Kérem e-mail-ben jelezze, ha hibát talál: (veanna@inf.elte.hu, vagy veanna@elte.hu)

1. feladat

$L_1 = \{ab,ba,b\}$ $L_2=b^*ab^*$ $L_3 = \{a, bb, aba\}$

$L_1L_3 = \{aba, abbb, ababa, baa, babb, baaba, ba, bbb, baba\}$
$(ab \mid b)^*$ -ra illő szavak: abbb, babb, bbb

L_2 szavai 4 hosszig: a, ba, ab, bba, bab, abb, bbba, bbab, babb, abbb

$L_1 \cup L_2 = L_2 \cup \{b\}$ vagy reguláris kifejezéssel: $b^*ab^* \mid b$
$L_2 \setminus L_1 = \{u \in \{a,b\}^* \mid (l(u) \geq 3$ és $u=b^nab^m$ $n,m\geq 0)$ vagy $u=a\}$ vagy reguláris kifejezéssel:
$a \mid b^*abb^+ \mid b^+ab^+ \mid b^+bab^*$
$L_2 \cap L_1^* =$ L2 \ {a} vagy reguláris kifejezéssel. $b^+ab^* \mid b^*ab^+$

2. feladat

ε mentesítse a következő 2-es típusú nyelvtant:

$G=<\{a,b,c\}, \{S,A,B,C,D\}, \mathbf{P}, S>$
P={ $S\to aAS \mid AD \mid CSC$
$A\to bB \mid aa \mid \varepsilon$
$B\to bABBa \mid BD$
$C\to AS \mid c$
$D\to bA \mid \varepsilon$ }

1. H halmaz meghatározása:
   $H_1=\{A,D\}$
   $H_2=\{A,D,S\}$
   $H_3=\{A,D,S,C\}$
   $H_4= H_3$

2. Az epszilon-mentesített nyelvtan:

S szerepel az H halmazban, tehát az üres szó eleme a generált nyelvnek. $S\to\varepsilon$ viszont nem vehető fel új szabályként, mert S szerepel a szabályok jobb oldalán is. Így egy új nyelvtani jelet választunk, legyen ez az S', és ez lesz a nyelvtan új kezdőjele. $S'\to\varepsilon$ szabállyal levezethető az üres szó (S' garantáltan nem fordul elő a szabályok jobb oldalán, így a szabály KES szabály), $S'\to S$ pedig lehetővé teszi az összes egyéb szó levezetését:

$G=<\{a,b,c\}, \{S',S,A,B,C,D\}, \mathbf{P}, S'>$
P={ $S'\to\varepsilon \mid S$
$S\to aAS \mid aS \mid aA \mid a \mid AD \mid A \mid D \mid CSC \mid CS \mid SC \mid CC \mid C$
$A\to bB \mid aa$
$B\to bABBa \mid \mid bBBa \mid BD$
$C\to AS \mid A \mid S \mid c$
$D\to bA \mid b$ }

<u>3. feladat</u> $L = (\varepsilon \mid c \mid (cab)^{+}c) ab$

Kezdjük el hossz szerint felsorolni a nyelv szavait, az első 5 szó:
L={ab, cab, cabcab, $(cab)^3$, $(cab)^4$…}

Ha megfigyeljük a szavak képzésének módját a következő nagyon egyszerű 3-as típusú grammatikát írhatjuk fel a nyelvhez:
G=<{a,b,c}, {S,V}, P, S>
P={ S→ab | cab | cabV
V→cabV | ε}

Lássuk be, hogy L(G)=L :

$L \subseteq L(G)$
Triviális: ’ab’ és ’cab’ szavak egy lépésben megkaphatók S-ből.
$(cab)^n$ szóhoz pedig 1-szer alkalmazzuk az S→cabV, majd (n-1)-szer: V→cabV szabályt, ekkor szavunk $(cab)^nV$ alakú, majd 1-szer befejezésként V→ε szabályt, és ezzel megkaptuk a $(cab)^n$ szót.

$L(G) \subseteq L$
Ez is könnyen látszik P-ből. S-ből vagy egy lépésben levezetjük az ’ab’ és ’cab’ szavakat, ezek elemei L-nek, vagy az S→cabV majd V→cabV szabályokkal hosszabbítjuk a levezetett szót, és befejezzük a levezetést a V→ε szabállyal, de ekkor jól láthatóan $(cab)^k$ alakú szavakat kapunk, amik szintén elemei a nyelvnek. Más alakú szó nem vezethető le a nyelvtannal.

Megjegyzések:

- észrevehető, hogy a reguláris kifejezés egyszerűbben is írható:
  L= ab | $(cab)^+$ amiből szintén hasonló grammatikához jutunk.
- S→cab szabály elhagyható.

<u>4. feladat</u> Vezessen le néhány helyes szót az alábbi nyelvtannal! Milyen nyelvet generál az alábbi grammatika! Állítását indokolja!

S→aSBC | aBC	(1) és (2)
CB→BC	(3)
aB→ab	(4)
bB→bb	(5)
bC→bc	(6)
cC→cc	(7)

L(G)=?

Legrövidebb szó:
S→aBC→abC→abc

Következő:
S→aSBC→aaBCBC→aabCBC→ha most bC→bc szabállyal folytatjuk, elakad a levezetés, nem juthatunk terminálisokhoz az "aabcBC" mondatformából, így a CB→BC csere szabállyal kell folytatni→aabBCC→aabbCC→aabbcC→aabbcc

Ebből sejthető, hogy a keresett nyelv $L=\{u \mid u=a^nb^nc^n\ n>0\}$

Indoklás:

$L \subseteq L(G)$
Receptet adunk, hogyan tudjuk előállítani a nyelv tetszőleges szavát:
Alkalmazzuk az első szabályt (n-1)-szer és a másodikat 1-szer, ekkor $S \Rightarrow a^n(BC)^n$ alakú a mondatformánk. A (3) szabállyal rendezzük $\Rightarrow a^nB^nC^n$ alakba. Most (4) majd (5) szabályokkal cseréljük le a B nyelvtani jeleket $\Rightarrow a^nb^nC^n$, végül (6) és (7) szabályokkal cseréljük le a C nyelvtani jeleket $\Rightarrow a^nb^nc^n$ szót előállítottuk.

$L(G) \subseteq L$
Azt kell belátnunk, hogy csak ilyen alakú szavak vezethetők le a grammatikával:
világos, hogy csak (1) és (2) szabályok növelik a szó hosszát, és csak (1) ad lehetőséget újabb növelésre, tehát indulásként csak egy $a^n(BC)^n$ mondatformához juthatunk. B csak 'a' vagy 'b' terminális jobb oldalán cserélhető le, így valamennyi B-nek el kell jutni ebbe a környezetbe. Ez csak úgy lehetséges, hogy közben a C-k átrendeződnek a szó jobb oldalára. C nyelvtani jelektől pedig csak úgy lehet megszabadulni, ha először a 'b' jobb oldalán lévő C-t, majd a többit lecseréljük, (6) vagy (7) szabály korábbi alkalmazása esetén a levezetés elakad, mert B nem tud eljutni a megfelelő környezetbe, így más alakú szavak nem generálhatók a grammatika szabályaival.

<u>5. feladat</u> 2-es típusú nyelvtan készítése az alábbi nyelvhez:

$T=\{a,b,c,d\}$. $L=\{(ba)^n u(ab)^n \mid n \geq 0 \text{ és } l_d(u) \leq 2 \text{ és } u \in \{b,c,d\}^*\}$

Vizsgáljunk meg a nyelv néhány lehetséges szavát:

- Látható, hogy $\varepsilon \in L$.
- A szó közepének néhány lehetséges alakja: ε, b, c, d, bb, bc, bd, cb, cc, cd, db, dc, dd, stb. (Arra ügyeljünk, hogy három 'd' betű már nem lehet a szó közepében!)
- Ezt veszi „körbe" „ba" és „ab" szótagból tetszőleges, de ugyanannyi darab, lehet 0 darab is!

Tehát a grammatika a következő lehet:

G=<{a,b,c,d},{S, K, H, J},P,S>
P={ $S \to baSab \mid \varepsilon \mid K$
$K \to bK \mid cK \mid dH \mid \varepsilon$ *(ha jött egy d, akkor már csak legfeljebb egy jöhet!)*
$H \to bH \mid cH \mid dJ \mid \varepsilon$ *(ha a második d is bekerült, több d már nem jöhet!)*
$J \to bJ \mid cJ \mid \varepsilon$ }

L(G)=L indoklása:
Számozzuk meg a nyelvtan szabályait:
P={ $S \to baSab \mid \varepsilon \mid K$ (1) (2) (3)
$K \to bK \mid cK \mid dH \mid \varepsilon$ (4) (5) (6) (7)
$H \to bH \mid cH \mid dJ \mid \varepsilon$ (8) (9) (10) (11)
$J \to bJ \mid cJ \mid \varepsilon$ } (12) (13) (14)

$L \subseteq L(G)$
Világos, hogy az ε szó levezethető a (2)-es szabály miatt.
Vegyünk L-ből egy tetszőleges nem ε szót: $v=(ba)^k u(ab)^k$

ennek előállítása: k-szor alkalmazzuk az (1)-es szabályt. Ha u=ε, akkor (2)-vel fejezzük be a levezetést. Ha u≠ε, a (3)-as szabállyal S helyére írjunk K-t. Majd a középső (u-val jelölt) rész betűit vegyük sorra, ha u első betűje ’b’, akkor (4)-es, ha ’c’, akkor az (5)-ös stb. szabályt használjuk, majd vegyük a második betűt, és hasonlóan folytassuk a levezetést. Ha elfogytak u betűi, (7), (11) vagy (14) szabállyal fejezhetjük be a levezetést.

$L(G) \subseteq L$
(1) szabály miatt a szavak elejére és végére azonos számú ’ba’ és ’ab’ szótag kerül, tehát $(ba)^n \ldots (ab)^n$ teljesül a generált nyelv szavaira. A szó közepe a K nyelvtani jelből állítható elő: (4) (5) (8) (9) (12) és (13) szabályok tetszőleges sok és tetszőleges sorrendű ’b’ és ’c’ betűt állítanak elő. A kikötés, hogy a szó közepén legfeljebb 2 db ’d’ betű lehet pedig azért teljesül, mert amikor behozzuk az első ’d’ betűt, K nyelvtani jel helyére H-t írunk, H helyére újra írhatunk ’d’-t, de akkor J-ből folytatódik a levezetés, J-ből viszont nem kerülhet több ’d’ betű a szóba, tehát legfeljebb két ’d’ fog a szó középső részében előfordulni.

<u>6. feladat</u> A feladat reguláris kifejezés keresése az L = { u | u ∈ {a,b}* és „bb” nem része u-nak} nyelv szavaihoz, illetve 3-as típusú nyelvtan készítése a szavak generálásához.

A reguláris kifejezés: (a | ba)*(b | ε)
Helyes megoldások a nyelvtanra például:
G_1=<{a,b},{S},P,S>
P={ S→aS | baS | ε | b}

G_2=<{a,b},{S,A},P,S>
P={ S→aS | bA | b | ε
A→aS }
Persze itt sem maradhat el az L(G)=L indoklása, ami hasonló lehet, mint azt az előbbi feladatoknál láttuk.

<u>7. feladat</u> Készítsen a tanult módszerrel 3-as típusú grammatikát az alábbi reguláris kifejezés által leírt nyelv szavainak generálásához. (*Alkalmazza* a tanult módon az S → t (S∈ N, t∈ T∪{ε}) elemi nyelvtanokra az *unió, konkatenáció, lezárás* műveleteit, ezek segítségével lépésről lépésre építse fel a nyelvtant.)

$L = b^*(ca \mid ac)^*(c \mid \varepsilon)$

A megoldás előállításának lépései. A nyelvtan kezdőjelét megvastagítva jelezzük, az új, vagy a régiből kapott, módosított szabályt pirossal kiemeltük:

b*	**S_1**→b Iterált: **S_2**→ε | S_1 S_1→bS_1| b
ca	**S_3**→c **S_4**→a Konkatenált: **S_3**→cS_4 S_4→a

<table>
<tr><td>ac</td><td colspan="2">$\mathbf{S_5} \to a$ $\mathbf{S_6} \to c$
Konkatenált:
$\mathbf{S_5} \to aS_6$
$S_6 \to c$</td></tr>
<tr><td>ca | ac</td><td colspan="2">ca és ac nyelvtanok uniója:
$\mathbf{S_7} \to S_3 \mid S_5$
$S_3 \to cS_4$
$S_4 \to a$
$S_5 \to aS_6$
$S_6 \to c$</td></tr>
<tr><td>(ca | ac)*</td><td colspan="2">ca | ac nyelvtan iteráltja:
$S_8 \to \varepsilon \mid S_7$
$S_7 \to S_3 \mid S_5$
$S_3 \to cS_4$
$S_4 \to aS_7 \mid a$
$S_5 \to aS_6$
$S_6 \to cS_7 \mid c$</td></tr>
<tr><td>c | ε</td><td colspan="2">$S_9 \to c$ és $S_{10} \to \varepsilon$ nyelvtanok uniója:
$\mathbf{S_{11}} \to S_9 \mid S_{10}$
$S_9 \to c$
$S_{10} \to \varepsilon$</td></tr>
<tr><td>b*(ca | ac)*(c | ε)</td><td colspan="2">Végül a három nyelvtan konkatenáltja adja a kész megoldást:
$\mathbf{S_2} \to S_8 \mid S_1$
$S_1 \to bS_1 \mid bS_8$
$S_8 \to S_{11} \mid S_7$
$S_7 \to S_3 \mid S_5$
$S_3 \to cS_4$
$S_4 \to aS_7 \mid aS_{11}$
$S_5 \to aS_6$
$S_6 \to cS_7 \mid cS_{11}$
$S_{11} \to S_9 \mid S_{10}$
$S_9 \to c$
$S_{10} \to \varepsilon$</td></tr>
<tr><td rowspan="2"></td><td colspan="2">Néhány láncszabályt a láncmentesítés algoritmusa szerint érdemes eltávolítani (pirossal jelölve), jobban áttekinthetővé válik a nyelvtan:</td></tr>
<tr><td>$\mathbf{S_2} \to S_8 \mid S_1$
$S_1 \to bS_1 \mid bS_8$
$S_8 \to S_{11} \mid S_7$
$S_7 \to S_3 \mid S_5$
$S_3 \to cS_4$
$S_4 \to aS_7 \mid aS_{11}$
$S_5 \to aS_6$
$S_6 \to cS_7 \mid cS_{11}$
$S_{11} \to S_9 \mid S_{10}$
$S_9 \to c$
$S_{10} \to \varepsilon$</td><td>$\mathbf{S_2} \to S_8 \mid S_1$
$S_1 \to bS_1 \mid bS_8$
$S_8 \to S_{11} \mid S_7$
$S_7 \to cS_4 \mid aS_6$
$S_4 \to aS_7 \mid aS_{11}$
$S_6 \to cS_7 \mid cS_{11}$
$S_{11} \to c \mid \varepsilon$

S_3, S_5, S_9, S_{10} kiesnek, nem érhetők el S_2-ből.</td></tr>
</table>

Szigorúan követve a tanult konstrukciókat egy helyes, de „terjengős”, sok nyelvtani jelet, sok láncszabályt tartalmazó nyelvtant kaptunk. Ha „kézzel” készítjük a reguláris nyelvtant, csak az elveket követve, a lánc mentesítéseket menet közben elvégezve, akkor egy „emészthetőbb”, jobban átlátható nyelvtant állíthatunk elő. Ezt mutatja be a következő megoldás. Ebben az iterált képzés egy másik lehetséges módszerét használjuk:

Legyen G=(N,T,P,S) egy hármas típusú nyelvtan, ennek iteráltját a következő módszerrel is képezhetjük:

Veszünk egy új kezdőjelet: $S' \notin N$, ebből az $\{ S' \rightarrow \varepsilon \mid S \}$ szabályokat vesszük fel, valamint az eredeti szabályhalmaz $A \rightarrow u$ ($u \in T^*$) alakú szabályait $A \rightarrow uS'$ szabályokra cseréljük. Könnyen megmutatható, hogy ez a konstrukció is pontosan az iterált nyelv szavait állítja elő.

Például: eredeti nyelvtan: $\{ S \rightarrow bbS \mid a \}$
Iterált nyelvtan: $\{S' \rightarrow \varepsilon \mid S,\ S \rightarrow bbS \mid aS' \}$

Az így konstruált nyelvtan:

(1) b*	Elemi nyelvtan: $\{ \mathbf{S_1} \rightarrow b \}$ kezdőjel: S_1 Ebből iterált: $\{ \mathbf{S_2} \rightarrow \varepsilon \mid S_1$ $S_1 \rightarrow bS_2 \}$ kezdőjel: S_2 Láncmentesítve: $\{ \mathbf{S_2} \rightarrow \varepsilon \mid bS_2 \}$ kezdőjel: S_2
(2) ca + ac	Elemi nyelvtanok: $\{ \mathbf{S_3} \rightarrow ca \}$ kezdőjel: S_3 $\{ \mathbf{S_4} \rightarrow ac \}$ kezdőjel: S_4 Unió: $\{ \mathbf{S_5} \rightarrow S_3 \mid S_4$ $S_3 \rightarrow ca$ $S_4 \rightarrow ac \}$ kezdőjel: S_5 Láncmentesítve: $\{ \mathbf{S_5} \rightarrow ca \mid ac \}$ kezdőjel: S_5
(3) $(ca + ac)^*$	Iterált a fenti nyelvtanból: $\{ \mathbf{S_6} \rightarrow \varepsilon \mid S_5$ $S_5 \rightarrow caS_6 \mid acS_6 \}$ kezdőjel: S_6 Láncmentesítve: $\{ \mathbf{S_6} \rightarrow \varepsilon \mid caS_6 \mid acS_6 \}$ kezdőjel: S_6
(4) $(c + \varepsilon)$	Elemi nyelvtanok: $\{ \mathbf{S_7} \rightarrow c \}$ kezdőjel: S_7 $\{ \mathbf{S_8} \rightarrow \varepsilon \}$ kezdőjel: S_8 Unió: $\{ \mathbf{S_9} \rightarrow S_7 \mid S_8$ $S_8 \rightarrow c$ $S_9 \rightarrow \varepsilon \}$ kezdőjel: S_9 Láncmentesítve: $\{ \mathbf{S_9} \rightarrow c \mid \varepsilon \}$ kezdőjel: S_9
(5) b*(ca+ac)*	Konkatenált: (1) után (3) nyelvtan konkatenálása: $\{ \mathbf{S_2} \rightarrow S_6 \mid bS_2$ $S_6 \rightarrow \varepsilon \mid caS_6 \mid acS_6 \}$ kezdőjel: S_2
$b^*(ca+ac)^*(c+\varepsilon)$	Konkatenált: (5) után (4) nyelvtan konkatenálása: $\{ \mathbf{S_2} \rightarrow S_6 \mid bS_2$ $S_6 \rightarrow S_9 \mid caS_6 \mid acS_6$ $S_9 \rightarrow c \mid \varepsilon \}$ kezdőjel: S_2 Láncmentesítve (csak $S_6 \rightarrow S_9$ lánc mentén): $\{ \mathbf{S_2} \rightarrow S_6 \mid bS_2$ $S_6 \rightarrow c \mid \varepsilon \mid caS_6 \mid acS_6\}$ kezdőjel: S_2

<u>8. feladat</u> Készítsen *véges determinisztikus automatát a tanult algoritmussal*, mely az alábbi 3-as típusú grammatika által generált szavakat fogadja el:

G=<{a,b,c},{S,A,B,C},**P**,S>
P={
$S \rightarrow aA \mid acS$
$A \rightarrow aS \mid bC \mid B$
$B \rightarrow aS \mid c$
$C \rightarrow bA \mid S \mid \varepsilon$
}

A 3-as normál forma kialakításának lépései:

Láncmentesítés
H(S)={S} H(A)={A,B} H(B)={B} H(C)={C,S}
$S \rightarrow aA \mid acS$ $A \rightarrow aS \mid bC \mid c$ $B \rightarrow aS \mid c$ $C \rightarrow bA \mid aA \mid acS \mid \varepsilon$ Vegyük észre, hogy B kiesik, ha meghatározzuk az elérhető nyelvtani jeleket: H={S,A,C}
Univerzális epszilon szabály felvétele
$S \rightarrow aA \mid acS$ $A \rightarrow aS \mid bC \mid cF$ $C \rightarrow bA \mid aA \mid acS \mid \varepsilon$ $F \rightarrow \varepsilon$
Hosszúság redukció
$S \rightarrow aA \mid aD$ $A \rightarrow aS \mid bC \mid cF$ $C \rightarrow bA \mid aA \mid aD \mid \varepsilon$ $D \rightarrow cS$ $F \rightarrow \varepsilon$

A 3 NF-ből kapott automata (NDA):

		a	b	c
→	S	A,D		
	A	S	C	F
←	C	A,D	A	
	D			S
←	F			

Az NDA-ból kapott VDA:

		a	b	c
→	{S}	{A,D}	{}	{}
	{A,D}	{S}	{C}	{S,F}
←	{C}	{A,D}	{A}	{}
←	{S,F}	{A,D}	{}	{}
	{A}	{S}	{C}	{F}
←	{F}	{}	{}	{}
	{}	{}	{}	{}

<u>9. feladat</u> A tanult módszerrel alakítsa át az alábbi ballineáris grammatikát jobblineárisra (S a kezdőszimbólum)!

$S \rightarrow Cb \mid Aab \mid \varepsilon \mid bc$
$A \rightarrow Bba \mid Ac \mid cc$
$B \rightarrow Abc \mid Ca$
$C \rightarrow Ccc \mid Bb \mid aa$

Bal lineáris	Jobb lineáris
$S \rightarrow Cb \mid Aab \mid \varepsilon \mid bc$	$S \rightarrow \varepsilon \mid bc$ $A \rightarrow ab$ $C \rightarrow b$
$A \rightarrow Bba \mid Ac \mid cc$	$S \rightarrow ccA$ $A \rightarrow cA$ $B \rightarrow baA$
$B \rightarrow Abc \mid Ca$	$A \rightarrow bcB$ $C \rightarrow aB$
$C \rightarrow Ccc \mid Bb \mid aa$	$S \rightarrow aaC$ $B \rightarrow bC$ $C \rightarrow ccC$

Összefoglalva a kapott szabályok:

$S \rightarrow ccA \mid aaC \mid bc \mid \varepsilon$
$A \rightarrow cA \mid bcB \mid ab$
$B \rightarrow baA \mid bC$
$C \rightarrow aB \mid ccC \mid b$