6. Algoritmikus kérdések, Chomsky-normálforma

1. Adott $L_1, L_2 \subseteq \{\mathtt{a}, \mathtt{b}\}^*$ nyelvekhez legyen $L_\oplus$ az azokból a szavakból álló nyelv, melyek az L_1 és L_2 nyelvek közül pontosan az egyikben vannak benne. Igazolja, hogy van olyan L_1 és L_2 környezetfüggetlen nyelv, melyre
   (a) $L_\oplus$ is környezetfüggetlen!
   (b) $L_\oplus$ nem környezetfüggetlen!

Megoldás:

(a) Legyen pl. L_1 tetszőleges CF nyelv és $L_2 = L_1$. Ekkor $L_\oplus = \emptyset$, ami környezetfüggetlen (reguláris is). Egy másik egyszerű példa: $L_1 = \emptyset$, L_2 tetszőleges CF nyelv. Ekkor $L_\oplus = L_2$.

(b) Legyen $L_1 = \{\mathtt{a}, \mathtt{b}\}^*$. Ekkor $L_\oplus = \overline{L_2}$. Tehát ha L_2 egy olyan CF nyelv, aminek komplementere nem CF (tudjuk, hogy van ilyen), akkor $L_\oplus$ nem környezetfüggetlen.

2. Lehetséges-e, hogy ha az $L_1, L_2 \subseteq \Sigma^*$ nyelvek környezetfüggetlenek, akkor az $L_1 \cup L_2$ és az $L_1 \cap L_2$ nyelvek közül
   (a) pontosan az egyik környezetfüggetlen?
   (b) mindkettő környezetfüggetlen?
   (c) egyik sem környezetfüggetlen?

Megoldás: Tudjuk, hogy a CF nyelvek zártak az unióra, de a metszetre nem. Ennek alapján

(a) Lehet: L_1 és L_2 legyen olyan CF, amiknek metszete nem CF.

(b) Lehet, pl. $L_1 = L_2$ esetén az uniójuk és a metszetük is az L_1, tehát CF.

(c) Nem lehet, hiszen az unió mindig CF.

3. Adott egy $L \subseteq \{\mathtt{0}, \mathtt{1}\}^*$ reguláris nyelv. Az a kérdés, hogy tartalmaz-e minden `0`-val kezdődő szót. Adjon eljárást, ami ezt a kérdést eldönti, ha a nyelv megadásának módja
   (a) véges automata
   (b) reguláris kifejezés
   (c) reguláris nyelvtan

Megoldás: Jelölje L_0 a `0`-val kezdődő szavak nyelvét. Ekkor a kérdés: Igaz-e, hogy $L \cap L_0 = L_0$?

(a) Vegyük L_0-hoz a minimálautomatát. Ebből és az L-et megadó véges automatából elkészítjük a metszethez tartozó véges automatát, majd ezt minimalizáljuk. Ha az eredmény izomorf az L_0 minimálautomatájával, akkor a válasz igen, különben a válasz nem.

(b) Készítsünk a reguláris kifejezésből véges automatát. Innen ugyanaz, mint az (a). (Előzőleg elhagyhatjuk a reguláris kifejezésből az `1`-gyel kezdődő tagokat, így a kezdő automatánk kisebb lehet.)

(c) Ha elhagyjuk az $S \to \mathtt{1}A$ típusú szabályokat, az nem elég, mert az is kell, hogy minden L_0-beli generálható. De az automatára áttérés itt is működik – ezt akár csinálhatjuk a teljes nyelvtanból vagy az $S \to \mathtt{1}A$ alakú szabályok elhagyása után.

4. Adott egy M nemdeterminisztikus véges automata és egy R reguláris kifejezés. Vázoljon algoritmust annak eldöntésére, hogy
   (a) az M által elfogadott $L(M)$ nyelv megegyezik-e a reguláris kifejezés $L(R)$ nyelvével,
   (b) a két nyelv csak véges sok szóban különbözik-e!

Megoldás:

(a) Készítsünk a reguláris kifejezésből a tanult algoritmussal véges automatát. Ezt determinizáljuk és minimalizáljuk. Most már csak azt kell ellenőrizni, hogy az M determinizálása és minimalizálása után kapott automatával izomorf DVA-t kaptunk-e.

(b) Kezdjük, mint az előbb: készítsünk a reguláris kifejezésből a tanult algoritmussal véges automatát. Ezt determinizáljuk és minimalizáljuk, legyen ez M_R. Determinizáljuk az M automatát is, ez legyen M'. Az a kérdés, hogy az $L = (L(M_R) - L(M')) \cup (L(M') - L(M_R))$ nyelv véges-e. Ennek eldöntéséhez készítsük el az L-hez tartozó minimálautomatát, felhasználva a különbséghez és az unióhoz tartozó konstrukciót (amiből kapott automatát aztán determinizálni, minimalizálni kell). Most már csak azt kell ellenőrizni, hogy ez a DVA nem tartalmaz olyan kört, amiből elérhető elfogadó állapot.

Az utóbbi rész (minimalizálás, kör keresés) helyettesíthető annak ellenőrzésével, hogy a kapott automata nem fogad el egyetlen olyan szót sem, melynek hossza p és $2p$ közé esik. (Itt a pumpálási hossz helyett választhatunk annál nagyobb p értéket is amire pl. a determinisztikus teljes automata állapotainak száma megfelelő.)

5. Az $L \subseteq \{\mathtt{0}, \mathtt{1}\}^*$ nyelvben csak véges sok szó van, és ezek közé tartozik a 10 db 0-ból álló szó is. Igazolja, hogy a nyelv reguláris de a minimálautomatája legalább 11 állapotú!

Megoldás: Jelölje s a minimálautomata állapotszámát.

1. változat – használjuk a tanultakat: Minden véges nyelv reguláris (lehet pl. NVA-t készíteni, amiben minden nyelvbeli szóhoz egy-egy külön út tartozik, amikben csak a kezdőállapot közös. Hasonlóan lehet egyszerű reguláris nyelvtant is adni). Tudjuk, hogy a nyelv pontosan akkor véges, ha minden szavának a hossza legfeljebb p, ami a pumpálási hosszt jelöli. Tehát most $10 < p$. A pumpálási lemma bizonyításában p egy determinisztikus teljes véges automata állapotszáma, azaz lehet $p = s$. Ezért $10 < s$.

2. változat – elemi: A 10 hosszú szóhoz egy 11 állapotot tartalmazó elfogadó számítási út tartozik. Ha ennek állapotai között van ismétlődés, akkor van egy kör a számítási útban, ami akárhányszor ismételhető, azaz végtelen a nyelv. Tehát minden VA legalább 11 állapotú.

6. Hozza a következő nyelvtanokat Chomsky-normálformájúra! Milyen nyelvet generálnak a felsorolt nyelvtanok?
   (a) $S \to \mathtt{a}S\mathtt{a} \mid \mathtt{ab}$
   (b) $S \to \mathtt{a}S\mathtt{a} \mid \mathtt{b}S\mathtt{a} \mid \varepsilon$
   (c) $S \to \mathtt{a}A\mathtt{b}B\mathtt{c} \mid \mathtt{a}C\mathtt{b}D\mathtt{c}$, $A \to \mathtt{a}A\mathtt{b} \mid \mathtt{ab}$, $B \to B\mathtt{c} \mid \mathtt{c}$, $C \to \mathtt{a}C \mid \mathtt{a}$, $D \to \mathtt{b}D\mathtt{c} \mid \mathtt{bc}$
   (d) $E \to E + E \mid E * E \mid \mathtt{a}$

Megoldás:

(a) A nyelv: $\{\mathtt{a}^{k+1}\mathtt{ba}^k : k \geq 0\}$

1. rész (minden legalább 2 hosszú jobb oldal csak változókból áll):

$$S \to X_a S X_a \mid X_a X_b \qquad X_a \to \mathtt{a} \qquad X_b \to \mathtt{b}$$

2. rész (feldarabolás):

$$S \to X_a Y \mid X_a X_b \qquad Y \to S X_a \qquad X_a \to \mathtt{a} \qquad X_b \to \mathtt{b}$$

(b) A nyelv: páros hosszú szavak, a felétől kezdve csupa a betű.

Ez nem egy szabályos CF nyelvtan, előbb az ε-szabálytól meg kell szabadulni. Ennek eredménye:

$$S' \to \varepsilon \mid S \qquad S \to \mathtt{a}S\mathtt{a} \mid \mathtt{aa} \mid \mathtt{b}S\mathtt{a} \mid \mathtt{ba}$$

A keletkezett egyszeres szabályt is meg kell szüntetni:

$$S' \to \varepsilon \mid \mathtt{a}S\mathtt{a} \mid \mathtt{aa} \mid \mathtt{b}S\mathtt{a} \mid \mathtt{ba} \qquad S \to \mathtt{a}S\mathtt{a} \mid \mathtt{aa} \mid \mathtt{b}S\mathtt{a} \mid \mathtt{ba}$$

Most jöhet az 1. rész:

$$S' \to \varepsilon \mid X_a S X_a \mid X_a X_a \mid X_b S X_a \mid X_b X_a \qquad S \to X_a S X_a \mid X_a X_a \mid X_b S X_a \mid X_b X_a \qquad X_a \to \mathtt{a} \qquad X_b \to \mathtt{b}$$

2. rész (az algoritmust követve, nem optimalizálunk!)

$$S' \to \varepsilon \mid X_a Y \mid X_a X_a \mid X_b Z \mid X_b X_a \qquad S \to X_a R \mid X_a X_a \mid X_b T \mid X_b X_a \qquad X_a \to \mathtt{a} \qquad X_b \to \mathtt{b}$$

$$Y \to SX_a \qquad Z \to SX_a \qquad R \to SX_a \qquad T \to SX_a$$

(A 4 legutóbb bevezetett változót lehetne egyetlen változóval helyettesíteni.)

(c) Az A változóból levezethetők: $L_A = \{\mathtt{a}^k\mathtt{b}^k : k \geq 1\}$, a B változóból levezethetők: $L_B = \{\mathtt{c}^\ell : \ell \geq 1\}$, a C változóból levezethetők: $L_C = \{\mathtt{a}^m : m \geq 1\}$, a D változóból levezethetők: $L_D = \{\mathtt{b}^n\mathtt{c}^n : n \geq 1\}$. Ezek alapján a nyelv: $\mathtt{a}L_a\,\mathtt{b}\,L_B\,\mathtt{c} \cup \mathtt{a}\,L_C\,\mathtt{b}\,L_D\,\mathtt{c} = \{\mathtt{a}^p\mathtt{b}^q\mathtt{c}^r : p, q, r \geq 2, p = q \text{ vagy } q = r\}$

Átalakítás 1. rész:

$$S \to X_aAX_bBX_c \mid X_aCX_bDX_c$$

$$A \to X_aAX_b \mid X_aX_b \qquad B \to BX_c \mid \mathtt{c} \qquad C \to X_aC \mid \mathtt{a} \qquad D \to X_bDX_c \mid X_bX_c$$

$$X_a \to \mathtt{a} \qquad X_b \to \mathtt{b} \qquad X_c \to \mathtt{c}$$

2. rész

$$S \to X_aY_1 \mid X_aZ_1 \qquad Y_1 \to AY_2 \qquad Y_2 \to X_bY_3 \qquad Y_3 \to BX_c \qquad Z_1 \to CZ_2 \qquad Z_2 \to X_bZ_3 \qquad Z_3 \to DX_c$$

$$A \to X_aR \mid X_aX_b \qquad R \to AX_b \qquad B \to BX_c \mid \mathtt{c} \qquad C \to X_aC \mid \mathtt{a} \qquad D \to X_bT \mid X_bX_c \qquad T \to DX_c$$

$$X_a \to \mathtt{a} \qquad X_b \to \mathtt{b} \qquad X_c \to \mathtt{c}$$

(d) A nyelv az (egyszerű, zárójel nélküli) aritmetikai kifejezések nyelve.

1. rész:

$$E \to EX_+E \mid EX_*E \mid \mathtt{a} \qquad X_+ \to + \qquad X_* \to *$$

2. rész

$$E \to EY \mid EZ \mid \mathtt{a} \qquad Y \to X_+E \qquad Z \to X_*E \qquad X_+ \to + \qquad X_* \to *$$

7. Egy $A \to \alpha$ nyelvtani szabály hossza legyen $1 + |\alpha|$. Egy CF nyelvtan hossza jelentse a benne levő szabályok hosszainak összegét.

   Legyen $\Sigma = \{\mathtt{0}, \mathtt{1}\}$ és G egy egyszeres szabályok nélküli CF nyelvtan, amiben n változó van és a nyelvtan hossza N. A G-ből a tanult módon elkészítjük a Chomsky-normálformájú G' nyelvtant. Az n és N paraméterek segítségével adjon felső becslést G' változóinak számára és a G' nyelvtan hosszára!

   Megoldás: A régi változók mellett Σ minden karakteréhez tartozhat új változó és egy új szabály. Egy $A \to \alpha$ szabályból pedig $|\alpha| - 2$ új változó keletkezik és ezt $|\alpha| - 1$ új 3 hosszú szabállyal helyettesítjük. Tehát $h = 1 + |\alpha|$ hosszból $3(|\alpha| - 1) = 3(h - 2) \leq 3h$ hossz lesz. Vegyük észre, hogy ez $|\alpha| = 2$ esetében is igaz, és a végső becslés az $A \to \mathtt{a}$ alakúaknál is helyes (az utóbbi esetekben a hossz nem változik).

   A CNF-ben a változók száma $n' \leq n + |\Sigma| + N - 3 = n + N - 1$.

   Az új nyelvtan hossza $N' \leq 2|\Sigma| + 3N \leq 3N + 6$.

. . . .