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

Formális nyelvek vizsgadolgozat 2011. december 28.
Megoldás

A maximálisan 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: [összesített pontszám/10], de legalább 1 és legfeljebb 5.

Figyelem: A teszt jellegű kérdéseknél indoklást nem kérünk, javítást viszont nem fogadunk el**!!!!!**
Kérjük, csak a válasznak kihagyott üres helyre írjanak!

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

I. $L_1 \cap L_2 = \emptyset \Rightarrow L_1L \cap L_2L = \emptyset$ **igaz** **nem igaz**

(Ellenpélda adható: $L_1 = \{\varepsilon\}$, $L_2 = \{a\}$, és $L = \{\varepsilon, a\}$ nyelvekre $L_1 \cap L_2 = \emptyset$ és mégis $L_1L \cap L_2L = \{\varepsilon, a\} \cap \{a, aa\} = \{a\} \neq \emptyset$.)

II. $(L_2^* L_1^*)^* = (L_1 \cup L_2)^*$ **igaz** **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. $\overline{L^*} = \overline{L}^*$ **igaz** **nem igaz**

(Itt $\overline{L^*}$ nem tartalmazza ε –t (hiszen L^* tartalmazza), míg $\overline{L}^*$ definíció szerint igen.)

2. Adja meg a következő definíciók, jelölések jelentését (kérdésenként 1 pont, +1, ha mind jó)!

a. 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 $c_j = \delta(c_{j-1}, t_j)$.

b. L(A), ahol A véges, standard jelölésű nemdeterminisztikus automata:
$L(A) = \{u;\ u \in T^* \text{ és } \delta(q_0, u) \cap F \neq \emptyset\}$

c.: $\mathcal{L}_{\text{RekFel}}$:
Olyan nyelvek osztálya, amelyekhez létezik a nyelvet felsoroló algoritmus. A felsorolt nyelv a felsoroló algoritmus által a futás során a szó típusú 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.: $L^F(V)$, ahol V egy (1)veremautomata a standard jelöléssel:
$L^F(V) = \{u;\ u \in T^*$ és $[q_0, u, \sigma_0] \underset{V}{\to}^* [f, \varepsilon, \alpha]$ valamely $f \in F$ és $\alpha \in \Sigma^*$ 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 előbbi tulajdonságait kihasználva belátható, hogy A_*^L jól definiált és $L = L(A_*^L)$.

b. A CYK algoritmusban használt $H_{i,j}$ halmazok és rekurzív előállításuk:
A CYK algoritmust Chomsky-normálformájú 2. típusú nyelvtanok esetében a szóprobléma eldöntésére használjuk. Legyen $G=\langle T,N, P, S\rangle$ az adott Chomsky-normálformájú nyelvtan, és számozzuk meg szabályait az $A_k \to q_k$ alakban ($k= 1,2,\ldots,|P|$). Legyen az adott terminális szó $u= t_1t_2 \ldots t_n$, $n\geq 1$.
A $H_{i,j}$ halmazokat ($1\leq i\leq j\leq n$) a következő módon definiáljuk: $H_{i,j} = \{A;\ A \underset{G}{\to}^* t_it_{i+1}\ldots t_j.\}$.
A rekurzív előállítás:
$H_{i,i} = \{A_k;\ q_k = t_i\}$, $H_{i,j} = \{A_k;\ \exists r\in[i\ldots j-1], B\in H_{i,r}, C\in H_{r+1,j}\ (q_k = BC)\}$, ha $i < j$.
(Ez utóbbi a $H_{i,j} = \{A_k;\ q_k \in \bigcup_{r=i}^{j-1} H_{i,r}H_{r+1,j}\}$ tömörebb alakban is felírható.)

c. Egy G=<T,N, *P*, S> 2. típusú nyelvtan feletti szintaxisfa:
G-feletti szintaxisfán olyan nem üres 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 $Z_1, Z_2, \ldots, Z_k$ címkék vannak, akkor $Z\to Z_1 Z_2 \ldots Z_k$ szabály a G-ben.

4. Adja meg azt a konstrukciót, mely bizonyítja, hogy a 3. típusú nyelvek osztálya zárt a lezárás (*) műveletére nézve (5 pont)!
Azt kell kimutatnunk, hogy ha L tetszőleges harmadik típusú nyelv, akkor L^* is az.
Legyen $G = \langle T, N, P, S\rangle$ egy L – et generáló 3. típusú nyelvtan. A következő alakú G_* 3. típusú nyelvtan az L^* nyelvet fogja generálni:
$G_* = \langle T, N\cup\{S'\}, P\cup\Phi(P, S')\cup\{S'\to S|\varepsilon\}, S'\rangle$, ahol $\Phi(P, S')$ lényegében megegyezik P – vel, azzal az eltéréssel, hogy az $A \to v \in P$, $v \in T^*$ alakú termináló szabályok helyett az $A \to v S'$ szabály van benne. .

5. Mit ért a véges determinisztikus automaták analízis feladatán? Kleene tételének bizonyítását alapul véve javasoljon módszert az analízis feladat megoldására! (6 pont)
Véges determinisztikus automaták analízisén egy olyan reguláris kifejezés előállítását értjük, mely az adott VDA által felismert nyelvet írja le. Kleene tétele alapján ilyen reguláris kifejezés biztosan

létezik. Az említett tétel bizonyításának egyik iránya pontosan egy ilyen reguláris kifejezés alábbi előállításával történik.

Legyen A= <{q_1, q_2,…,q_n},T, δ, q_1, F> az adott VDA.

Képezzük rendre a felső index szerint növekvő sorrendben a következő $L^k_{i,j}$ nyelveket ($0 \le i,j,k \le n$ és $i,j \ne 0$) :

$L^0_{i,j} = \{t;\ \delta(q_i, t) = q_j\} \cup \Delta_{i,j}$, ahol $\Delta_{i,j} = \{\varepsilon\}$, ha i=j, egyébként pedig $\emptyset$.

$L^k_{i,j} = L^{k-1}_{i,j} \cup L^{k-1}_{i,k} (L^{k-1}_{k,k})^* L^{k-1}_{k,j}$

Ekkor L(A) = $\bigcup_{q_j \in F} L^n_{1,j}$, azaz megkapjuk L(A) egy reguláris kifejezését.

.

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

a. 3 zárójelpár feletti helyes zárójelezések nyelve **0 1 <u>2</u> 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 három zárójelpár esetében is. Az alábbi 2. típusú szabályrendszer viszont előállítja őket:
S → (S)| {S}| [S] | SS | ε)

b. {a^n; ahol n 0 és 1000 közé eső prímszám} **0 1 2 <u>3</u>**

(A nyelv nyilván véges, emiatt 3. típusú)

c. {$x^{n-2}y^{n-2}z^{n-2}$; n ≥3 } **0 <u>1</u> 2 3**

(A nagy Bar_Hillel lemmával bizonyítottuk, hogy az {αα ; α∈ T} dadogós nyelv (|T| ≥ 2) nem 2. típusú. Ez a bizonyítás analóg módon most is működik. Az alábbi 1. típusú szabályrendszer viszont generálja:*
S → xSYz | xYz ; zY→Yz; xY→ xy; ; yY→ yy)

d. Tetszőleges rekurzívan felsorolható nyelv **<u>0</u> 1 2 3**

(Tudjuk, hogy a rekurzívan felsorolható nyelvek osztálya megegyezik a 0. típusú nyelvek osztályával, mely a Chomsky-féle erős hierarchia tétel alapján bővebb, mint az 1. típusú nyelvek osztálya.)

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 (5, illetve 4 pont)!

a. S→BDcJ, Dc→ccccD, DJ→EJ, cE→Ecccc, BE→BD, BD→ε, EJ→ε, B→ε, J→ε.

A generált nyelv: {c^n; $n=4^k$, ahol k=0,1,2,….}

Indoklás: Ahhoz, hogy terminális szót kapjunk, el kell „tüntetnünk” a mondatformában lévő nyelvtani jeleket. A bal, illetve jobb szélt jelölő B és J jelek bármikor eltüntethetők a megfelelő epszilon szabállyal. D a B-hez ragadva a bal szélen, E pedig a J-hez ragadva a jobb szélen tűnhet el. Az S→BDcJ kezdés után azonnal az eltüntető szabályokat alkalmazva egy darab c-t kapunk. További terminális szavakat a D-nek a c betűkön való, J-ig történő átugratásával, illetve E-nek a c betűkön való, B-ig történő átugratásával kaphatunk. Mindkét esetben négyszereződik a c-k száma, hiszen minden átugrás egy már meglevő c helyett 4 darab c-t hoz be.

b. S → SBa, S→Y, YB→bY, aB→Ba, YB→bc

A generált nyelv: $\{b^n ca^n;\ n = 1,2,\dots\}$

Indoklás: Számozzuk az adott sorrendnek megfelelően 1-től 5-ig szabályainkat! Ahhoz, hogy terminális szót kapjunk, az S nyelvtani jelet a 2. (S→Y) szabállyal egyszer át kell írnunk Y-ra. A levezetésben ez előtt 1.-et és 4.-et használhattuk vegyesen, ezért a mondatforma a 2. szabály alkalmazása után $Y\alpha$ alakú lesz, ahol $\alpha\in\{a,B\}^+$ és benne ugyanannyi B van, mint a. A B-ket csak a szó baloldalán, Y-hoz tapadva tudjuk a 3. szabállyal b terminálissá átírni, melyek ezután az Y előtt halmozódnak fel. Az Y-t csak B mellett, mégpedig pontosan az utolsó B-vel együtt lehet csak az 5. szabállyal átírni (különben vagy B, vagy Y eltávolíthatatlanul maradna benne). Ez az átírás mindig hoz be egy b-t, és a szó közepére pedig egy c-t.

8. Bizonyítsa be MYHILL-NERODE tételét (4 pont a szükséges definíciók és a helyes kimondás, 7 a teljes bizonyítás)!

Közvetlenül a tételhez köthető definíciók:
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.) $(L_p)_r = L_{pr}$.
Egy $A=\langle Q,T,\delta,q_0,F\rangle$ VDA $q\in Q$ állapotra vonatkozó maradéknyelve:
$L(A,q) := \{v,\ L(A,q) \in F\}$.
Tulajdonságai: 1.) $L(A,q_0) = L(A)$; 2.) $\varepsilon \in L(A,q) \Leftrightarrow q \in F$; 3.) $L(A,q)_r = L(A, \delta(q,r))$

Tétel (MYHILL-NERODE): $L \in \mathcal{L}_3 \Leftrightarrow |\{L_p;\ p \in T^*\}| < \infty$
Bizonyítás:
$\Leftarrow$ Elég belátnunk, hogy L-hez van őt felismerő VDA.
Mivel feltételünk szerint $|\{L_p;\ p \in T^*\}| < \infty$, képezhetjük a következő A_*^L VDA –t:

$A_*^L = \langle\{L_p;\ p \in T^*\}, T, \delta, L_\varepsilon, \{L_p; \varepsilon \in L_p\}\rangle$, ahol $\delta(L_p, t) = L_{pt}$.

A maradéknyelvek tulajdonságait használva belátható, hogy A_*^L jól definiált és $\delta(L_\varepsilon, u) = L_u$.
Ez utóbbit és a 2. tulajdonságot használva:
$u \in L(A_*^L) \Leftrightarrow \varepsilon\in\delta(L_\varepsilon, u) \Leftrightarrow \varepsilon\in L_u \Leftrightarrow u \in L$, tehát $L = L(A_*^L)$.
$\Rightarrow$ Legyen $A=\langle Q,T,\delta,q_0,F\rangle$ L-et felismerő VDA ($\mathcal{L}_3 = \mathcal{L}_{DA}$ miatt ilyen mindig van).
Tetszőleges $p \in T^*$ szóra – az automata maradéknyelv tulajdonságait használva - számítsuk ki L_p-t: $L_p = L(A)_p = L(A,q_0)_p = L(A, \delta(q_0,p))$.
Ezt felhasználva felírhatjuk a következő összefüggést:
$\{L_p;\ p \in T^*\} = \{L(A, \delta(q_0, p));\ p \in T^*\} \subseteq \{L(A,q);\ q \in Q\}$
Itt áttérve a halmazok elemszámára:
$|\{L_p;\ p \in T^*\}| \le |\{L(A,q);\ q \in Q\}| \le |Q| < \infty$ (itt felhasználtuk, hogy egy VDA - nak nyilván legfeljebb annyi maradéknyelve lehet, amennyi az állapotainak száma).