3. Reguláris kifejezés, pumpálási lemma

1. Álljon az $L \subseteq \{\mathtt{a},\mathtt{b}\}^*$ nyelv azokból a szavakból, melyekben minden a-blokk páratlan hosszú. Adjon L-hez reguláris kifejezést!

   Megoldás: A páratlan hosszú blokkok így írhatók fel: $\mathtt{a(aa)}^*$. Amire vigyázni kell, hogy két ilyen blokk közé kerüljön legalább egy b, mert különben összeolvadnak egy páros blokká. Továbbá, hogy a szó elején és végén állhat b is, de lehet a-blokk is. Tehát ez jó lesz: $\mathtt{b}^*\ (\mathtt{a(aa)}^*\ \mathtt{bb}^*)^*\ (\varepsilon + \mathtt{a(aa)}^*)$.

2. A $(0+1)^*01(0+1)^* + 1^*0^*$ reguláris kifejezés által meghatározott nyelvhez adjon DVA-t!

   Megoldás: Követhetjük a tanult konstrukció lépéseit (a végén a kapott automatát determinizálni is kell!), de mivel a feladat szövege ezt nem kéri, inkább gondolkodjunk...

   A reguláris kifejezés első tagja azokat a szavakat írja le, amiben szerepel a 01 részszó. Ami ebben nem szerepel az vagy üres, vagy csak egyféle karakterből áll, vagy néhány 1-t követ néhány 0. Ezeket együtt leírja az 1^*0^* reguláris kifejezés. Azaz a feladatban megadott reguláris kifejezéshez tartozó nyelv a $\{\mathtt{0},\mathtt{1}\}^*$, amihez az egyetlen állapotú, $\delta(q,\mathtt{0}) = \delta(q,\mathtt{1})$ állapotátmenetű DVA, ahol q a kezdő és egyben az elfogadó állapot is.

3. Legyen $\Sigma = \{\mathtt{0},\mathtt{1}\}$. Az M véges automata állapotai $Q = \{1,2,3\}$, ezekből az 1 a kezdőállapot, $F = \{2,3\}$ az elfogadó állapotok, $\delta(i,\mathtt{0}) = i$, $\delta(1,\mathtt{1}) = 2$, $\delta(2,\mathtt{1}) = 3$, $\delta(3,\mathtt{1}) = 1$.

   (a) Mi az M által elfogadott nyelv?

   (b) Adja meg az $R(i,j,0)$ reguláris kifejezéseket!

   (c) A tanult eljárást használva adjon meg az $L(M)$ nyelvhez egy reguláris kifejezést!

   Megoldás: (a) $L(M) =$ az 1-k száma nem osztható 3-mal.
   (b) $R(i,i,0) = \varepsilon + \mathtt{0}$, $R(i,i+1,0) = \mathtt{1}$, $R(i,i+2,0) = \emptyset$. (Az összeadást mod 3 kell érteni.)
   (c) $R(i,j,1) = R(i,j,0) + R(i,1,0)R(1,1,0)^*R(1,j,0) = R(i,j,0) + R(i,1,0)\mathtt{0}^*R(1,j,0)$, ami egyszerűsítve:

$i \setminus j$	1	2	3
1	$\mathtt{0}^*$	$\mathtt{0}^*\mathtt{1}$	$\emptyset$
2	$\emptyset$	$\varepsilon + \mathtt{0}$	$\mathtt{1}$
3	$\mathtt{10}^*$	$\mathtt{10}^*\mathtt{1}$	$\varepsilon + \mathtt{0}$

   $R(i,j,2) = R(i,j,1) + R(i,2,1)R(2,2,1)^*R(2,j,1) = R(i,j,1) + R(i,2,1)\mathtt{0}^*R(2,j,1)$, ami egyszerűsítések után:

$i \setminus j$	1	2	3
1	$\mathtt{0}^*$	$\mathtt{0}^*\mathtt{10}^*$	$\mathtt{0}^*\mathtt{10}^*\mathtt{1}$
2	$\emptyset$	$\mathtt{0}^*$	$\mathtt{0}^*\mathtt{1}$
3	$\mathtt{10}^*$	$\mathtt{10}^*\mathtt{10}^*$	$\varepsilon + \mathtt{0} + \mathtt{10}^*\mathtt{10}^*\mathtt{1}$

   $R(i,j,3) = R(i,j,2) + R(i,3,2)R(3,3,2)^*R(3,j,2) = R(i,j,2) + R(i,3,2)(\mathtt{0} + \mathtt{10}^*\mathtt{10}^*\mathtt{1})^*R(3,j,2)$, amiből a nyelv $R(1,2,3) + R(1,3,3) = R(1,2,2) + R(1,3,2) + R(1,3,2)(\mathtt{0} + \mathtt{10}^*\mathtt{10}^*\mathtt{1})^*(R(3,2,2) + R(3,3,2)) = \mathtt{0}^*\mathtt{10}^* + \mathtt{0}^*\mathtt{10}^*\mathtt{1}(\mathtt{0} + \mathtt{10}^*\mathtt{10}^*\mathtt{1})^*(\mathtt{10}^*\mathtt{10}^* + \varepsilon)$.

4. Az $L = \{\mathtt{ab}^n : n \geq 1\} \cup \{\mathtt{baba}\}$ nyelvnél a pumpálási lemmát valaki a $z = \mathtt{baba}$ szóra akarja használni.

   (a) A z szó pumpáltjai benne lesznek-e a nyelvben?

   (b) Következik-e ebből hogy L reguláris vagy az, hogy L nem reguláris?

   (c) Reguláris-e az L nyelv?

   Megoldás:

(a) Nem, mivel ezek $\mathtt{b}$-vel kezdődnek és különböznek a nyelv egyetlen ilyen szavától, a $\mathtt{baba}$-tól.

(b) Ebből semmi nem következik, hiszen lehet, hogy a p pumpálási hossza $p > 4$, és akkor erről a szóról nem is szól a pumpálási lemma.

(c) A nyelv reguláris, mivel $L = L_1(L_2 - L_3) \cup L_4$, ahol $L_1 = \{\mathtt{a}\}$ reguláris (pl. mert véges), $L_2 = \mathtt{b}^*$ reguláris, $L_3 = \{\varepsilon\}$ reguláris. $L_4 = \{\mathtt{baba}\}$ reguláris (pl. mert véges), és a reguláris nyelvek zártak az itt használt műveletekre (konkatenálás, különbség, unió), tehát az eredményül kapott L is reguláris.

Megjegyzés: Egy determinisztikus véges automatát sem nehéz mutatni rá.

5. Legyen $L \subseteq \{\mathtt{a}, \mathtt{b}\}^*$ a palindromok nyelve. Valaki ebből a $z = \mathtt{ab}^p\mathtt{a}$ szót választotta pumpálásra, ahol p a pumpálási hossz.

   (a) A z szó pumpáltjai benne lesznek-e a nyelvben?

   (b) Következik-e ebből hogy L reguláris vagy az, hogy L nem reguláris?

   (c) Reguláris-e az L nyelv?

Megoldás:

(a) Legyen a felosztás $z = uvw$. A pumpálási lemma feltétele szerint $|uv| \leq p$.

Ha a pumpált rész $v = \mathtt{ab}^t$, akkor a pumpáltak $(\mathtt{ab}^t)^k\mathtt{b}^{p-t}\mathtt{a}$ alakúak. Ha $k > 1$, akkor a szó első $\mathtt{b}$-blokkjának hossza $t < p$, míg az utolsó $\mathtt{b}$-blokkjának hossza $p \neq t$, tehát ez nem palindrom. A $k = 0$ esetben a pumpálás eredménye $\mathtt{b}^{p-t}\mathtt{a}$, és mivel $p - t \geq 1$ ez sem palindrom. Azaz ebben az esetben mindig ($k \neq 1$ esetén) kikerülünk a nyelvből.

Ha a pumpált szakasz csupa $\mathtt{b}$-ből áll, $v = \mathtt{b}^t$ akkor viszont a szó minden k esetén palindrom marad, ilyenkor tehát a pumpálás során benne marad a nyelvben.

(b) Mivel a szónak volt olyan felosztása, aminél a pumpálás nem vezet ki a nyelvből, ezzel nem jutottunk aellentmondásra. Abból pedig, hogy egy szóból nem jutunk ellentmondásra még semmi sem következik. (Másként kellett volna választani!)

(c) A nyelv nem reguláris. A pmpálási lemma alkalmazásához válasszuk az előző helyett pl. a $z = \mathtt{a}^p\mathtt{ba}^p$ szót. Ennek minden, a pumpálási lemma szerinti $z = uvw$ felosztásában $v = \mathtt{a}^t$, ahol $1 \leq t \leq p$. Ekkor pl. $k = 2$ esetén a szó $uvvw = \mathtt{a}^{p+t}\mathtt{ba}^p$ lesz, ami nem palindrom, hiszen a szóban szereplő egyetlen $\mathtt{b}$ nem középen van.

6. A pumpálási lemma segítségével igazolja, hogy az alábbi nyelvek nem regulárisak!

   (a) $L_a = \{\mathtt{a}^m\mathtt{b}^n : 1 \leq m \leq n \leq 2m\}$

   (b) $L_b = \{\mathtt{a}^i\mathtt{b}^j\mathtt{c}^k : i > j \geq 0,\ \ k \geq 0\}$.

   (c) $L_c = \{\mathtt{0}^{n!} : n \geq 1\}$

   (d) $L_d = \{\mathtt{0}^p : p \text{ prímszám } \}$

   (e) $L_e\{(\mathtt{ab})^n(\mathtt{ba})^n : n \geq 0\}$

Megoldás:

(a) Tegyük fel, hogy a nyelv reguláris. Ekkor igaz rá a pumpálási lemma, legyen $p > 0$ az ebből adódó pumpálási hossz. Válasszuk pl. a $z = \mathtt{a}^p\mathtt{b}^p$ szót. Erre $z \in L_a$ és $|z| = 2p \geq p$ teljesül. Ahhoz, hogy ellentmondásra jussunk, azt kell megmutatni, hogy ennek minden $z = uvw$ felosztásához, ahol $|uv| \leq p$ és $|v| > 0$ van olyan $k \geq 0$ szám, amire $uv^kw \notin L_a$.

Minden ilyen felosztásban $|uv| \leq p$ miatt v csak $\mathtt{a}$ betűkből áll, $v = \mathtt{a}^t$ ahol $0 < t \leq p$. Ezért pl. $k = 2$ esetén $uv^2w = \mathtt{a}^{p+t}\mathtt{b}^p \notin L_a$, hiszen $m = p + t > p$, $n = p$, és így nem teljesül, hogy $m \leq n$.

Tehát erre az L_a-beli szóra nem teljesül a pumpálási lemma, a nyelv nem reguláris.

(b) Tegyük fel, hogy a nyelv reguláris. Ekkor igaz rá a pumpálási lemma, legyen $p > 0$ az ebből adódó pumpálási hossz. Válasszuk pl. a $z = \mathtt{a}^p\mathtt{b}^{p-1}\mathtt{c}^{p-1}$ szót. Erre $z \in L_b$ és $|z| = 3p - 2 \geq p$ teljesül. Azt kell megmutatni, hogy ennek minden $z = uvw$ felosztásához, ahol $|uv| \leq p$ és $|v| > 0$ van olyan $k \geq 0$ szám, amire $uv^kw \notin L_b$.

Minden ilyen felosztásban, $|uv| \leq p$ miatt, v csak $\mathtt{a}$ betűkből áll, $v = \mathtt{a}^t$ ahol $0 < t \leq p$. Legyen most $k = 0$. Ekkor $uv^0w = \mathtt{a}^{p-t}\mathtt{b}^{p-1}\mathtt{c}^{p-1}$ ami nincs L_b-ben, hiszen $p - t \leq p - 1$.

Tehát erre az L_b-beli szóra nem teljesül a pumpálási lemma, a nyelv nem reguláris.

(c) Tegyük fel, hogy a nyelv reguláris. Ekkor igaz rá a pumpálási lemma, legyen $p > 0$ az ebből adódó pumpálási hossz. Ha ez kisebb, mint 2, akkor legyen $p = 2$. (Nagyobb számot mindig választhatunk, mint a tényleges pumpálási hossz. Miért?) Válasszuk pl. a $z = \mathtt{0}^{p!}$ szót. Erre $z \in L_c$ és $|z| = p! \geq p$ teljesül. Azt kell megmutatni, hogy ennek minden $z = uvw$ felosztásához, ahol $|uv| \leq p$ és $|v| > 0$, van olyan $k \geq 0$ szám, amire $uv^k w \notin L_c$.

A v részszó csak $\mathtt{0}^t$ alakú lehet, ahol $0 < t \leq p$. Ekkor $uv^2w = \mathtt{0}^{p!+t}$. Mivel $p \geq 2$, ezért $p! < p! + t \leq p! + p \leq 2p! < (p+1)!$, tehát uv^2w hossza nem faktoriális, $uv^2w \notin L_c$.

Tehát erre az L_c-beli szóra nem teljesül a pumpálási lemma, a nyelv nem reguláris.

(d) Tegyük fel, hogy a nyelv reguláris. Ekkor igaz rá a pumpálási lemma, legyen $p > 0$ az ebből adódó pumpálási hossz. Válasszunk egy $z = \mathtt{0}^r$ szót, ahol $r \geq p$ prímszám. (Ez lehetséges, mivel végtelen sok prímszám van.)

Erre $z \in L_d$ és $|z| = r \geq p$ teljesül. Azt kell megmutatni, hogy ennek minden $z = uvw$ felosztásához, ahol $|uv| \leq p$ és $|v| > 0$ van olyan $k \geq 0$ szám, amire $uv^k w \notin L_d$.

A v részszó csak $\mathtt{0}^t$ alakú lehet, $0 < t \leq p$. Most a $k = 2$ nem biztos, hogy segít, hiszen lehet, hogy $r + t$ prímszám. De ha a $k = r + 1$ esetet nézzük, akkor $uv^k w = \mathtt{0}^{r+(k-1)t}$, aminek hossza $r + (k-1)t = r + rt = r(t+1)$ biztos nem prím. Ezért az $uv^{r+1}w = \mathtt{0}^{r(t+1)} \notin L_d$.

Tehát erre az L_d-beli szóra nem teljesül a pumpálási lemma, a nyelv nem reguláris.

(e) Tegyük fel, hogy a nyelv reguláris. Ekkor igaz rá a pumpálási lemma, legyen $p > 0$ az ebből adódó pumpálási hossz. Válasszuk a $z = (\mathtt{ab})^p(\mathtt{ba})^p$ szót.

Erre $z \in L_e$ és $|z| = 4p \geq p$ teljesül. Azt kell megmutatni, hogy ennek minden $z = uvw$ felosztásához, ahol $|uv| \leq p$ és $|v| > 0$ van olyan $k \geq 0$ szám, amire $uv^k w \notin L_e$.

Vegyük észre, hogy L_e minden nem üres szavában a bb egyetlen helyen, középen fordul elő, mint résszszó (egy $4n$ hosszú szóban a $(2n, 2n+1)$ helyeken). Mivel a pumpálandó v rész ez elé esik, ezért pl. $k = 2$-re ez a bb részszó középről hátrébb tolódik ($|v| = t \geq 1$ esetén $|uvvw| = 4n + t$, és a bb helye a $(2n + t, 2n + t + 1)$, nem pedig ahogy lennie kellene a $(2n + t/2, 2n + t/2 + 1)$, tehát kikerülünk a nyelvből.

Megjegyzés: az észrevétel nélkül esetszétválasztással is megoldhattuk volna.

7. Legyen $\Sigma = \{\mathtt{0},\mathtt{1}\}$. A pumpálási lemma segítségével igazolja, hogy az alábbi nyelvek nem regulárisak!
   (a) $L_a = \{s \in \Sigma^* :$ van olyan $x, y \in \Sigma^*,$ hogy $|x| = |y|$ és $s = x\mathtt{0}y\}$.
   (b) $L_b = \{s \in \Sigma^* :$ van olyan $x \in \Sigma^*,$ hogy $s = xx\}$.

Megoldás:

(a) Tegyük fel, hogy a nyelv reguláris. Ekkor igaz rá a pumpálási lemma, legyen $p > 0$ az ebből adódó pumpálási hossz. Mivel a nyelv azokból a szavakból áll, amiknek a közepén nulla van, célszerű olyan szót választani, amin pumpálás után könnyen látszik, hogy a középső karakter továbbra is nulla vagy nem nulla.

Válasszuk pl. a $z = \mathtt{1}^p\mathtt{0}\mathtt{1}^p$ szót. Ez persze benne van a nyelvben, és a hossza $2p + 1 \geq p$. A pumpálási lemma szerinti felosztásoknál a v rész a szó elején levő 1-k közül tartalmaz néhányat, legyen $v = \mathtt{1}^t$, ahol a feltételek szerint $0 < t \leq p$. Ezért pl. a $k = 2$ esetben az $uv^2w = \mathtt{1}^{p+t}\mathtt{0}\mathtt{1}^p$ szóban az egyetlen nulla már nem középen lesz, ezért ez a szó nincs az L_a nyelvben, ami ellentmond a pumpálási lemmának. Tehát a nyelv nem reguláris.

(b) Tegyük fel, hogy a nyelv reguláris. Ekkor igaz rá a pumpálási lemma, legyen $p > 0$ az ebből adódó pumpálási hossz. Itt is hasonló ötlet segít mint az előbb.

Legyen most a szó pl. $z = \mathtt{1}^p\mathtt{0}\mathtt{1}^p\mathtt{0}$. Ez az $x = \mathtt{1}^p\mathtt{0}$ választással $z = xx \in L_b$ és $|z| = 2p + 2 \geq p$.

A pumpált v részszó az első 1-blokk része lesz. Ha ezt megismételjük, akkor egy $\mathtt{1}^{p+t}\mathtt{0}\mathtt{1}^p\mathtt{0}$ alakú szóhoz jutunk. Ez viszont nincs a nyelvben, mert csak két 0 szerepel benne és az egyik a szó végén, így ha yy alakban szeretnénk felírni, az y-nak 0-ra kellene végződni, de a két 0 előtt nem ugyanannyi 1 áll, tehát a két rész nem lehet egyforma.

Tehát nem igaz rá a pumpálási lemma, a nyelv nem reguláris.

(Ha pl. a $z = 1^{2p} \in L_b$ szót választottuk volna, az nem működne, mert igaz ugyan, hogy a pumpáláskor az eredeti közepe a szónak „elcsúszik", de ettől még a szó a nyelvben maradhat. Valóban ez történik, ha olyan felbontást veszünk, amiben a pumpált v szakasz páros hosszú. Ezért ezzel a z választással nem kapunk ellentmondást.)

8. Legyen $L_1 = \{\mathtt{a}^n\mathtt{b}^k\mathtt{c}^k : n \geq 1, k \geq 0\}$, $L_2 = \{\mathtt{b}^k\mathtt{c}^m : k, m \geq 0\}$ és $L = L_1 \cup L_2$. Igazolja, hogy L_1 nem reguláris, L_2 reguláris, L pumpálható és L nem reguláris!

   Megoldás: L_1 nem reguláris: Tegyük fel, hogy reguláris. Akkor a metszete a reguláris $\mathtt{ab}^*\mathtt{c}^*$ nyelvvel is reguláris. Ez viszont az $\{\mathtt{ab}^k\mathtt{c}^k : k \geq 0\}$ nyelv, amiről a pumpálási lemma segítségével megmutatjuk, hogy nem reguláris. Ehhez legyen $p > 0$ a pumpálási hossz, és vegyük a $z = \mathtt{ab}^p\mathtt{c}^p \in L_1$ szót, aminek hosszára teljesül a lemma feltétele, $|z| = 2p + 1 \geq p$. Ennek minden, a pumpálási lemmának megfelelő $z = uvw$ felosztásában v vagy $v = \mathtt{a}$ vagy $v = \mathtt{ab}^t$ és $0 < t < p$ vagy $v = \mathtt{b}^t$ alakú, ahol pedig $0 < t \leq p$. Az uv^2w szóban az első két esetben két $\mathtt{a}$ betű is lesz, a harmadik esetben pedig a $\mathtt{b}$-k száma el fog térni a $\mathtt{c}$-k számától, tehát valóban minden felosztásra teljesül, hogy $uv^2w \notin L_1$, és ezért a nyelv nem reguláris.

   L_2 reguláris: ez a $\mathtt{b}^*\mathtt{c}^*$ nyelv, tehát reguláris.

   L pumpálható: ha egy L_2-beli szót veszünk, akkor a pumpáltjai is L_2-ben vannak, tehát L-ben is. Egy tetszőleges L_1-beli szónál, a pumpált szakasz álljon egyetlen $\mathtt{a}$ betűből (tehát pl. $u = \varepsilon, v = \mathtt{a}$, w pedig a szó további része.) Ekkor a pumpálással csak az $\mathtt{a}$-k száma változik. Ha ez nő ($k > 1$), akkor a pumpált szó is L_1-ben lesz. A lefelé pumpálásnál ($k = 0$) továbbra is L_1-ben marad, ha az eredeti szó legalább 2 db $\mathtt{a}$-val kezdődött. Abban az esetben, amikor csak 1 db $\mathtt{a}$ van a szó elején, a lefelé pumpálás átvisz az L_2 nyelvbe, tehát ilyenkor is L-ben lesz a pumpált szó. Ebből látszik, hogy valóban minden L-beli szónak van olyan felosztása, ami pumpálható.

   L nem reguláris: Az $\mathtt{a}$, $\mathtt{ab}$, $\mathtt{abb}$, $\mathtt{abbb}$ ... szavak páronként megkülönböztethetők, hiszen $\mathtt{a}$ $\mathtt{b}^k$ csak a $\mathtt{c}^k$-val kiegészítve lesz L-ben, más $\mathtt{c}^\ell$-lel kiegészítve nem. Azt meg tudjuk, hogy ha van végtelen sok páronként megkülönböztethető szó a nyelvben, akkor a nyelv nem reguláris („végtelen sok állapot kellene”)

....