**Formális nyelvek példadolgozat a vizsgára**

**(mintamegoldással)**

A maximálisam megszerezhető pontszám 62. Az osztályzat kialakításának algoritmusa:  
osztályzat = összesített pontszám tizedének egész része, de legalább 1 és legfeljebb 5.

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

*(Ebben a mintamegoldásban a teszt jellegű kérdésekre adott válaszokat is indokoljuk, dőlt betűvel szedve.)*

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

**I. L1∩L2=∅ ⇒ L1L3∩L2L3=∅ igaz nem igaz**

*(Ellenpélda adható: L1= {ε}, L2= {a}, és L3= {ε, a} nyelvekre L1∩L2=∅ és mégis*

*L1L3∩L2L3={ε, a}∩{a, aa}= {a}≠ ∅)*

**II. (L2\*L1\*)\* = (L1∪L2)\* igaz nem igaz**

*(Kihasználjuk a reguláris műveletek monotonitását és a \* lezárási tulajdonságát, (L\*)\*=L\* -ot.*

*⇒ L2\*L1\* ⊆ (L1∪L2)\* (L1∪L2)\* ⊆ ( (L1∪L2)\*)\*= (L1∪L2) \*, ahonnan*

*(L2\*L1\*)\* ⊆ ((L1∪L2)\*)\* = (L1∪L2)\*,*

*⇐ Világos, hogy L1⊆ L2\*L1\* és L2⊆ L2\*L1\*, ahonnan L1∪L2⊆ L2\*L1\*, amiből \* monotonitása miatt az állítás következik)*

**III. igaz nem igaz**

*(Ha ε∈L, akkor , de mindig tartalmazza az ε –t.)*

**2. Adja meg a következő jelölések jelentését (kérdésenként 2 pont)!**

**a. L(G), ahol G formális nyelvtan a standard jelöléssel**

L(G) = {u; u∈T\* és }

**b. VDA-k esetében a kiterjesztett állapot-átmeneti függvény**

Rekurzívan: ; , ahol q ∈ Q, v ∈ T\*, t∈T tetszőlegesek.

Iteratívan: Legyenek q ∈ Q, v ∈ T\*, k:= l(v), v= t1t2…tk. , ha léteznek olyan c0,c1,…ck  állapotok, hogy c0 = q, ck = q’ és minden 1 ≤ j ≤ k –ra cj = δ(cj-1,tj) )

**c.: L RekFel**

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

**d.: ~ , ahol q,q’ egy VDA állapotai**

⇔ ∀u ∈ T\* (δ(q, u) ∈ F ⇔δ(q’, u) ∈ F)

**e.: L(V), ahol V egy vermes veremautomata a standard jelöléssel**

L(V) = {u; u∈T\* és

**3. Adja meg a következő konstrukciókat, képleteket (4 pont kérdésenként)!**

**a. G=<T,N, *P*, S> 2. típusú nyelvtan epszilon mentesítése során használt H halmaz:**

H0:= {A; A→ε ∈ ***P*** }

Hi+1= Hi ∪{A; ∃Q∈ Hi\* A→Q∈ ***P*** }

H=Hj az első olyan j-re, melyre Hj+1= Hj.

**b. A harmadik típusú nyelvek osztályának a konkatenálás műveletére vonatkozó zártságát bizonyító nyelvtani konstrukció:**

Azt kell kimutatnunk, hogy ha L1 és L2 tetszőleges harmadik típusú nyelvek, akkor L1L2 is az. Legyenek i=1,2 mellett Gi = <T, Ni, Pi ,Si> az Li –t generáló 3. típusú nyelvtanok, ahol N1 ∩ N2 =∅. A következő alakú Gkonk 3. típusú nyelvtan az L1L2 nyelvet fogja generálni:

Gkonk= <T, N1 ∪ N2, Φ(P1 ,S2)∪ P2, S1> , ahol Φ(P1 ,S2) lényegében megegyezik P1 –el, azzal az eltéréssel, hogy az A → v ∈ P1, v ∈ T\* alakú termináló szabályok helyett az A → v S2 szabály van benne.

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

G-feletti szintaxisfán olyan fát értünk, melynek

1. csúcsai a T∪N∪{ε} halmaz elemeivel címkézettek,

2. belső pontjai N elemeivel címkézettek,

3. ε címkéjű pontnak nincsen testvére,

4. ha egy pont címkéje A, közvetlen leszármazottjaiban balról jobbra a q szó olvasható, akkor A→ qszabály G-ben.

**4. 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∈L és l(u)≥n, akkor u felírható az u=xyz alakban a következő tulajdonságokkal: 1. l(y)>0, 2. l(xy) ≤ n, 3. ∀i = 0,1,… esetén xyiz ∈ L.

Bizonyítás: Mivel L 3. típusú, ezért létezik olyan A=<T,Q,δ,q0,F> 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=t1t2…tm L-beli szót, melyre m≥n. Legyen (q0=) c0, c1, c2, … , cm (∈ F) az u felismerése során érintett állapotok sorozata. c0, c1, c2, … , cn között a skatulya-elv alapján van két állapot, melyek egyenlők (Q –ban csak n különböző állapot van). Legyen ezek indexe 0 ≤ j <k ≤ n, azaz cj=ck. Ekkor az x:= t1t2…tj, y:= tj+1tj+2…tk és z:= tk+1tk+2…tm szavakra nyilván u=xyz, l(y)=k-j > 0, l(xy)=k ≤ n, továbbá δ(c0,x)=cj , δ(cj,y)=ck=cj , δ(ck,z)=cm. Tetszőleges i = 0,1,… mellett az xyiz szó hatására az automata ugyanazt az utat fogja bejárni, mint xyz, csak a δ(cj,y)= cj hurkot i-szer. Emiatt δ(c0, xyiz) = cm. Mivel q0= c0 és cm ∈ F ez azt jelenti, hogy xyiz ∈ L.

**5. 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! 3 pont kérdésenként.)**

**a. {an; n=k2, ahol 0≤ k≤ 1000 } 0 1 2 3**

*(a nyelv véges, végesekre bizonyítottuk, hogy felismerhetők VDA-val, tehát 3. típusú)*

**b. 2 zárójelpár feletti helyes zárójelezések nyelve 0 1 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 → {S} | [S] | SS | ε )*

**c. {xnynzn ; n ≥1} 0 1 2 3**

*(A nagy Bar\_Hillel lemmával bizonyítottuk, hogy a {vv ; v∈ 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 parciálisan rekurzív nyelv 0 1 2 3**

*(Tudjuk, hogy L ParRek = L 0 és a Chomsky erős hierarchia miatt L 0 ⊃ L 1)*

6. **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 (7, illetve 5 pont)!**

**a. S→LA; A→zZA|zZ|yYA|yY; zZ→Zz; zY→Yz; yZ→Zy; yY→Yy;**

**LZ→zL; LY→yL; L→ε.**

L1 = {vv; v∈{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 → Dd|Yy, D→Yy|cy, Y→Dd|cd.**

L2 = {cv; v∈ {d,y}\*, l(v)≥2 és ) vagy reguláris kifejezéssel

L2 = c(d+ε)(yd)+(y+ε).

Indoklás: Az S-ből legalább 1 lépésben generált, nyelvtani jelet még tartalmazó szavak alakja Zv, ahol pre (v,1) = z, z∈{d,y}, Z a z nagybetűs változata, továbbá .

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.