Név (olvashatóan): Név (aláírás): EHA KÓD:

Formális nyelvek példadolgozat a vizsgára (mintamegoldással)

A maximálisam megszerezhető pontszám 62. A sikeres vizsga előfeltétele, hogy az első 5 kérdésből el kell érni legalább 17 pontot (a 34-ből). Az előfeltételt teljesítők osztályzata: [összpontszám/10], de legalább 1 és legfeljebb 5.

Figyelem: A teszt jellegű kérdéseknél indoklást nem kérünk, javítást viszont nem fogadunk el!!!!!

1. Melyik állítás igaz tetszőleges L, L1, L2, L3 nyelvek esetén az alábbi állítások közül? (Húzza alá a megfelelő választ! 2 pont kérdésenként, +1, ha mind jó, -1 a rosszért.)

I. $L_1 \cap L_2 = \emptyset \Rightarrow L_1L_3 \cap L_2L_3 = \emptyset$ igaz <u>nem igaz</u>

(Ellenpélda adható: $L_1 = \{\varepsilon\}$, $L_2 = \{a\}$, és $L_3 = \{\varepsilon, a\}$ nyelvekre $L_1 \cap L_2 = \emptyset$ és mégis $L_1L_3 \cap L_2L_3 = \{\varepsilon, a\} \cap \{a, aa\} = \{a\} \neq \emptyset$)

II. $(L_2^*L_1^*)^* = (L_1 \cup L_2)^*$ <u>igaz</u> nem igaz

*(Kihasználjuk a reguláris műveletek monotonitását és a * lezárási tulajdonságát, $(L^*)^* = L^*$ -ot.*
$\Rightarrow L_2^*L_1^* \subseteq (L_1 \cup L_2)^* (L_1 \cup L_2)^* \subseteq ((L_1 \cup L_2)^*)^* = (L_1 \cup L_2)^*$, ahonnan
$(L_2^*L_1^*)^* \subseteq ((L_1 \cup L_2)^*)^* = (L_1 \cup L_2)^*$,
*$\Leftarrow$ Világos, hogy $L_1 \subseteq L_2^*L_1^*$ és $L_2 \subseteq L_2^*L_1^*$, ahonnan $L_1 \cup L_2 \subseteq L_2^*L_1^*$, amiből * monotonitása miatt az állítás következik)*

III. $\overline{L}^{\,+} = \overline{L - \{\varepsilon\}}^{\,*}$ igaz <u>nem igaz</u>

(Ha $\varepsilon \in L$, akkor $\varepsilon \notin \overline{L}^{\,+}$, de $\overline{L - \{\varepsilon\}}^{\,*}$ mindig tartalmazza az ε –t.)

2. Adja meg a következő definíciók, jelölések jelentését (kérdésenként 1 pont, +1 ha mind jó)!
a. L(G), ahol G formális nyelvtan a standard jelöléssel

$L(G) = \{u;\ u \in T^* \text{ és } S \xrightarrow[G]{*} u\}$

b. VDA-k esetében a $\hat{\delta}: Q \times T^* \rightarrow Q$ kiterjesztett állapot-átmeneti függvény

Rekurzívan: $\hat{\delta}(q, \varepsilon) = q$; $\hat{\delta}(q, vt) = \delta(\hat{\delta}(q, v), t)$, ahol $q \in Q$, $v \in T^*$, $t \in T$ tetszőlegesek.

Iteratívan: Legyenek $q \in Q$, $v \in T^*$, $k := l(v)$, $v = t_1t_2\ldots t_k$. $\hat{\delta}(q, v) = q'$, ha léteznek olyan $c_0, c_1, \ldots c_k$ állapotok, hogy $c_0 = q$, $c_k = q'$ és minden $1 \leq j \leq k$ –ra $c_j = \delta(c_{j-1}, t_j)$)

c.: $\mathcal{L}_{\text{RekFel}}$

Név (olvashatóan): Név (aláírás): EHA KÓD:

Olyan nyelvek osztálya, amelyekhez létezik a nyelvet felsoroló algoritmus. A felsorolt nyelv az algoritmus által az output változóba kiírt szavak összessége.

d.: $q \overset{i}{\sim} q'$, ahol q,q' egy VDA állapotai

$q \overset{i}{\sim} q' \Leftrightarrow \forall r \in T^{\leq i} (\delta(q, u) \in F \Leftrightarrow \delta(q', u) \in F)$

e.: $L^{\varepsilon}(V)$, ahol V egy (1)veremautomata a standard jelöléssel

$L^{\varepsilon}(V) = \{u;\ u \in T^*$ és $[q_0, u, \sigma_0] \underset{V}{\to}^* [q, \varepsilon, \varepsilon]$ valamely $q \in Q$ esetén$\}$

3. Adja meg a következő konstrukciókat, képleteket (3 pont kérdésenként, +1, ha mind jó)!
a. Az L reguláris nyelvhez készített A_*^L automata:

Egy $L \subseteq T^*$ nyelv valamely $p \in T^*$ szóra vonatkozó maradéknyelve: $L_p := \{v, pv \in L\}$.
Tulajdonságai: 1. $L_\varepsilon = L$; 2. $\varepsilon \in L_p \Leftrightarrow p \in L$; 3. $(L_p)_r = L_{pr}$.
MYHILL-NERODE tétele szerint a maradéknyelveinek halmazára teljesül $|\{L_p; p \in T^*\}| < \infty$. Ez a halmaz lesz az alább definiált A_*^L VDA állapothalmaza:
$A_*^L = \langle\{L_p; p \in T^*\}, T, \delta, L_\varepsilon, \{L_p; \varepsilon \in L_p\}\rangle$, ahol $\delta(L_p, t) = L_{pt}$. A maradéknyelvek 2. tulajdonságait kihasználva belátható, hogy A_*^L jól definiált és $L = L(A_*^L)$.

b. A Kleene-tétel bizonyítása során használt $L_{i,j}^k$ nyelvek és rekurzív előállításuk:
Legyen $A = \langle\{q_1, q_2, \ldots, q_n\}, T, \delta, q_1, F\rangle$ valamely VDA. A $0 \leq i,j,k \leq n$ és $i,j \neq 0$ indexek mellett képezzük az $L_{i,j}^k = \{u;\ \delta(q_i, t) = q_j$ és a közbülső állapotok indexe $\leq k\}$ definícióval adott halmazokat. Ezeket a következő rekurzióval állíthatjuk elő:
$L_{i,j}^0 = \{t;\ \delta(q_i, t) = q_j\} \cup \Delta_{i,j}$, ahol $\Delta_{i,j} = \{\varepsilon\}$, ha i=j, egyébként pedig $\varnothing$.
$L_{i,j}^k = L_{i,j}^{k-1} \cup L_{i,k}^{k-1} (L_{k,k}^{k-1})^* L_{k,j}^{k-1}$

c. Egy G=<T,N, *P*, S> 2. típusú nyelvtan feletti szintaxisfa:

G-feletti szintaxisfán olyan fát értünk, melynek
1. csúcsai a $T \cup N \cup \{\varepsilon\}$ halmaz elemeivel címkézettek,
2. belső pontjai N elemeivel címkézettek,
3. ε címkéjű pontnak nincsen testvére,
4. ha egy pont címkéje Z, közvetlen leszármazottjaiban balról jobbra rendre az $Z_1, Z_2, \ldots, Z_l$ címkék vannak, akkor $Z \to Z_1 Z_2 \ldots Z_l$ szabály G-ben.

4. Mutassa meg, hogy a 3. típusú nyelvek osztálya zárt a konkatenáció műveletére nézve (5 pont)!

Azt kell kimutatnunk, hogy ha L_1 és L_2 tetszőleges harmadik típusú nyelvek, akkor L_1L_2 is az.
Legyenek i=1,2 mellett $G_i = \langle T_i, N_i, P_i, S_i\rangle$ az L_i –t generáló 3. típusú nyelvtanok, ahol $N_1 \cap N_2 = \varnothing$. A következő alakú G_{konk} 3. típusú nyelvtan az L_1L_2 nyelvet fogja generálni:

$G_{konk} = \langle T_1 \cup T_2, N_1 \cup N_2, \Phi(P_1, S_2) \cup P_2, S_1 \rangle$, ahol $\Phi(P_1, S_2)$ lényegében megegyezik P_1 –el, azzal az eltéréssel, hogy az $A \to v \in P_1$, $v \in T_1{}^*$ alakú termináló szabályok helyett $A \to v\, S_2$ szabály van benne.

5. Milyen a CYK algoritmus és mire használjuk? (6 pont)

A CYK algoritmust Chomsky-normálformájú 2. típusú nyelvtanok esetében a szóprobléma eldöntésére használjuk. A Chomsky-normálforma szabályalakjai: $S \to \varepsilon$ KES, $A \to BC$, $A \to t$.
Legyen $G = \langle T, N, P, S \rangle$ egy Chomsky-normálformájú nyelvtan és $u \in T^*$ tetszőleges szó.
Eldöntendő, hogy $u \in L$ vagy sem. Az eldöntő algoritmus a következő:
a. $u = \varepsilon$ esetében $u \in L \Leftrightarrow S \to \varepsilon \in P$, ezért elegendő $S \to \varepsilon \in P$ -t ellenőrizni.
b. Legyen $u = t_1 t_2 \ldots t_k$, $k \geq 1$.
Számozzuk meg P elemeit az $A_i \to q_i$ alakban, $i = 1,2,\ldots,n$. és képezzük a $H_{i,j}$ halmazokat ($1 \leq i \leq j \leq k$) a következő módon:
$H_{i,i} = \{A_x;\ q_x = t_i\}$, $H_{i,j} = \{A_x;\ q_x \in \bigcup_{k=i}^{j-1} H_{i,k} H_{k+1,j}\}$, ha $i < j$.
A szabályok alakjából és a képzés módjából következik, hogy $A_x \in H_{i,j} \Leftrightarrow A_x \underset{G}{\to^*} t_i t_{i+1} \ldots t_j$.
Ennek alapján $S \in H_{1,k} \Leftrightarrow S \underset{G}{\to^*} t_1 t_2 \ldots t_k \Leftrightarrow u \in L(G)$, ezért elegendő $S \in H_{1,k}$ -t ellenőrizni.

6. Az alább felsorolt nyelvek esetén azt a legnagyobb típust kell megnevezni, amilyen típusú nyelvtannal az adott nyelvet biztosan generálni tudja. (Húzza alá a megfelelő választ! 2 pont kérdésenként, -1 a rossz válasz.)

a. $\{a^n;\ n = k^2$, ahol $0 \leq k \leq 1000\}$ **0 1 2 $\underline{3}$**
(a nyelv véges, tehát 3. típusú)

b. 2 zárójelpár feletti helyes zárójelezések nyelve **0 1 $\underline{2}$ 3**
(Legyenek a zárójelek például {,}, [,]. A kis Bar-Hillel lemmával bizonyítottuk, hogy az egy zárójelpár feletti helyes zárójelezések nyelve nem harmadik típusú. Ez a bizonyítás változatlanul működik két zárójelpár esetében is. Az alábbi 2. típusú szabályrendszer viszont előállítja őket:
$S \to \{S\} \mid [S] \mid SS \mid \varepsilon$)

c. $\{x^n y^n z^n\ ;\ n \geq 1\}$ **0 $\underline{1}$ 2 3**
(A nagy Bar_Hillel lemmával bizonyítottuk, hogy az $\{\alpha\alpha\ ;\ \alpha \in T^*\}$ dadogós nyelv ($|T| \geq 2$) nem 2. típusú. Ez a bizonyítás analóg módon most is működik. Az alábbi 1. típusú szabályrendszer viszont generálja:
$S \to xSYz \mid xYz$; $zY \to Yz$; $xY \to xy$; ; $yY \to yy$)

d. Tetszőleges parciálisan rekurzív nyelv **$\underline{0}$ 1 2 3**
(Tudjuk, hogy $\mathcal{L}_{ParRek} = \mathcal{L}_0$ és $\mathcal{L}_0 \supset \mathcal{L}_1$)

7. **Mely nyelveket generálják az alábbi nyelvtanok (csak a szabályokat adjuk meg, a nagy betűk a nyelvtani jelek, a kicsik a terminálisok, S a kezdőjel)? Indokolja röviden a válaszát (6, illetve 4 pont)!**

a. $S \to LA$; $A \to zZA|zZ|yYA|yY$; $zZ \to Zz$; $zY \to Yz$; $yZ \to Zy$; $yY \to Yy$; $LZ \to zL$; $LY \to yL$; $L \to \varepsilon$.

$L_1 = \{vv;\ v \in \{y,z\}^+\}$, vagyis az $\{y,z\}$ ábécé feletti, nem üres dadogós szavakat tartalmazó nyelv.

Indoklás: Az A nyelvtani jel eltűnése után a mondatformánk $L\{zZ,yY\}^+$ alakú lesz. Legyenek v, illetve V azok a szavak, melyeket a terminálisok (kis betűk), illetve nyelvtani jelek (nagy betűk) összeolvasásával kapunk. Ezek nyilván egymás kis és nagybetűs változatai. Mivel Z és Y csak az L mellett alakulhat át a neki megfelelő terminálissá, mindegyiküket muszáj az L mellé balra cserélgetni. Ezért L előtt a V kisbetűs változata, v keletkezik. L-et el lehet hagyni, de csak a teljes V v-re való cseréje után, mert különben a szóban maradna nyelvtani jel.

b. $S \to Dd|Yy$, $D \to Yy|cy$, $Y \to Dd|cd$.

$L_2 = \{cv;\ v \in \{d,y\}^*,\ l(v) \geq 2$ és $dd \not\subseteq v$ és $yy \not\subseteq v)$ vagy reguláris kifejezéssel
$L_2 = c(d+\varepsilon)(yd)^+(y+\varepsilon)$.

Indoklás: Az S-ből legalább 1 lépésben generált, nyelvtani jelet még tartalmazó szavak alakja Zv, ahol $pre(v,1) = z$, $z \in \{d,y\}$, Z a z nagybetűs változata, továbbá $dd \not\subseteq v$ és $yy \not\subseteq v$.
Az S-re vonatkozó szabályok alakja alapján az 1 hosszú levezetésekre ez biztosan igaz, míg amiatt, hogy D, illetve Y jobboldalai rendre a másiknak megfelelő kisbetűkkel végződnek, a tulajdonság öröklődik. Az utolsó lépés is lényegében megtartja az előbbi tulajdonságot, csak a nyelvtani jel helyett a baloldalon c jön be. A legrövidebb levezetés is legalább két lépésből áll, ami legalább 3 terminális jelet hoz be.

8. Bizonyítsa be a kis Bar-Hillel lemmát(3 pont a helyes kimondás, 7 a teljes bizonyítás)!

Tétel (kis Bar-Hillel lemma) Tetszőleges L 3. típusú nyelv esetében van olyan $n=n(L)>0$ egész, hogy valahányszor $u \in L$ és $l(u) \geq n$, akkor u felírható az $u=xyz$ alakban a következő tulajdonságokkal: 1. $l(y)>0$, 2. $l(xy) \leq n$, 3. $\forall i = 0,1,\ldots$ esetén $xy^iz \in L$.

Bizonyítás: Mivel L 3. típusú, ezért létezik olyan $A=\langle T,Q,\delta,q_0,F\rangle$ VDA, melyre $L(A) = L$. Legyen $n := |Q|$, mely valóban egész, 0-nál nagyobb és nyilván függ L-től (mert maga A függ L-től). Tekintsünk egy $u=t_1t_2\ldots t_m$ L-beli szót, melyre $m \geq n$. Legyen ($q_0=$) $c_0, c_1, c_2, \ldots, c_m$ ($\in F$) az u felismerése során érintett állapotok sorozata. $c_0, c_1, c_2, \ldots, c_n$ között a skatulya-elv alapján van két állapot, melyek egyenlők (Q –ban csak n különböző állapot van). Legyen ezek indexe $0 \leq j < k \leq n$, azaz $c_j=c_k$. Ekkor az $x:= t_1t_2\ldots t_j$, $y:= t_{j+1}t_{j+2}\ldots t_k$ és $z:= t_{k+1}t_{k+2}\ldots t_m$ szavakra nyilván $u=xyz$, $l(y)=k-j > 0$, $l(xy)=k \leq n$, továbbá $\delta(c_0,x)=c_j$, $\delta(c_j,y)=c_k=c_j$, $\delta(c_k,z)=c_m$. Tetszőleges $i = 0,1,\ldots$ mellett az xy^iz szó hatására az automata ugyanazt az utat fogja bejárni, mint xyz, csak a $\delta(c_j,y)= c_j$ hurkot i-szer. Emiatt $\delta(c_0, xy^iz) = c_m$. Mivel $q_0= c_0$ és $c_m \in F$ ez azt jelenti, hogy $xy^iz \in L$.