1. Legyen $\Sigma = \{\mathtt{a}, \mathtt{b}\}$ és álljon L azokból a szavakból, melyekben az $\mathtt{a}$ és a $\mathtt{b}$ betűk száma megegyezik. Reguláris-e ez az L nyelv?

   Megoldás: Nem. Indirekt bizonyítunk, az előadáson az $\{a^n b^n\}$ nyelvre elmondott bizonyítás itt is bizonyít.

   Indirekt bizonyítunk. Tegyük fel, hogy az L nyelv reguláris, azaz van olyan M determinisztikus véges automata, ami L-et fogadja el. Jelölje t az M állapotainak számát. Tekintsük a $0, 1, 2, \ldots, t$ hosszú, csupa $\mathtt{a}$ betűből álló szavakat. Mivel csak t állapot van, ez a $t+1$ szó között van legalább kettő, amelyek ugyanabban az állapotban érnek véget. Ha a k és az ℓ hosszú $\mathtt{a}$-sorozat is a q állapotban ér véget ($k \neq \ell$), akkor innen folytatva k darab és ℓ darab $\mathtt{b}$ betűvel is elfogadó állapotba kell érjünk. Viszont ekkor az $\mathtt{a}^k\mathtt{b}^\ell$ szóval is elfogadó állapotba jutunk, pedig ez nincs is a megadott nyelvben.

2. Álljon az ábécé a nyitó és a csukó zárójelből. Igazolja, hogy a helyes zárójelsorozatokból álló nyelv nem reguláris!

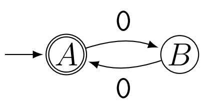
   Megoldás: Az 1. feladat bizonyítása itt is érvényes, csak az $\mathtt{a}$ helyett nyitó, a $\mathtt{b}$ helyett csukó zárójellel.

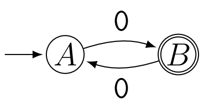
   Fontos: *Nem jó bizonyítás: Ez (az eltérő abc-től eltekintve) az 1. feladat nyelvének részhalmaza. Mivel az nem reguláris, ez sem az.* **Általában nem igaz, hogy egy nem reguláris nyelvet tartalmazó nyelv sem reguláris!** Miért?

3. Reguláris-e az a nyelv, ami az olyan, csupa $\mathtt{0}$ sorozatból áll, amelyeknek a hossza

   (a) páros szám? (b) páratlan szám?
   (c) négyzetszám? (d) kettő hatvány?

   Megoldás:

   (a) Igen: 

   (b) Igen: 

   A (c) és (d) esetén a válasz nem: gondoljuk meg, hogyan néz ki egy DVA ha csak egy elemű az ábécé! Minden állapotból egyetlen nyíl (átmenet) indul ki. A kezdőállapotból a gráfban van egy mondjuk t hosszú út, aminek utolsó csúcsán vagy egy hurok van vagy innen az él visszamutat egy korábbi állapotba. Tehát a gráf egy kezdeti útból és egy az út végén levő körből áll (az út lehet 0 hosszú, a kör meg 1 hosszú, utóbbi ha hurok van). Ha nincs elfogadó állapot a körön, akkor csak véges sok különböző szót tud elfogadni. Ha van (lehet akár több is), akkor végtelen sok szót. Méghozzá, ha c jelöli a kör hosszát, és $\mathtt{0}^k \in L$ egy a körön levő elfogadó állapotban ér véget, akkor körbe érve $\mathtt{0}^{k+c} \in L$ is teljesül. Ezért az nem fordulhat elő, hogy egy adott nyelvnél az elfogadott szavak hosszai között tetszőlegesen nagy ugrás előforduljon.

   A korábbi bizonyítási technika is működik, például így ha t állapota van a DVA-nak, akkor vegyük a nyelv $t+1$ darab legrövidebb szavát. Biztos van közöttük kettő, ami ugyanabban az (elfogadó) állapotban ér véget. Tegyük fel, hogy $|w_1| = k^2$ és $|w_2| = \ell^2$ ugyanabban a q állapotban végződik és $k < \ell$. Ha q-ból még $2k+1$ lépést teszünk, akkor elfogadó állapothoz kell jussunk, hiszen ha a w_1-et folytatjuk, akkor egy $k^2 + 2k + 1 = (k+1)^2$ hosszú szót kapunk. Viszont ha a w_2-t folytatjuk, akkor is elfogad az automata, pedig a szó hossza $\ell^2 + 2k + 1 < \ell^2 + 2\ell + 1 = (\ell+1)^2$, és mivel ugyanakkor nagyobb mint ℓ^2, ezért nem négyzetszám.

   A (d)-re ugyanez az ötlet (de más számolás) működik.

4. Legyen az ábécé $\Sigma = \{\mathtt{0}, \mathtt{1}\}$. Határozza meg az alábbi reguláris kifejezésekhez tartozó nyelveket!

   (a) $(\mathtt{0}+\mathtt{1})^*\mathtt{011}(\mathtt{0}+\mathtt{1})^*$ (b) $\mathtt{1}(\mathtt{0}+\mathtt{1})^*\mathtt{0}$ (c) $((\mathtt{0}+\mathtt{1})(\mathtt{0}+\mathtt{1}))^*$

   Megoldás: (a) A $\mathtt{011}$-et tartalmazó szavak. (b) Az $\mathtt{1}$-gyel kezdődő és $\mathtt{0}$-ra végződő szavak. (c) A páros hosszú szavak (tetszőleges kettő hosszú szavakból rakunk egymás után valahány, akár nulla darabot).

5. Adjon reguláris kifejezést azokra a nyelvekre, amelyek a $\{0, 1\}$ ábécé felett a következő szavakból állnak!

   (a) páratlan hosszú szavak;

   (b) páros hosszú nem üres szavak, melyeknek első és utolsó karaktere is 1;

   (c) legalább 3 db 0-t tartalmazó szavak;

   (d) páros sok 0-t tartalmazó szavak;

   (e) a 0-val kezdődő és páratlan hosszú, valamint az 1-gyel kezdődő és páros hosszú szavak;

   (f) a 00 részszót tartalmazó páratlan hosszú szavak.

   Megoldás: (a) Az üres szó nem jó. Tetszőleges első karakter után egy tetszőleges páros hosszú szó következik: $(0+1)((0+1)(0+1))^*$ (vagy persze az utolsó karaktert is levághatjuk az első helyett).
   (b) Az első és utolsó karakter között tetszőleges páros hosszú szó állhat: $1((0+1)(0+1))^*1$
   (c) A három kiválasztott 0 karakter előtt, után és között is bármi állhat: $(0+1)^*0(0+1)^*0(0+1)^*0(0+1)^*$
   (d) A 0-kat párosával tesszük le, két szomszédos között, előttük, utánuk tetszőleges számú 1 állhat: $(1^*01^*01^*)^*1^*$ (a legvégén lévő 1^*-ra azért van szükség, hogy nulla darab 0-ra is működjön), vagy pl. a 0-k után álló 1-eket elhagyhatjuk, a következő pár elején úgyis van 1^*, ilyenkor a legvégén lévő 1^* gondoskodni arról, hogy végződhessen valahány egyesre is: $(1^*01^*0)^*1^*$.
   (e) Azt, hogy a két lehetőség legalább egyike teljesül a két reguláris kifejezés összege fejezi ki: $0((0+1)(0+1))^* + 1(0+1)((0+1)(0+1))^*$.
   (f) Vagy páros sok karakter van a 00 előtt és akkor utána páratlan, vagy előtte van páratlan és utána páros, azaz $((0+1)(0+1))^*00(0+1)((0+1)(0+1))^* + (0+1)((0+1)(0+1))^*00((0+1)(0+1))^*$, vagy kicsit másként csoportosítva, valamivel rövidebben: $((0+1)(0+1))^*(00(0+1)+(0+1)00)((0+1)(0+1))^*$.

6. Adjon olyan reguláris kifejezéseket, amelyek rövidebbek az itt szereplőknél, de ugyanazt a nyelvet írják le!

   (a) $(0+\varepsilon)^*$ (b) $((0+\varepsilon)(0+\varepsilon))^*$ (c) $(0+1)^*01(0+1)^* + 1^*0^*$

   Megoldás: (a) A kifejezés tetszőleges számú 0-t generál, a 0^* is pont ezt csinálja.
   (b) A két zárójel együtt nulla, egy vagy kettő 0 karaktert ad, ezt lehet tetszőleges számszor ismételni, azaz minden, 0-kból álló szót generál, ami leírható a 0^* kifejezéssel is.
   (c) A leírt szavak: van benne 01 vagy előbb 1-ek és utána 0-k állnak (vagyis nincs benne 01), azaz bármilyen szó lehet, tehát a $(0+1)^*$ jó.

7. Adjon reguláris kifejezést arra a nyelvre, ami az összes, az 110 részszót nem tartalmazó $\{0, 1\}$ feletti szóból áll!

   Megoldás: A reguláris kifejezésnél nincs művelet a kivonásra, azt kell kitalálnunk, hogyan néznek ki a megengedett szavak. Egy ilyen szó állhat csupa 0-ból, vagy kezdődhet tetszőleges számú 0 karakterrel (akár nulla darabbal is). Ha van benne 1, akkor két 1 között kell legyen 0, kivéve, ha a szó végén vagyunk, ott akárhány 1 lehet egymás után. Ha nincs két 1 egymás után, akkor a végén még lehetnek 0-k. Ezek alapján egy lehetséges megoldás: $0^*(\varepsilon + 1(0^*01)^*(0^* + 1^*))$, vagy egy elegánsabb: $(0+10)^*1^*$.

8. Mit generál az $X \to cXd \mid Y; \quad Y \to cY \mid c$ nyelvtan?

   Megoldás: Sejtés: ez a nyelvtan az olyan c^id^j alakú szavakat generálja, amikben $j \geq 0$ és $i > j$.

   Ennek belátására két dolgot kell megmutatni:

   Csak ilyen szavakat tud generálni:
   A levezetés elején egy darabig az X baloldalú első szabályt használjuk j-szer ($j \geq 0$), eközben c^jXd^j alakú szavak állnak elő. Ezután, ha terminális karakterekből álló szót akarunk, akkor használnunk kell az $X \to Y$ szabályt, így kapunk egy c^jYd^j alakú szót. Ezután y-ból tetszőleges számú, de

legalább egy darab c betűt generálunk, így végül egy $c^j c^t d^j$ alakú szót kapunk, ahol $j \geq 0$ és $t \geq 1$, azaz pont egy olyan szó, amit sejtettünk.

Minden ilyen szót generál:
Ha kapunk egy olyan $c^i d^j$ alakú szót, amiben $j \geq 0$ és $i > j$, akkor az előző részben vázolt levezetéssel ezt elő is tudjuk állítani.

9. Határozza meg az $S \to A \mid B$; $A \to \mathtt{0}A\mathtt{1} \mid \mathtt{01}$; $B \to \mathtt{1}B\mathtt{0} \mid \mathtt{10}$ nyelvtan által generált nyelvet!

   Megoldás: a nyelv az A-ból, illetve a B-ből generálható nyelvek uniója, és ezért
   $L = \{\mathtt{0}^n\mathtt{1}^n : n \geq 1\} \cup \{\mathtt{1}^n\mathtt{0}^n : n \geq 1\}$.

10. Adjon környezetfüggetlen nyelvtant az 4. feladatban szereplő nyelvekre!

    Megoldás: Sok helyes nyelvtan van, mutatunk egyet-egyet, ami a reguláris kifejezés szerkezetét tükrözi.
    (a) $S \to A\mathtt{011}A$, $A \to \mathtt{0}A \mid \mathtt{1}A \mid \varepsilon$ (A-ból a $(\mathtt{0}+\mathtt{1})^*$ generálható, előadáson is volt)
    (b) $S \to \mathtt{1}A\mathtt{0}$, $A \to \mathtt{0}A \mid \mathtt{1}A \mid \varepsilon$
    (c) $S \to \mathtt{00}S \mid \mathtt{01}S \mid \mathtt{10}S \mid \mathtt{11}S \mid \varepsilon$, de például $S \to \mathtt{0}S\mathtt{0} \mid \mathtt{0}S\mathtt{1} \mid \mathtt{1}S\mathtt{0} \mid \mathtt{1}S\mathtt{1} \mid \varepsilon$ is jó.

11. Adjon környezetfüggetlen nyelvtant az alábbi nyelvekre!
    (a) $L = \{\mathtt{a}^n\mathtt{b}^{n+1} : n \geq 0\}$ (b) $L = \{\mathtt{a}^n\mathtt{b}^{2n} : n \geq 0\}$

    Megoldás: Mindegyiknél több jó megoldás is van, pl.:
    (a) $S \to T\mathtt{b}$, $T \to \mathtt{a}T\mathtt{b} \mid \varepsilon$
    Egy másik lehetőség például, ha lényegében az $\mathtt{a}^n\mathtt{b}^n$ nyelvet generáljuk, és a végén rakjuk a szó „közepére” az extra $\mathtt{b}$ betűt, azaz $S \to \mathtt{a}S\mathtt{b} \mid \mathtt{b}$.
    (b) Az ötlet, hogy most egy lépésben egy $\mathtt{a}$ betűt teszünk előre és két $\mathtt{b}$-t hátra, $S \to \mathtt{a}S\mathtt{bb} \mid \varepsilon$. Világos, hogy az így kapott nyelvtan jó.

12. Adjon környezetfüggetlen nyelvtant a jó zárójelezések nyelvéhez!

    Megoldás: A zárójelsorozatot elképzelhetjük úgy, hogy vannak a külső szintű zárójelek (ezekből akárhány), és minden ilyen külső szintű zárójelpáron belül is egy jó sorozatnak kell lenni. Ebből a következő nyelvtant kaphatjuk:
    $Z \to ZZ \mid (Z) \mid \varepsilon$
    Az első szabállyal legyárthatjuk a tetszőleges számú külső szintű zárójel helyét, a második szabály kirakja a zárójelpárokat, és lehetőséget ad, hogy a belsejükben folytassuk az eljárást.

    Kezdhetjük az első külső zárójelpárral is, aminek a belsejében, és utána is jó sorozatnak kell lenni:
    $Z \to (Z)Z \mid \varepsilon$

    Vagy kezdhetjük egy tetszőleges külső zárójelpárral. Ekkor előtte, benne és utána is helyes sorozat kell álljon:
    $Z \to Z(Z)Z \mid \varepsilon$

    (Ráadás: a fentiek közül melyik nyelvtan egyértelmű és melyik nem? (A kérdés megválaszolásához a 4. heti előadáson elhangzó anyag ismerete szükséges.))

13. Határozza meg az alábbi környezetfüggetlen nyelvtanok által generált nyelveket!
    (a) $T \to TT \mid \mathtt{a}T\mathtt{b} \mid \mathtt{b}T\mathtt{a} \mid \mathtt{a} \mid \varepsilon$
    (b) $R \to T\mathtt{a}T$; $T \to TT \mid \mathtt{a}T\mathtt{b} \mid \mathtt{b}T\mathtt{a} \mid \mathtt{a} \mid \varepsilon$

    Megoldás: (a) Az világos, hogy a keletkezett szóban, ha nem az üres szó, akkor legalább annyi $\mathtt{a}$ betű lesz, mint $\mathtt{b}$. (Ez utóbbi tulajdonság valójában az üres szóra is teljesül.) Megmutatjuk, hogy minden ilyen szó levezethető, ezt a hossz szerinti teljes indukcióval csináljuk. Tekintsünk egy ilyen w szót. Ha a hossza $|w| \leq 1$, akkor vagy $w = \varepsilon$ vagy $w = \mathtt{a}$, és mindkettő valóban levezethető. Tegyük fel, hogy minden n-nél rövidebb, legalább annyi $\mathtt{a}$ betűt, mint $\mathtt{b}$ betűt tartalmazó szó levezethető.

Legyen most $w = x_1 x_2 \cdots x_n$ egy $n \geq 2$ hosszú szó, amiben legalább annyi `a` van, mint `b`.

Legyen i a legkisebb olyan pozitív szám, melyre teljesül, hogy az $x_1 \cdots x_i$ részszóban ugyanannyi `a` van, mint `b`.

Ha nincs ilyen i, akkor minden kezdőszeletben, és az egész szóban is több az `a`, mint a `b`. Tehát biztos, hogy $x_1 = \mathtt{a}$ és ha ezt a betűt levágjuk, akkor is a nyelvben maradunk, ezért az indukciós feltevés miatt a $T \Rightarrow TT \Rightarrow \mathtt{a}T$ kezdés folytatható úgy, hogy a végén jó levezetést kapjunk.

Vegyük észre, hogy ha van jó i, akkor $x_1 \neq x_i$ (mert az i-ediknél lesz pont ugyanannyi `a`, mint `b`). Ezért ha a levezetés első lépése után az első T-ből a 2. vagy 3. szabállyal megkapjuk az x_1 és x_i karaktereket, közéjük a többi (itt is ugyanannyi `a` van, mint `b`) a T-ből generálható. A szó végén is legalább annyi `a` van, mint `b`, ezért ez megkapható az első lépésben kapott második T-ből.

Másik megfontolás („alulról felfelé”): kiindulunk a szóból, és alkalmazzuk a következő átírási szabályokat: tetszőleges `ab` vagy `ba` részszót helyettesítsünk T-vel (a $T \Rightarrow \mathtt{a}T\mathtt{b} \Rightarrow \mathtt{ab}$ vagy a $T \Rightarrow \mathtt{b}T\mathtt{a} \Rightarrow \mathtt{ba}$ lépéssorozatok megfordításai); TT-t helyettesítsünk T-vel; $\mathtt{a}T\mathtt{b}$ és $\mathtt{b}T\mathtt{a}$ szintén helyettesíthető T-vel. Mit kapunk, amikor ezek egyike sem alkalmazható: `b` betű biztos nem marad (mert akkor `a` is kell, hogy maradjon, és lesz, esetleg T-vel elválasztott, `a` és `b` is). Ha már csak `a` és T maradt, helyettesítsük az `a` betűket T-vel, a szomszédos T-ket meg egyetlen T-vel. Így a végén egyetlen T marad csak, és ekkor megkaptunk (visszafele) egy levezetést.

(b) A nyelv azokból a szavakból áll, amelyekben több `a` van, mint `b`.

Vegyük észre, hogy a T-re az előző szabályok maradtak, azaz T-ből azok a szavak vezethetők le, amelyekben legalább annyi `a` van, mint `b`. A kezdő szabállyal még egy `a` betűt hozzáteszünk, tehát biztos, hogy a levezetett szavakban több `a` lesz, mint `b`. Azt kell még megindokolni, hogy minden ilyet megkapunk: Egy ilyen w szót bontsunk úgy fel, hogy $w = w_1 \mathtt{a} w_2$, ahol w_1-ben ugyanannyi `a` van, mint `b`, w_2-ben meg legalább ugyanannyi. Ez a w_1 és w_2 is levezethető T-ből, tehát az egész szó is.

Ilyen felbontás mindig van, mert tekintsük a legrövidebb kezdőszeletet, amiben az `a`-k száma több, mint a `b`-k száma. Ennek az utolsó karaktere `a`, az ez előtti rész legyen w_1, az utána levő pedig w_2. (Lehet, hogy $w_1 = \varepsilon$ vagy $w_2 = \varepsilon$.)

14. Határozza meg a következő nyelvtan által generált nyelvet!

$$\begin{aligned} R &\to XRX \mid S \\ S &\to \mathtt{a}T\mathtt{b} \mid \mathtt{b}T\mathtt{a} \\ T &\to XTX \mid X \mid \varepsilon \\ X &\to \mathtt{a} \mid \mathtt{b} \end{aligned}$$

Megoldás: Ez a nem palindromokból álló nyelv. Egy szó pontosan akkor nem palindrom, ha van olyan i, hogy az elölről és a hátulról számított i-edik karaktere eltérő. (Több ilyen i is lehet, de most válasszunk egyet.) Egy ilyen szó úgy vezethető le a nyelvtanból, ha $(i-1)$-szer alkalmazzuk az 1. szabályt, amivel megcsináljuk a helyet az első és utolsó $i-1$ karakternek, utána a 2. szabály segítségével kapunk egy S-et, amivel a 3., illetve 4. szabály segítségével létrehozzuk az i-edik pozíciókba az eltérő karaktereket, utána az 5-9. szabályokkal tetszőlegesen kitölthetjük a többi helyet.

Azt is könnyű látni, hogy csak ilyen szavak vezethetők le a nyelvtanból, mert X-ből az $(\mathtt{a}+\mathtt{b})$, T-ből az $(\mathtt{a}+\mathtt{b})^*$ reguláris kifejezéssel leírható nyelvek vezethetők le, S-ből az olyanok, amiknek az első és utolsó betűje különbözik, e köré rak R ugyanannyi karaktert előre, mint hátulra.