{array}
$$

Az *abc* szó levezetése az 1, 2, 3, 5 helyettesítési szabályok alkalmazásával:

$S \Rightarrow aS \Rightarrow aA \Rightarrow abAc \Rightarrow abc$

Elemzés az *aba* szóra:

$$
\begin{array}{llrrrll}
(aba\#, S\#, \varepsilon) & \xrightarrow{(aS,1)} & (& aba\#, & aS\#, & 1 &) \\
& \xrightarrow{pop} & (& ba\#, & S\#, & 1 &) \\
& \xrightarrow{(A,2)} & (& ba\#, & A\#, & 12 &) \\
& \xrightarrow{(bAa,3)} & (& ba\#, & bAc\#, & 123 &) \\
& \xrightarrow{pop} & (& a\#, & Ac\#, & 123 &) \\
& \rightarrow & \textbf{HIBA} & & & &
\end{array}
$$

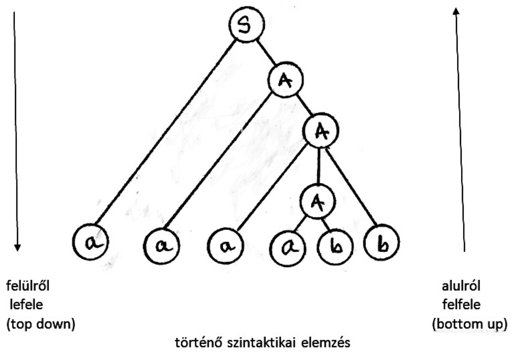
Rekurzív leszállás módszere (Példa C++-ban)

$S' \to S$
$S \to aS \mid A$
$A \to bA \mid \varepsilon$

```
# include < iostream >
using namespace std;

char aktszimb;
char w[1024];
int i;

void S();
void A();

void Vizsgal(char a){
    if ( a == aktszimb )
    aktszimb = w[++i];
    else {
          cout << "HIBA ! " << endl;
          exit(1);
    }
}

void S(){
    switch (aktszimb){
         case 'a' : Vizsgal('a') ; S(); break;
         case 'b' : A(); break;
         case '#' : ;
    }
}
```

```
void A(){
    switch (aktszimb) {
        case 'b' : Vizsgal('b'); A(); break;
        case '#' : ;
     }
}

int main(){
  cout << "w = ";
  cin >> w;
  i = 0 ;
  aktszimb = w[i];
  S() ;
  Vizsgal ('#');
  return 0;
}
```

Alulról felfele haladó (bottom up) elemzések:



- **LR(1)-elemzés**
- **elemzés elsőbbségi (precedencia) nyelvtannal**

LR(1)-elemzés

$G = (\{S', S, A\}, \{a, b\}, P, S')$ **helyettesítési szabályai (P):**
(0) $S' \to S$
(1) $S \to AA$
(2) $A \to aA$
(3) $A \to b$

a táblázata pedig

LR(1)

állapot	action			goto	
	a	b	$\#$	S	A
0	$s3$	$s4$		1	2
1			*elfogad*		
2	$s6$	$s7$			5
3	$s3$	$s4$			8
4	$r3$	$r3$			
5			$r1$		
6	$s6$	$s7$			9
7			$r3$		
8	$r2$	$r2$			
9			$r2$		

LALR(1)

állapot	action			goto	
	a	b	$\#$	S	A
0	$s3$	$s4$		1	2
1			$elfogad$		
2	$s3$	$s4$			5
3	$s3$	$s4$			6
4	$r3$	$r3$	$r3$		
5			$r1$		
6	$r2$	$r2$	$r2$		

Példa. Elemezzük az előző példában megadott táblázat felhasználásával az $abb\#$ szöveget.

állapot	action			goto	
	a	b	$\#$	S	A
0	$s3$	$s4$		1	2
1			$elfogad$		
2	$s3$	$s4$			5
3	$s3$	$s4$			6
4	$r3$	$r3$	$r3$		
5			$r1$		
6	$r2$	$r2$	$r2$		

			szabály
$(\#0, \quad abb\#)$	$\xrightarrow{s3}$	$(\#0a3, \quad bb\#)$	
	$\xrightarrow{s4}$	$(\#0a3b4, \quad b\#)$	
	$\xrightarrow{r3}$	$(\#0a3A6, \quad b\#)$	$A \rightarrow b$
	$\xrightarrow{r2}$	$(\#0A2, \quad b\#)$	$A \rightarrow aA$
	$\xrightarrow{s4}$	$(\#0A2b4, \quad \#)$	
	$\xrightarrow{r3}$	$(\#0A2A5, \quad \#)$	$A \rightarrow b$
	$\xrightarrow{r1}$	$(\#0S1, \quad \#)$	$S \rightarrow AA$
	$\xrightarrow{\textit{elfogad}}$	**O.K.**	

Precedencia-elemzés

Legyen $G = (\{S, A, B\}, \{+, *, a\}, P, S)$**, ahol a** P **szabályai:**

$$S \to A$$
$$A \to A + B \mid B$$
$$B \to C * B \mid C$$
$$C \to a$$

	S	A	B	C	$+$	$*$	a	$\$$
S								
A					$\doteq$			$\gtrdot$
B					$\gtrdot$			$\gtrdot$
C					$\gtrdot$	$\doteq$		$\gtrdot$
$+$			$\doteq$	$\lessdot$			$\lessdot$	
$*$			$\doteq$	$\lessdot$			$\lessdot$	
a					$\gtrdot$	$\gtrdot$		$\gtrdot$
$\$$		$\lessdot$	$\lessdot$	$\lessdot$			$\lessdot$	

Az $a + a * a$ **szót vizsgáljuk.**

$$\$ \lessdot a \gtrdot + \quad a \quad * \quad a \quad \$$$ **redukálás** $C \to a$

$$\$ \lessdot C \gtrdot + \quad a \quad * \quad a \quad \$$$ **redukálás** $B \to C$

$$\$ \lessdot B \gtrdot + \quad a \quad * \quad a \quad \$$$ **redukálás** $A \to B$

$$\$ \lessdot A \doteq + \lessdot a \gtrdot * \quad a \quad \$$$ **redukálás** $C \to a$

$\$ \lessdot A \doteq + \lessdot C \doteq * \lessdot a \gtrdot \$$	**redukálás** $C \to a$
$\$ \lessdot A \doteq + \lessdot C \doteq * \lessdot C \gtrdot \$$	**redukálás** $B \to C$
$\$ \lessdot A \doteq + \lessdot C \doteq * \doteq B \gtrdot \$$	**redukálás** $B \to C * B$
$\$ \lessdot A \doteq + \doteq B \gtrdot \$$	**redukálás** $A \to A + B$
$\$ \lessdot A \gtrdot \$$	**redukálás** $S \to A$
$\$ \; S \; \$$	

Ez megfelel a következő (legjobboldalibb) levezetésnek:

$S \Rightarrow A \Rightarrow A + B \Rightarrow A + C * B \Rightarrow A + C * C \Rightarrow A + C * a \Rightarrow A + a * a$
$\Rightarrow B + a * a \Rightarrow C + a * a \Rightarrow a + a * a$