T-79.148 **Kevät 2006**
Tietojenkäsittelyteorian perusteet
Harjoitus 12
Demonstraatiotehtävien ratkaisut

4. **Tehtävä**: Osoita, että yhteysherkät kielet voidaan tunnistaa lineaarisesti rajoitetuilla automaateilla. (Käytä hyväksesi sitä, että kieliopin produktioita sovellettaessa lausejohdoksen pituus ei voi koskaan lyhentyä, paitsi tyhjän merkkijonon muodostamassa erikoistapauksessa.) Päättele edellisen perusteella, että kaikki yhteysherkät kielet ovat rekursiivisia.

   Vastaus: Yhteysherkän kieliopin määritelmän mukaan produktiot ovat sellaisia, että yksittäisessä säännössä oikea puoli on aina vähintään yhtä pitkä kuin vasen.

   Yhteysherkkä kieli voidaan tunnistaa lineaarisesti rajoitetulla automaatilla, joka epädeterminisesti siirtyy johonkin kohtaan syötettä ja soveltaa jotain sääntöä oikealta vasemmalle. Koska merkkijono tällöin lyhenee, lisätilaa ei tarvita. Toisaalta, jos väliin jäisi tyhjiä merkkejä, voidaan implementoida suoraviivainen tiivistys. Jos nauhalle jää pelkkä kieliopin aloitussymboli, siirrytään hyväksyvään lopputilaan.

   Tarkastellaan esimerkiksi yhteysherkkää kielioppia:

$$\begin{aligned} S &\rightarrow aA \mid bB \\ aA &\rightarrow abB \mid ab \\ bB &\rightarrow baA \mid ba \end{aligned}$$

   Tässä kielen tunnistava lineaarisesti rajoitettu automaatti muokkaisi syötenauhaansa seuraavaan tapaan (syötteellä $abab$):

$$\begin{array}{|c|c|c|c|c|c|}\hline > & \underline{\mathrm{a}} & \mathrm{b} & \mathrm{a} & \mathrm{b} & < \\ \hline\end{array} \vdash^* \begin{array}{|c|c|c|c|c|c|}\hline > & \mathrm{a} & \mathrm{b} & \underline{\mathrm{a}} & \mathrm{A} & < \\ \hline\end{array}$$

$$\vdash^* \begin{array}{|c|c|c|c|c|}\hline > & \mathrm{a} & \underline{\mathrm{b}} & \mathrm{B} & < \\ \hline\end{array}$$

$$\vdash^* \begin{array}{|c|c|c|c|}\hline > & \underline{\mathrm{a}} & \mathrm{A} & < \\ \hline\end{array}$$

$$\vdash^* \begin{array}{|c|c|c|}\hline > & \underline{\mathrm{S}} & < \\ \hline\end{array}$$

   Yllä kuvattu kone ei kuitenkaan ole totaalinen, sillä on mahdollista, että sen laskenta ei koskaan pysähdy. Esimerkiksi, jos käsiteltävä kielioppi on:

$$\begin{aligned} S &\rightarrow \varepsilon \\ ab &\rightarrow ba \\ ba &\rightarrow ab, \end{aligned}$$

   niin kaikki laskennat syötteellä, jossa on ainakin yksi ab tai ba pari jäävät ikuiseen silmukkaan.

   Asian voi korjata huomaamalla, että koska koneen nauhan pituus ei voi kasvaa, on koneella on vain rajallinen määrä mahdollisia tilanteita. Näiden tilanteiden lukumäärä on suuruusluokkaa $q \times n \times |\Gamma|^n$, missä q on tilojen lukumäärä, n syötteen pituus ja $|\Gamma|$ aakkoston koko.

   Edellä mainitulla tavalla rakennettu kone saadaan totaaliseksi liittämällä siihen laskuri, joka laskee suoritettujen askelten määrän. Helpoiten tämä käy siten, että tehdään koneesta kaksiurainen, jonka toiselle uralle kirjoitetaan askelten määrä binäärilukuna, jota kasvatetaan aina kunkin alkuperäisen koneen laskenta-askeleen yhteydessä yhdellä. Kun laskuri ylittää raja-arvon, sana hylätään, sillä tällöin kone on joutunut silmukkaan.

Lopuksi täytyy tarkistaa, että laskuri voidaan toteuttaa rikkomatta vaatimusta lineaarisesta tilankäytöstä. Luvun $q \times n \times |\Gamma|^n$ esittämiseen tarvitaan $k = \log_2(q) + \log_2(n) + n\log(|\Gamma|)$ bittiä, joka kasvaa lineaarisesti n:n kasvaessa. Vaikka tässä $k > n$, niin laskuri saadaan puristettua käytössä olevaan tilaan esittämällä se sopivassa kannassa.

5. **Tehtävä**: Osoita, että jokainen rajoittamattomalla kieliopilla tuotettava kieli voidaan tuottaa kieliopilla, jossa produktioiden vasemmalla puolella ei esiinny päätemerkkejä.

   Vastaus: Kieliopin G kanssa saman kielen tuottavan tehtävänannon mukaisen kieliopin G' voi muodostaa systemaattisesti siten, että jokaista kielessä esiintyvää päätemerkkiä a kohden otetaan uusi välike A_a ja tälle sääntö $A_a \to a$. Tällä uudella välikkeellä korvataan kaikki päätemerkit a, jotka esiintyvät G:n säännöissä.

   Seuraavaksi sama formaalimmin: Olkoon $G = (V, \Sigma, P, S)$ yleinen kielioppi. Muodostetaan kielioppi $G' = (V', \Sigma, P', S')$, missä

$$V' = V \cup \{A_a \mid a \in \Sigma\}$$

   Kukin kieliopin G produktio $r = x_1 \cdots x_n \to x_{n+1} \cdots x_{n+m}$, missä $x_i \in V$ muutetaan muotoon:

$$c(r) = y_1 \cdots y_n \to y_{n+1} \cdots y_{n+m}$$

   missä

$$y_i = \begin{cases} x_i, & x_i \in V - \Sigma \\ A_{x_i}, & x_i \in \Sigma \ . \end{cases}$$

   Nyt voidaan määritellä sääntöjoukko P' seuraavasti:

$$P' = \{c(r) \mid r \in P\} \cup \{A_a \to a \mid a \in \Sigma\} \ .$$

   Tarkastellaan uudelleen edellisessä tehtävässä esiintynyttä kielioppia:

$$\begin{aligned} S &\to aA \mid bB \\ aA &\to abB \mid ab \\ bB &\to baA \mid ba \end{aligned}$$

   Konstruktion tuloksena syntyy kielioppi:

$$\begin{aligned} S &\to A_aA \mid A_bB \\ A_aA &\to A_aA_bB \mid A_aA_b \\ A_bB &\to A_bA_aA \mid A_bA_a \\ A_a &\to a \\ A_b &\to b \end{aligned}$$

6. **Tehtävä**: Osoita, että jokainen yhteysherkkä kielioppi voidaan saattaa normaalimuotoon, jossa produktiot ovat muotoa $S \to \varepsilon$ tai $\alpha A\beta \to \alpha\omega\beta$, missä A on kieliopin välike ja $\omega \neq \varepsilon$. (S on tässä kieliopin lähtösymboli.)

   Vastaus: Normaalimuotoon saattamisessa on kolme vaihetta:

   i) Poistetaan alkusymboli S sääntöjen oikealta puolelta.

   ii) Poistetaan kaikki päätemerkit sääntöjen vasemmalta puolelta.

   iii) Muutetaan väärää muotoa olevat säännöt oikeanlaisiksi.

   Vaiheet määritellään seuraavasti:

i) Mikäli S esiintyy jonkin säännön oikealla puolella, niin lisätään kielioppiin uusi alkusymboli S' ja sääntö $S' \to S$. (Yhteysherkkien kielioppien määritelmän mukaan tällöin ei sääntöjoukossa voi olla sääntöä $S \to \varepsilon$, joten tätä tapausta ei tarvitse käsitellä.)

ii) Päätemerkkien poistaminen sääntöjen vasemmalta puolelta tapahtuu käyttäen tehtävän 5. vastauksessa esitettyä konstruktiota.

iii) Jokaista väärää muotoa olevaa sääntöä:

$$X_1 \cdots X_n \to Y_1 \cdots Y_m \ ,$$

missä $m \geq n$, kohden lisätään kielioppiin $n-1$ uutta välikettä $(Z_1, \ldots, Z_{n-1})$, ja sääntö korvataan joukolla sääntöjä:

$$\begin{aligned}
X_{n-1}X_n &\to Z_1X_n \\
Z_1X_n &\to Z_1Y_n \cdots Y_m \\
X_{n-2}Z_1 &\to Z_2Z_1 \\
Z_2Z_1 &\to Z_2Y_{n-1} \\
&\vdots \\
Z_{n-1}Z_{n-2} &\to Z_{n-1}Y_2 \\
Z_{n-1}Y_2 &\to Y_1Y_2
\end{aligned}$$

Tarkastellaan, miten iii)-kohta käsittelee sääntöä:

$$ABBA \to BAABA$$

Koska $n = 4$, tarvitaan kolme uutta välikettä: Z_1, Z_2 ja Z_3. Sääntöjoukoksi tulee:

$$\begin{aligned}
BA &\to Z_1A \\
Z_1A &\to Z_1BA \\
BZ_1 &\to Z_2Z_1 \\
Z_2Z_1 &\to Z_2A \\
AZ_2 &\to Z_3Z_2 \\
Z_3Z_2 &\to Z_3A \\
Z_3A &\to BA \ .
\end{aligned}$$

Tällöin vastaavaksi johdoksi tulee:

$$\begin{aligned}
AB\underline{BA} &\Rightarrow AB\underline{Z_1A} \Rightarrow A\underline{BZ_1}BA \Rightarrow A\underline{Z_2Z_1}BA \Rightarrow \underline{AZ_2}ABA \\
&\Rightarrow \underline{Z_3Z_2}ABA \Rightarrow \underline{Z_3A}ABA \Rightarrow BAABA \ .
\end{aligned}$$