Formális nyelvek és automaták

Nagy Sára gyakorlatai alapján
Készítette: Nagy Krisztián

6. gyakorlat

Visszatérés egy előző órai beadható feladatra:

1. feladat:

$S \rightarrow XAKBY$
$K \rightarrow AKB|AB$
$AB \rightarrow BaA$
$XB \rightarrow aX$
$AY \rightarrow Ya$
$aB \rightarrow Ba$
$Aa \rightarrow aA$
$X \rightarrow aa$
$Ya \rightarrow aa$

Milyen nyelvet generál a fentebbi grammatika?

Megoldás:

$$S \rightarrow XAKBY \rightarrow^* \begin{matrix} XA^nB^nY \\ n \geq 2 \end{matrix} \rightarrow XA^{n-1}BaAB^{n-1}Y \rightarrow^* XA^{n-2}BaAaBaAB^{n-2}Y \rightarrow^*$$
$$\rightarrow^* XB^na^{n^2}A^nY \rightarrow aXB^{n-1}a^{n^2}A^nY \rightarrow^* a^nXa^{n^2}A^nY \rightarrow^* a^nXa^{n^2}Ya^n \rightarrow^* a^na^2a^{n^2}aa^n$$

$L(G) = \left\{a^{n^2+2n+3} |\ n \geq 2\right\}$

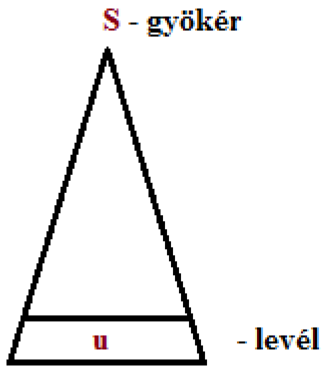
Emlékeztető:

2. típusú nyelvtanok körében:

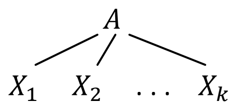
$A \rightarrow p \;\; p \in (T \cup N)^*$
$A \in N$ típusú szabályok találhatóak.

Szintaxisfa építés

A mondat szintaxisfájának levelei a nyelvtan terminális szimbólumai, a szintaxisfa többi pontja a nemterminális szimbólumokat reprezentálja, a gyökérelem pedig a nyelvtan kezdőszimbóluma.



$A \rightarrow X_1 \dots X_k \in P$ típusú szabályokat a szintaxisfában az alábbi módon ábrázoljuk:



$u \in L(G)$

Nézzünk egy régebbi példát:

Egy kifejezés (expression) sémája:
$E \rightarrow T|T + E$
$T \rightarrow F|F * T$
$F \rightarrow i|(E)$

Terminálisok: $X = \{+,*, i, (,)\}$

Helyesek-e az alábbi kifejezések?

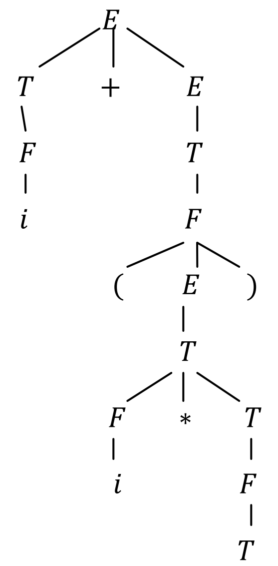
$a)\ i * i + i * i$ ✔
$b)\ i + (i * i)$ ✔
$c)\ (i + +i) * i$ ✘
$d)\ (i * i) + i + (i)$ ✔

Vezessük le $b)\ i + (i * i)$-t a jól ismert módszerünkkel!

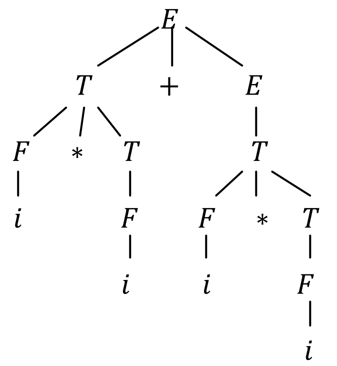
$E \rightarrow T + E \rightarrow F + E \rightarrow i + E \rightarrow i + T \rightarrow i + F \rightarrow i + (E) \rightarrow i + (T) \rightarrow i + (F * T) \rightarrow i + (i * T) \rightarrow$
$\rightarrow i + (i * F) \rightarrow i + (i * i)$

Nézzük meg ugyan ennek a b) részfeladatnak a felírását szintaxisfával!

$b)\ i + (i * i)$



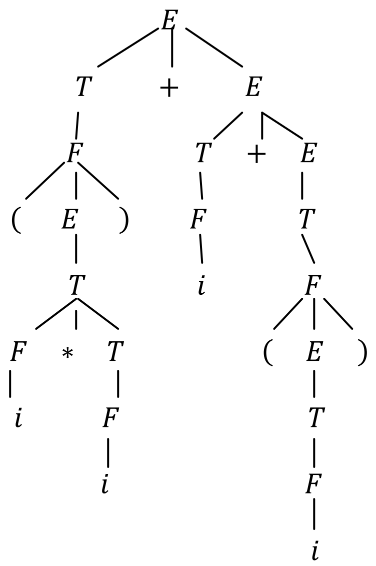
$a)\ i * i + i * i$



Megjegyzések:

A
$|$
ε → Nem lehet testvére. Csak a legalján szerepelhet egy szintaxisfában.

$d)\ (i * i) + i + (i)$



Chomsky normálforma (Legfeljebb bináris lehet a szintaxisfa)

Megengedett szabályok:

$$A \to BC \quad A, B, C \in N$$
$$A \to a \quad A \in N, a \in T$$
$$+\ KES\ szabály$$

*(**<u>Emlékeztető</u>** ismét: 2. típusú **nyelvtanok körében:***
$A \to p \;\; p \in (T \cup N)^*$
$A \in N$ *típusú szabályok találhatóak.)*

Készítsünk egy táblázatot, melyben meg vizsgáljuk, hogy a 2. típusú nyelvtanok körében mely szabályoknak rossz az alakja és milyen algoritmus nyújt megoldást a problémára.

CHOMSKY NORMÁLFORMA

Rossz alakú szabály	**Algoritmus**	**Példa**

$A \rightarrow \varepsilon$	ε-mentesítés	**Lásd 4. gyakorlat!**
$A \rightarrow B$	Lánctalanítás	$A \rightarrow B\|\alpha_1\| \dots \|\alpha_k$ $B \rightarrow \beta_1\| \dots \|\beta_t$ **Mentesítés:** $A \rightarrow \beta_1\| \dots \|\beta_t\|\alpha_1\| \dots \|\alpha_k$ $B \rightarrow \beta_1\| \dots \|\beta_t$ B szabály megmarad, mert A-ban lehet, hogy valahol még hivatkoznak rá. Lánc végéről kell visszafelé helyettesíteni
$A \rightarrow X_1 \dots X_k$ $x_i \in (T \cup N)$ $i \in [1, k]$	Álterminálisok bevezetése	$A \rightarrow aAbB$ **Álterminálisokkal:** $A \rightarrow Q_a A Q_b B$ $Q_a \rightarrow a$ $Q_b \rightarrow b$
$A \rightarrow B_1 B_2 \dots B_k \quad k \geq 3$	Hosszredukció	$A \rightarrow B_1 B_2 \dots B_k \quad k \geq 3$ **Redukció:** $A \rightarrow B_1 V_1$ $V_1 \rightarrow B_2 V_2$ . . . $V_{k-2} \rightarrow B_{k-1} B_k$

Feladatok (Chomsky-normálformára hozás):

1. Feladat: Hozzuk Chomsky-normálformára az alábbi grammatikát!

$S \to aAB|AB$
$A \to \varepsilon|B|Aa$
$B \to bA|\varepsilon$

1. lépés: ε-mentesítés

Írjuk fel a H halmaz elemeit!

$S \to aAB|AB$
$A \to \varepsilon|B|Aa$
$B \to bA|\varepsilon$

$H_1 = \{A, B\}$
$H_2 = H_1 \cup \{S\} = \{S, A, B\}$
Mivel S nem szerepel szabály jobb oldalán, így nem fontos bevezetnünk új kezdőszimbólumot, hanem $S \to \varepsilon$ szerepelhet.

$S \to aAB|aB|aA|a|AB|A|B|\varepsilon$
$A \to B|Aa|a$
$B \to bA|b$

2. lépés: Lánctalanítás

$$\begin{array}{ll} S & S \\ | & | \\ A & B \\ | & \\ B & \end{array}$$

$S \to aAB|aB|aA|a|AB|A|B|\varepsilon$
$A \to B|Aa|a$
$B \to bA|b$

$S \to aAB|aB|aA|a|AB|bA|b|Aa|\varepsilon$
$A \to bA|b|Aa|a$
$B \to bA|b$

Megjegyzés: Ha valamelyik szabályból kétszer lehet levezetni ugyan azt a terminálist, akkor csak egyszer írjuk le. (Jelen esetben $S \to a$ volt ilyen szabály.)
Továbbá az $A \to A$ tpusú szabályokat elhagyjuk, hiszen önmagába mutató szabály, mely nem csinál semmit.

3. lépés: Álterminálisok bevezetése

Az $A \to a \qquad A \in N, a \in T$ típusú szabályban található terminálisokat leszámítva, az összes többi típusú szabályokban található terminálisokra, egy új nem terminális jelet vezetünk be úgy, hogy $Q_a \;\; a \in T$.

$S \to aAB|aB|aA|a|AB|bA|b|Aa|\varepsilon$
$A \to bA|b|Aa|a$
$B \to bA|b$

$S \to Q_aAB|Q_aB|Q_aA|a|AB|Q_bA|b|AQ_a|\varepsilon$
$A \to Q_bA|b|AQ_a|a$
$B \to Q_bA|b$
$Q_a \to a$
$Q_b \to b$

4. lépés: Hosszredukció

$S \to Q_aAB|Q_aB|Q_aA|a|AB|Q_bA|b|AQ_a|\varepsilon$
$A \to Q_bA|b|AQ_a|a$
$B \to Q_bA|b$
$Q_a \to a$
$Q_b \to b$

Mivel ez az egy olyan szabály van ami legalább 3 hosszú, így csak ezt kell redukálnunk.

$S \to Q_aX|Q_aB|Q_aA|a|AB|Q_bA|b|AQ_a|\varepsilon$
$A \to Q_bA|b|AQ_a|a$
$B \to Q_bA|b$
$Q_a \to a$
$Q_b \to b$
$X \to AB$

2. Feladat: Hozzuk Chomsky-normálformára az alábbi grammatikát!

$S \to aSbS|\varepsilon$ (Helyes zárójelezés nyelvét írja le)

1. lépés: ε-mentesítés

$S \to aSbS|\varepsilon$

$H_1 = \{S\} = H$

Mivel a kezdő szimbólumból levezethető ε és a kezdő szimbólum a szerepel szabály jobboldalán, így új kezdő szimbólumot vezetünk be.

$S' \to S|\varepsilon$
$S \to aSbS|abS|aSb|ab$

2. lépés: Lánctalanítás

$S' \to S|\varepsilon$
$S \to aSbS|abS|aSb|ab$

$S' \to aSbS|abS|aSb|ab|\varepsilon$
$S \to aSbS|abS|aSb|ab$

3. lépés: Álterminálisok bevezetése

$S' \to aSbS|abS|aSb|ab|\varepsilon$
$S \to aSbS|abS|aSb|ab$

$S' \to Q_aSQ_bS|Q_aQ_bS|Q_aSQ_b|Q_aQ_b|\varepsilon$
$S \to Q_aSQ_bS|Q_aQ_bS|Q_aSQ_b|Q_aQ_b$
$Q_a \to a$
$Q_b \to b$

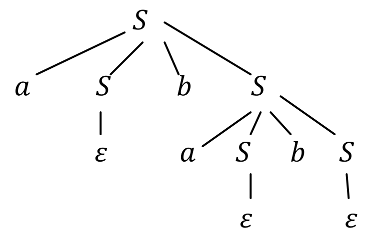
4. lépés: Hosszredukció

$S' \to Q_aSQ_bS|Q_aQ_bS|Q_aSQ_b|Q_aQ_b|\varepsilon$
$S \to Q_aSQ_bS|Q_aQ_bS|Q_aSQ_b|Q_aQ_b$

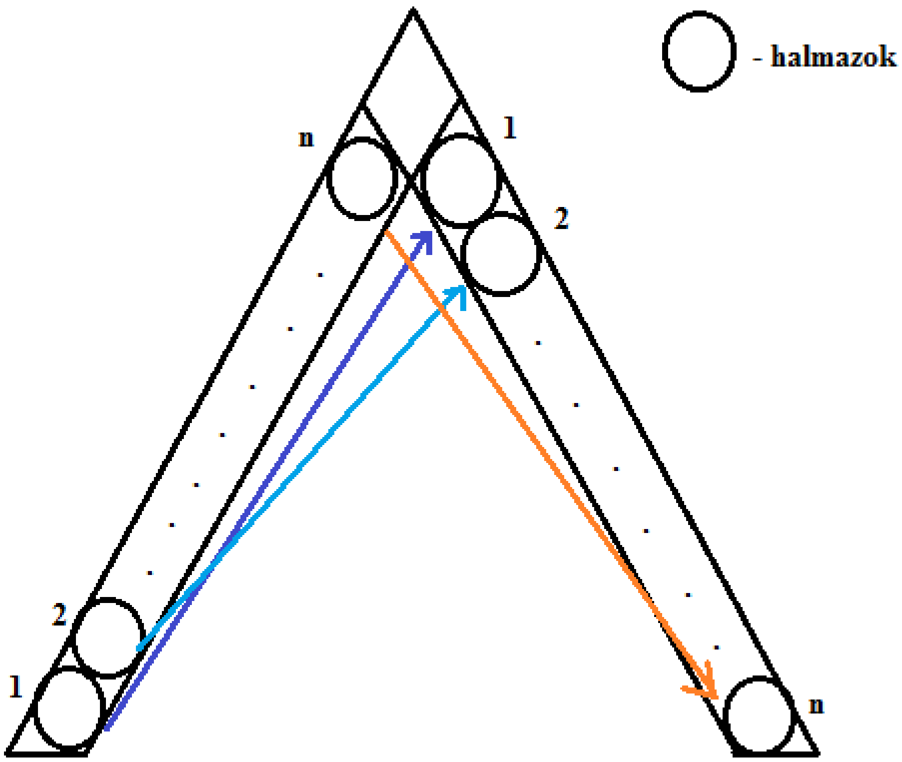
$S' \to XY|Q_aY|XQ_b|Q_aQ_b|\varepsilon$
$S \to XY|Q_aY|XQ_b|Q_aQ_b$
$X \to Q_aS$
$Y \to Q_bS$
$Q_a \to a$
$Q_b \to b$

A helyes zárójelezés nyelve alapján írjuk föl az $abab$ szót szintaxisfával!

$abab$



CYK – algoritmus



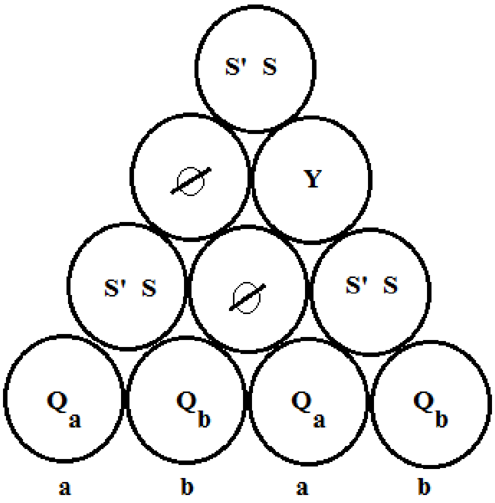
A nyilak mutatják, hogy kell szorozni a halmazokat, hogy megkapjuk a felette levő halmazt.

- Chomsky- normálforma szükséges a használatához.
- Alulról- felfele építkezünk

Amennyiben megjelenik a kezdőszimbólum a „piramis” tetején úgy az adott szó levezethető a grammatikával, tehát a szó eleme az adott nyelvnek. Ha nem jelenik meg, akkor nem eleme.

Feladat (CYK - algoritmussal): Az $abab$ szó eleme-e az alábbi nyelvnek?

$S' \rightarrow XY|Q_aY|XQ_b|Q_aQ_b|\varepsilon$
$S \rightarrow XY|Q_aY|XQ_b|Q_aQ_b$
$X \rightarrow Q_aS$
$Y \rightarrow Q_bS$
$Q_a \rightarrow a$
$Q_b \rightarrow b$



Mivel S' kezdőszimbólum megjelenik a „piramis" tetején, így az $abab$ szó eleme a nyelvnek.

Szintaxisfa visszaépítés