T-79.148 **Kevät 2004**
Tietojenkäsittelyteorian perusteet
Harjoitus 6
Demonstraatiotehtävien ratkaisut

Lemma (Pumppauslemma). Olkoon A säännöllinen kieli. Tällöin on olemassa $n \geq 1$ siten, että kaikki A:n merkkijonot x , joiden pituus $|x| \geq n$ ovat ilmaistavissa muodossa $x = uvw$, missä $|uv| \leq n$, $|v| \geq 1$ ja merkkijonot muotoa uv^iw kuuluvat kieleen A kaikilla $i \geq 0$.
Kompaktimmin, painottaen pumppauslemman asettamia vaatimuksia, voimme kirjoittaa seuraavasti:

$\forall$ säännöllisille kielille A

$\exists n \geq 1$ s.e.
$\forall x \in A : |x| \geq n$
$\exists$ osiinjako $x = uvw$, missä $|uv| \leq n$ ja $|v| \geq 1$
$\forall i \geq 0 \; uv^iw \in A$.

Pumppauslemmaa voidaan käyttää hyväksi, kun halutaan osoittaa kieli L *ei-säännölliseksi*. Tehdään ensin vastaoletus, eli oletetaan L säännölliseksi kieleksi. Tavoitteena on päästä ristiriitaan tämän oletuksen kanssa seuraten pumppauslemman asettamia vaatimuksia säännöllisille kielille.
Pumppauslemmaa käytettäessä täytyy aina muistaa, että sillä voi osoittaa vain kielen epäsäännöllisyyden, ja sitä *ei* voi käyttää toiseen suuntaan. Esimerkiksi kieli

$$I = \{c^i a^n b^n \mid i > 0 \land n \geq 0\} \cup L(a^* b^*)$$

ei ole säännöllinen, mutta kaikki siihen kuuluvat sanat (tyhjää sanaa lukuunottamatta) voidaan osittaa pumppauslemman ehtojen mukaisesti. Näin ollen kieltä I ei voida suoraan todistaa epäsäännölliseksi, vaan todistuksessa täytyy käyttää apuna säännöllisten kielten sulkeumaominaisuuksia 5. tehtävän vastauksessa esitettävään tapaan.
Jos halutaan osoittaa kieli säännölliseksi, voidaan muodostaa sen hyväksyvä äärellinen automaatti, sillä pätee: Kieli L on säännöllinen $\Leftrightarrow$ on olemassa äärellinen automaatti M, joka hyväksyy kielen L (merkitään $L(M) = L$).

4. **Tehtävä**: *Hahmolausekkeet* ovat esimerkiksi UN*X-järjestelmien tekstityökaluissa käytetty säännöllisten lausekkeiden yleistys, jossa sallitaan merkkijonoarvoisten muuttujien käyttö lausekkeissa. Sovitettaessa merkkijonoa annettuun lausekkeeseen vaaditaan, että tietynnimisen muuttujan arvoksi tulee eri kohdissa sama osamerkkijono. Siten esimerkiksi aXb^*Xa ja $aX(a \cup b)^*YX(a \cup b)^*Ya$ ovat aakkoston $\{a, b\}$ hahmolausekkeita, joista ensimmäinen kuvaa kielen $\{awb^nwa \mid w \in \{a, b\}^*,\ n \geq 0\}$. Osoita, että hahmolausekkeet ovat säännöllisten lausekkeiden aito yleistys, so. niillä voidaan kuvata myös joitakin ei-säännöllisiä kieliä.

   Vastaus: Osoittaaksemme, että tehtävän hahmolausekkeet on säännöllisten lausekkeiden aito yleistys, tulee meidän löytää hahmolauseke, jonka määrittämä kieli ei ole säännöllinen.

   Tarkastellaan hahmolauseketta XX vastaavaa kieltä $L = \{zz \mid z = \{a, b\}^*\}$. Oletetaan, että L on säännöllinen. Valitaan $x = a^n ba^n b \in L$, missä n on pumppauslemmassa esiintyvä kielestä L riippuva kokonaisluku. Nyt $|x| = 2n + 2 > n$. Pumppauslemman mukaan voidaan kirjoittaa $x = uvw$, missä $|uv| \leq n$ ja $|v| \geq 1$. Siis $u = a^{n-|v|-k}$, $v = a^{|v|}$ ja $w = a^k ba^n b$, missä $0 \leq k < n$. Nyt pumppauslemman mukaan kaikille $i \geq 0$ tulisi päteä $uv^iw \in L$. Kuitenkin $uv^0w = uw = a^{n-|v|}ba^n b \notin L$, sillä se ei ole muotoa zz, koska vaadittiin $|v| \geq 1$. Päädyttiin siis ristiriitaan oletuksen kanssa. L ei näin ollen voi olla säännöllinen.

   Löydettiin siis ei-säännöllinen kieli, joka voidaan kuvata hahmolausekkeella. Näin ollen hahmolausekkeet ovat säännöllisten lausekkeiden aito yleistys. $\square$

5. **Tehtävä**: Osoita, että kieli $\{w \in \{a,b\}^* \mid w\text{:ssä on yhtä monta } a\text{:ta ja } b\text{:tä}\}$ ei ole säännöllinen, ja laadi yhteydetön kielioppi sen kuvaamiseen.

   Vastaus: Kielen $L = \{w \in \{a,b\}^* \mid w\text{:ssä on yhtä monta } a\text{:ta ja } b\text{:tä}\}$ voisi todistaa ei-säännölliseksi suoraan pumppauslemmalla. Tässä esitetään kuitenkin hieman monimutkaisempi ratkaisu esimerkkinä siitä, miten "hankalia" kieliä voidaan käsitellä.

   Määritellään kieli $L' = L \cap L(a^*b^*)$. Oletetaan, että L on säännöllinen. Koska $L(a^*b^*)$ on säännöllinen ja säännöllisten kielten joukko on suljettu leikkauksen suhteen, täytyy myös L':n olla säännöllinen. (Toisinpäin ehto ei päde: L' voi olla säännöllinen vaikka L ei olisi, sillä esim. $A \cap \emptyset = \emptyset$ kaikille kielille A).

   Huomataan, että $L' = \{a^k b^k \mid k \geq 0\}$. Tarkastellaan sanaa $w = a^n b^n$, missä n on pumppauslemmassa esiintyvä parametri. Yritetään osittaa w lemman ehtojen mukaisesti. Koska $|xy| \leq n$, osituksen täytyy olla muotoa;

   $$\begin{aligned} x &= a^{n-i-k} \\ y &= a^i \\ z &= a^k b^n, \end{aligned}$$

   missä $0 < i \leq n$ ja $i + k \leq n$. Nyt $xz = a^{n-i} b^n$, joten $xz \notin L'$. Näin ollen sanaa w ei voida pumpata, eikä L' ole säännöllinen, joten myöskään L ei ole säännöllinen.

   Alla kielen L kuvaava yhteydetön kielioppi G:

   $G = (V, \Sigma, P, S)$, missä
   $V = \{S, T, a, b\}$,
   $\Sigma = \{a, b\}$,
   $P = \{\ S \to SS \mid aT \mid Ta \mid \varepsilon,$
   $\quad\quad T \to ST \mid TS \mid b\}$

   Tässä kieliopissa välikkeellä S johdetaan merkkijonot, joissa on molempia aakkosia yhtä monta, ja T:llä kaikki joissa on b-merkkejä yksi enemmän kuin a-merkkejä.

   Esimerkki. Annetaan johto merkkijonolle $aababb \in L$.

   $$\begin{aligned} S &\Rightarrow aT \\ &\Rightarrow aST \\ &\Rightarrow aaTT \\ &\Rightarrow aabT \\ &\Rightarrow aabST \\ &\Rightarrow aabaTT \\ &\Rightarrow aababT \\ &\Rightarrow aababb \end{aligned}$$

6. **Tehtävä**: Laadi yhteydetön kielioppi, joka tuottaa kaikki seuraavan esimerkin tapaiset, yksinkertaisista sisäkkäisistä `for`-silmukoista, `begin`- ja `end`-sulkeilla kootuista lauseista ja alkeisoperaatioista `a` rakentuvat "ohjelmat":

   ```
   a;
   for 3 times do
   begin
     for 5 times do a;
     a; a
   end
   ```

   Silmukkalaskureiden voit olettaa olevan kokonaislukuja väliltä 0,..., 9.

   Vastaus: Ohjelmointikielten kieliopit määritellään useimmiten siten, että aakkostoksi otetaan kielessä esiintyvät syntaktiset elementit (lekseemit). Tässä tapauksessa niitä ovat numerot, `a` sekä varatut sanat. Ohjelman jäsentäminen jaetaan kahteen osaan:

(a) Muutetaan ohjelman teksti jonoksi lekseemeitä tilakoneiden avulla.

(b) Muodostetaan lekseemijonon jäsennyspuu.

Tehtävän kieliopin voi määritellä monellakin eri tapaa, tässä on yksi mahdollinen tulkinta:

$$
\begin{aligned}
G =& (V, \Sigma, P, C) \\
V =& \{C, S, N, \textbf{begin}, \textbf{do}, \textbf{end}, \textbf{for}, \textbf{times}, 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, ;, a\} \\
\Sigma =& \{\textbf{begin}, \textbf{do}, \textbf{end}, \textbf{for}, \textbf{times}, 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, ;, a\}
\end{aligned}
$$

Tässä välikkeen S tulkintana on "lause" (*statement*), C:n "yhdistetty lause" (*compound statement*) ja N:n "numero". Kieliopin säännöt määritellään seuraavasti:

$$
\begin{aligned}
P = \{&C \to S \mid S; C \\
&S \to a \mid \textbf{begin}\ C\ \textbf{end} \mid \textbf{for}\ N\ \textbf{times}\ \textbf{do}\ S \\
&N \to 0 \mid 1 \mid 2 \mid 3 \mid 4 \mid 5 \mid 6 \mid 7 \mid 8 \mid 9\}
\end{aligned}
$$

Esimerkki. Tehtävänannossa esiintyneen ohjelman jäsennyspuu: