

Formális nyelvek és automaták

Nagy Sára gyakorlatai alapján

Készítette: Nagy Krisztián

4. gyakorlat

2. megszorított grammatika/nyelv:

$$A \rightarrow q \quad A \in N \quad q \in (T \cup N)(T \cup N)^*$$

+ KES szabály

$$L_2 \text{ megsz} = L_2$$

ε mentesítés:

- Új kezdő szimbólum, melyből levezethető az S és az ε (KES szabály miatt):

$$S' \rightarrow \varepsilon$$

$$S' \rightarrow S$$

- $H = \{A \in N \mid A \xrightarrow{*}_G \varepsilon\}$

$H_1 = \{A \in N \mid \exists A \rightarrow \varepsilon\}$ Azon nem terminálisok tartoznak ebbe a halmazba, melyekből közvetlenül levezethető ε (Csak az $A \rightarrow \varepsilon$ típusú)

$H_2 = \{A \in N \mid \exists A \rightarrow x_1 \dots x_k \text{ és } x_i \in H_1 \forall i \in [1, k] \cup H_1\}$ Azon nem terminálisok tartoznak ebbe a halmazba, amelyek valamely H_1 halmazban levő nem terminálissal együtt szerepelnek a grammatikában, továbbá a H_1 halmaz elemei is.

.

.

.

$$H_i = H_{i+1} = H$$

(Példa:

$$A \rightarrow BC$$

$$B \rightarrow \varepsilon$$

$$C \rightarrow \varepsilon$$

$$H_1 = \{B, C\} \text{ (epszilon szabályok miatt)}$$

$$H_2 = \{A, B, C\}$$

Megjegyzés: Adódik, hogy maximum annyi H halmaz lehet, amennyi az adott grammatikában található nem terminálisok száma.

Folyamat: $A \rightarrow \gamma_1 B \gamma_2$

$B \in H$, akkor az alábbi átalakítást kell elvégezni:

$$A \rightarrow \gamma_1 B \gamma_2 \mid \gamma_1 \gamma_2 \text{ (kihagyjuk belőle az } \varepsilon\text{-t)}$$

Feladatok: (Epsilon mentesítés)

1. Feladat:

G_1 :

$S \rightarrow AB|b$

$A \rightarrow aB|AS$

$B \rightarrow AA|\varepsilon$

Írjuk fel a H halmaz elemeit.

1. lépés:

H_1 halmaz felírása: Nézzük meg miből vezethető le közvetlen az epsilon. Hol van epsilon szabály.

$S \rightarrow AB|b$

$A \rightarrow aB|AS$

$B \rightarrow AA|\varepsilon$

Ezért $H_1 = \{B\}$

2. lépés:

H_2 halmaz felírása: Nézzük meg hol található olyan szabály, melyből egy másik szabállyal levezethető az epsilon. Továbbá ne felejtjük el, hogy az előző halmaz elemei is beletartoznak ebbe a halmazba.

$S \rightarrow AB|b$

$A \rightarrow aB|AS$

$B \rightarrow AA|\varepsilon$

A késsel jelölt nem terminálisok is bekerülnek a H_2 halmazba, abban az esetben, ha $\varepsilon - t$ behelyettesítve egy újabb epsilon szabályt kapunk.

Jelen esetben az alábbi szabályokat kell megvizsgálni:

$S \rightarrow AB$

$A \rightarrow aB$

$A \rightarrow aB$ eset nem problémás, mivel B mellett egy terminális áll, ezért nem kapunk egy újabb epsilon szabályt.

Ennek tudatában pedig $S \rightarrow AB$ se problémás, mivel A olyan szabály, ami nem vezet epsilon-ra, így ez a szabály se fog epsilon-ra vezetni.

Ezek alapján: $H_2 = H_1 = \{B\} = H$

Most pedig alkalmazzuk az epsilon mentesítés / Folyamat pontjában leírt szabályunkat és írjuk fel epsilon mentes formában a nyelvet:

G_2 :

$S \rightarrow AB|b|A$

$A \rightarrow aB|AS|a$

$B \rightarrow AA$

Lényegében ez azt jelentette, hogy szabályba kell foglalni azokat az eseteket, amikor a H halmaz beli elemek eltűnnének.

$L(G_1) = L(G_2)$ és ezzel az áttalakítással elértük, hogy megszorított alakú legyen.

2. Feladat:

$S \rightarrow AB|b$

$A \rightarrow \varepsilon|AS$

$B \rightarrow AA|bbB$

Írjuk fel a H halmazzt:

$H_1 = \{A\}$

$S \rightarrow AB|b$

$A \rightarrow \varepsilon|AS$

$B \rightarrow AA|bbB$ miatt.

$H_2 = H_1 \cup \{B\} = \{A, B\}$

$S \rightarrow AB|b$

$A \rightarrow \varepsilon|AS$

$B \rightarrow AA|bbB$ miatt.

(Miért? – Azért, mert $B \rightarrow AA$ -ra az $A \rightarrow \varepsilon$ szabályt alkalmazzuk mind a két A-ra, akkor $B \rightarrow \varepsilon$ -t kapnánk)

$H_3 = H_1 \cup \{S\} = \{A, B, S\}$

$S \rightarrow AB|b$

$A \rightarrow \varepsilon|AS$

$B \rightarrow AA|bbB$ miatt.

(Miért? – Azért, mert $B \rightarrow AA$ -ra az $A \rightarrow \varepsilon$ szabályt alkalmazzuk mind a két A-ra, akkor $B \rightarrow \varepsilon$ -t kapnánk, továbbá ha az $A \rightarrow \varepsilon$ szabályt alkalmazzuk A-ra, akkor $S \rightarrow \varepsilon$ -t kapnánk.)

Mivel 3 db nem terminális található a szabályban, így biztos, hogy $H_3 = H$

Mivel a kezdő szimbólum (S) is be került a halmazunkba, így meg kell vizsgálnunk, hogy áll-e a kezdő szimbólum valahol a szabályok jobb oldalán. Amennyiben áll, úgy új kezdő szimbólumot kell bevezetni (KES szabály), majd alkalmazzuk az ε mentesítés / Folyamat pontjában leírt szabályunkat. Abban az esetben, ha nem áll a kezdő szimbólum valamelyik szabályunk jobb oldalán, akkor simán alkalmazzuk az ε mentesítés / Folyamat pontjában leírt szabályunkat és készen vagyunk.

A mi esetünkben „sajnos” $A \rightarrow AS$ miatt, áll kezdő szimbólum a szabályunk jobb oldalán, így új kezdő szimbólumot vezetünk be. Azért kell bevezetnünk, hogy ne veszítsük el magát az ε -t, mivel az üres szó is levezethető a jelenlegi nyelvünkben.

Így a szabályunk:

$S' \rightarrow \varepsilon | S$ (KES szabály miatt)

$S \rightarrow AB | B | A | b$ (Ki kellett egészítenünk azokkal az esetekkel, amikor vagy A-t vagy B-t tüntetjük el az ε miatt)

$A \rightarrow AS | S$ (Azért nincs A a szabály jobb oldalán, mert $A \rightarrow A$ folyton önmagába képező szabály.)

$B \rightarrow AA | A | b | bB | bb$

3. Feladat:

$S \rightarrow aSbS | bSaS | \varepsilon$

Írjuk fel a H halmazt:

$H_1 = \{S\}$

$S \rightarrow aSbS | bSaS | \varepsilon$

$H_2 = H_1 = H$

$S' \rightarrow \varepsilon | S$

$S \rightarrow aSbS | aSb | abS | ab | bSaS | bSa | baS | ba$

Mit csináltunk?

Mivel a kezdő szimbólum benne van a H halmazban és a szabályunk jobb oldalán is előfordult, ezért új kezdő szimbólumot vezettünk be, hogy ne vesszen el az üres szavunk (KES szabály). Ezek után pedig hozzáadtuk azokat a lehetőségeket, amikor az adott S kiesik.

- $L_{2\text{ megsz}} = L_2$ ($L_{2\text{ megsz}}$ esetén ε szabály megtiltásra kerül, de a KES szabály nem!)

3. típusú grammatikák ε -mentesítése

Emlékeztető: 3. típusú grammatika szabályrendszere:

$A \rightarrow uB$ ($A \in N, u \in T^*, B \in N$)

$A \rightarrow u$ ($A \in N, u \in T^*$) („záró szabály”)

Most: $A \rightarrow \varepsilon$ (~~epszilon szabály~~)

Működik a 2-es típusú nyelveknél vett ε -mentesítése, mivel ha megtiltjuk az epszilon szabályt, mikor eltöröljük B-t, akkor pont a záró szabályt kapjuk.

A következő példa egy második típusú nyelv és ϵ –mentesített változatát tartalmazza, majd ugyan ez a nyelv harmadik típusúra átírva és ϵ –mentesítve. Szemléltetés, hogy ugyan úgy működik a mentesítés mind a két nyelvtípusnál.

Második típusú:

$S \rightarrow BV$
 $V \rightarrow \epsilon|BV|AV$
 $B \rightarrow a|b|c| \dots |z$
 $A \rightarrow 0|1|2|3|4|5|6|7|8|9$

ϵ –mentesített:

$S \rightarrow BV|B$
 $V \rightarrow BV|B|AV|A$
 $B \rightarrow a|b|c| \dots |z$
 $A \rightarrow 0|1|2|3|4|5|6|7|8|9$

Harmadik típusú:

$S \rightarrow aV|bV| \dots |zV$
 $V \rightarrow \epsilon|aV|bV| \dots |zV|0V|1V| \dots |9V$

ϵ –mentesített:

$S \rightarrow a|b| \dots |z|aV|bV| \dots |zV$
 $V \rightarrow a|b| \dots |z|0|1| \dots |9$
 $V \rightarrow aV|bV| \dots |zV|0V|1V| \dots |9V$

BNF (Backus–Naur forma)

Szintaxisa:

<fogalom>	nem terminálisokra hivatkozik (< ... >)
::=	elválasztja a szabály jobb és bal oldalát
a	terminálisok
	karakterrel elválasztott sorozatok (vagy)

Megjegyzés: terminális nem szerepelhet a szabály bal oldalán.

(Egyéb dolgok a BNF-ről: http://hu.wikipedia.org/wiki/Backus%E2%80%93Naur_forma)

Feladatok: (BNF-es)

1. Feladat: Írjuk fel az alábbi grammatikát BNF-el!

$$S \rightarrow BV$$
$$V \rightarrow \varepsilon|BV|AV$$
$$B \rightarrow a|b|c| \dots |z$$
$$A \rightarrow 0|1|2|3|4|5|6|7|8|9$$

<azonosító> ::= <betű><vége>

<vége> ::= ε | <betű><vége> | <számjegy><vége>

<betű> ::= a | b | ... | z

<számjegy> ::= 0 | 1 | ... | 9

2. Feladat: Írjuk le a természetes számok halmazát BNF-el!

<természetes_szárok> ::= 0 | <nem_nulla><vége> | <nem_nulla>

<nem_nulla> ::= 1 | 2 | ... | 9

<vége> ::= <számjegy><vége> | <számjegy>

<számjegy> ::= 0 | 1 | 2 | ... | 9

3. Feladat:

Programozási nyelvekben:

<kif> ::= <tag> | <tag>+<tag>

<tag> ::= <faktor> | <faktor>*<faktor>

<faktor> ::= a | (<kif>) Megjegyzés: a – adat/ szám

Terminálisok: $T = \{+, *, a, (,)\}$

Egy fordító program dolga leellenőrizni, hogy jó-e egy adott szintaxis alapján az adott kifejezés.

(a) $a + (a * a)$ kifejezés helyes-e a fenti szintaxissal?

$\langle kif \rangle \rightarrow \langle tag \rangle + \langle tag \rangle \rightarrow \langle faktor \rangle + \langle tag \rangle \rightarrow a + \langle tag \rangle \rightarrow a + \langle faktor \rangle \rightarrow$
 $\rightarrow a + (\langle kif \rangle) \rightarrow a + (\langle tag \rangle) \rightarrow a + (\langle faktor \rangle * \langle tag \rangle) \rightarrow a + (a * \langle tag \rangle) \rightarrow$
 $\rightarrow a + (a * \langle faktor \rangle) \rightarrow a + (a * a) \checkmark$

Levezethető, tehát jó a kifejezés.

Grammatikával felírva:

$$K \rightarrow T | T + T$$
$$T \rightarrow F | F * T$$
$$F \rightarrow a | (K)$$

(b) $a + a + a$ kifejezés helyes-e a fenti szintaxissal?

Nem helyes a második + miatt.

Ha $(a + a) + a$ vagy $a + (a + a)$ lenne akkor viszont helyes lenne.

4. Feladat:

Írjuk le BNF-el a függvényeket.

Instrukciók:

- azonosítóval kezdődik
- 1 vagy több argumentuma van
- az argumentumokat vessző választja el
- az argumentum lehet azonosító vagy függvény kifejezés
- az azonosító betűkből áll

Például: $\sin(x)$

$f(g(h, y), x, z)$

Megoldás:

$\langle \text{függvény} \rangle ::= \langle \text{azonosító} \rangle (\langle \text{argumentumok} \rangle)$

$\langle \text{argumentumok} \rangle ::= \langle \text{argumentum} \rangle | \langle \text{argumentumok} \rangle \langle \text{argumentum} \rangle$

$\langle \text{argumentum} \rangle ::= \langle \text{azonosító} \rangle | \langle \text{függvény} \rangle$

$\langle \text{azonosító} \rangle ::= \langle \text{betű} \rangle | \langle \text{betű} \rangle \langle \text{azonosító} \rangle$

$\langle \text{betű} \rangle ::= a | b | \dots | z$

Gyakorló feladat: BNF-es 3. feladat szintaxisa alapján jó kifejezés-e az $(a + a) * a$?

Beadható feladat:

Írjuk le BNF-el az e-mail címet az alábbi instrukciók alapján!

- két részből áll, melyet @ választ el
- mind a két részben szavak találhatóak
- a szavak .-al vannak elválasztva
- az első részben legalább egy szó van (amennyiben több, ponttal vannak elválasztva)
- a második részben legalább két szó van (szintén ponttal elválasztva)
- a vége kettő vagy három karakterre végződik
- a szavak csak betűket tartalmazhatnak