Tik-79.148 **Kevät 1999**
Tietojenkäsittelyteorian perusteet
Laskuharjoitus 3
Ratkaisut

1. (1.8.2 / 1.9.2)

   Säännölliset lausekkeet määritellään induktiivisesti:

   - $\emptyset$ ja kaikki $a \in \Sigma$ ovat säännöllisiä lausekkeita.
   - Mikäli α ja β ovat säännöllisiä lausekkeita, niin myös $(\alpha\beta), (\alpha \cup \beta)$ ja α^* ovat säännöllisiä lausekkeita.
   - Mitkään muut lausekkeet eivät ole säännöllisiä.

   Säännölliset lausekkeet on ehkä helpointa ajatella muottina. Sana kuuluu kieleen, jos sen saa sovitettua lausekkeen antamaan muottiin:

   - $ab \Rightarrow$ otetaan ensin a, sitten b
   - $a \cup b \Rightarrow$ joko a tai b
   - $a^* \Rightarrow$ otetaan kuinka monta a:ta tahansa (myös 0).

   Säännöllisten lausekkeiden unionia voidaan merkitä monilla eri tavoilla. Yleisimmät ovat $\cup$, $+$ sekä $|$. Useimmissa käytännön toteutuksissa määritellään lisäksi ylimääräisiä operaattoreita yksinkertaistamaan lausekkeita. Lausekkeiden ilmaisuvoima pysyy kuitenkin samana.

   Säännöllisiä lausekkeita käytetään mm. ohjelmointikielten kääntäjien syntaktiseen analyysiin sekä määrämuotoisen tekstin etsimiseen tekstitiedostoista (grep).

   a) Lausekkeen $\emptyset \cup a^* \cup b^* \cup (a \cup b)^*$ kuvaamaan kieleen kuuluu tyhjä sana e sekä kaikki sanat, joissa on joko pelkästään a-kirjaimia, pelkästään b-kirjaimia tai molempia sekaisin. Koska $\emptyset \subset (a \cup b)^*$, $a^* \subset (a \cup b)^*$ ja $b^* \subset (a \cup b)^*$, lauseke supistuu muotoon $(a \cup b)^*$.

   b) Lauseke $((a^*b^*)^*(b^*a^*)^*)^*$ kuvaa myös kielen $(a \cup b)^*$. Tämän voi nähdä jo ensimmäisestä sulkulausekkeesta: ensin otetaan 0 tai useampia a-kirjaimia, sitten 0 tai useampia b-kirjaimia ja toistetaan tätä prosessia niin monta kertaa kuin on tarpeen minkä tahansa sanan generoimiseksi.

c) Lausekkeen $(a^*b)^* \cup (b^*a)^*$ ensimmäinen osa kuvaa kaikki sanat, jotka päättyvät b-kirjaimeen, ja jälkimmäinen kaikki a-kirjaimeen päättyvät sanat. Näiden kahden unioni on taas kerran kaikkien sanojen joukko $(a \cup b)^*$.

d) Lauseke $(a \cup b)^*a(a \cup b)^*$ kuvaa kaikki sanat, joissa on ainakin yksi a. Keskellä olevaa a-kirjainta ei voi siirtää Kleenen tähden sisään, sillä silloin hyväksyttäisiin myös sanat, joissa ei ole yhtään a:ta. Yksi muutos sievennys lausekkeelle voidaan kuitenkin tehdä. Nykyisessä muodossa voidaan mikä tahansa sanassa olevista a-kirjaimista valita "pakolliseksi" merkiksi. Voimme kuitenkin valita aina viimeisen a-kirjaimen "pakolliseksi", jolloin lauseke supistuu muotoon: $(a \cup b)^*ab^*$.

2. (1.8.3 / 1.9.3)

Kaikissa tämän tehtävän kohdissa käytetään aakkostona $\Sigma = \{a, b\}$.

(a) Kieltä $L = \{w \,|\, w \in \Sigma^*, w\text{:ssä on korkeintaan 3 } a\text{:ta.}\}$ vastaava säännölinen lauseke on: $b^*(\emptyset \cup a)b^*(\emptyset \cup a)b^*(\emptyset \cup a)b^*$

Perusajatuksena on luetella kaikki mahdolliset a-kirjamen esiintymät, ja ympäröidä ne b-kirjaimilla. Jokaisen a:n kohdalla on myös mahdollisuus jättää se pois valitsemalla $\emptyset$.

(b) Kieltä $L = \{w \,|\, w : \text{ssä olevien } a : \text{den määrä on jaollinen kolmella}\}$ vastaa lauseke $b^*(ab^*ab^*ab^*)^*$

(c) Kieltä $L = \{w \,|\, w\text{:ssä on tasan yksi alimerkkinono } aaa\}$ vastaa lauseke: $(b \cup ab \cup aab)^*aaa(b \cup ba \cup baa)^*$. Alimerkkijonoiksi lasketaan myös päällekäiset esiintymät, eli aaa sisältyy kaksi kertaa merkkijonoon $aaaa$.

3. (1.8.5 / 1.9.5)

a) Väite $baa \in a^*b^*a^*b^*$ on tosi, sillä valitsemalla yksi ensimmäisistä b-kirjaimista ja kaksi toisista a-kirjaimista saadaan sana baa.

b) Väite $b^*a^* \cap a^*b^* = a^* \cup b^*$ on tosi, sillä

$$\begin{gathered} b^*a^* \cap a^*b^* = \\ \{e, b^+, a^+, b^+a^+\} \cap \{e, a^+, b^+, a^+b^+\} = \\ \{e, a^+, b^+\} = a^* \cup b^* \end{gathered}$$

Merkintä $a^+ = aa^*$

c) Väite $a^*b^* \cap c^*d^* = \emptyset$ on väärä, sillä

$$\begin{gathered} a^*b^* \cup c^*d^* = \\ \{e, b^+, a^+, b^+a^+\} \cap \{e, c^+, d^+, c^+d^+\} = \\ \{e\} \neq \emptyset \end{gathered}$$

Tässä kannattaa huomata, että yhtälön vasemmalla puolella oleva merkintä **ei** ole säännöllinen lauseke, sillä säännöllisessä lausekkeessa ei voi esiintyä leikkausta, vaan kyseessä on kahden säännöllisen kielen leikkaus.

d) Väite $abcd \in (a(cd)^*b)^*$ on myös väärä, sillä kaikki säännöllisen lausekkeen hyväksymät sanat (tyhjää sanaa lukuunottamatta) päättyvät b-kirjaimeen.

4. Väite: Jos kieli L on säännöllinen, niin myös kieli

$$L' = \{w \,|\, uw \in L \text{ jollekin merkkijonolle } u\}$$

on säännöllinen.

Kieleen L' kuuluvat kaikki ne merkkijonot, jotka esiintyvät kieleen L kuuluvien sanojen suffikseina. Esimerkiksi: jos

$$\begin{gathered} L = \{a, aba, abb\}, \text{ niin} \\ L' = \{e, a, b, ba, bb, aba, abb\} \end{gathered}$$

Tässä konstruoidaan systemaattinen tapa muodostaa L:n määrittelemästä säännöllisestä lausekkeesta R (eli $L(R) = L$) uusi lauseke R', jonka määrittelemä kieli on L' (eli $L(R') = L'$).

Formaalimmin, merkitään kaikkien säännöllisten lausekkeiden joukkoa symbolilla S. Nyt halutaan löytää jokin funktio $f : S \to S$, siten, että $f(R) = R'$.

Määritellään funktio f rekursiivisesti:

i) (perustapaus) $f(a) = a \cup \emptyset, a \in \Sigma \cup \{\emptyset\}$. Toisin sanoen, mikäli säännöllinen lauseke R on tyhjä tai pelkästään yksi aakkoston merkki, on lauseke R' saman merkin ja tyhjän lausekkeen unioni.

ii) $f(\alpha \cup \beta) = f(\alpha) \cup f(\beta), \alpha, \beta \in S$. Toisin sanoen, mikäli R on kahden lausekkeen α ja β unioni, niin sen määrittelemän kielen sanojen lopussa voi olla mitä tahansa, mitä voi olla jomman kumman alilausekkeen lopussa.

iii) $f(\alpha^*) = f(\alpha)\alpha^*$, eli sanan lopussa voi olla tähden vaikutusalueella oleva alilauseke kokonaisena mielivaltaisen monta kertaa, ja sen edellä jotain, mihin alilauseke voi päättyä

iv) $f(\alpha\beta) = f(\alpha)\beta \cup f(\beta)$, eli katenaatiolla muodostetun lausekkeen kuvaaman kielen sanojen lopussa voi joko olla jotain, mitä voi olla jälkimmäisen osan lopussa, tai sitten viimeinen osa on kokonaisuudessaan mukana, ja sen edellä on jotain, mihin ensimmäinen osa voi loppua.

Tämän jälkeen puuttuu enää formaali todistus ylläolevien sääntöjen oikeellisuudesta. Todistus sivuutetaan tässä, mutta sen voi rakentaa osoittamalla erikseen jokaisen säännön oikeellisuuden induktiolla.

Esimerkki: Olkoon $R = ba(aa \cup bab)^*$. Nyt

$$\begin{aligned}
f(R) &= f(ba(aa \cup bab)^*) \\
&= f(ba)(aa \cup bab)^* \cup f((aa \cup bab)^*) \\
&= (f(b)a \cup f(a))(aa \cup bab)^* \cup f(aa \cup bab)(aa \cup bab)^* \\
&= ((\emptyset \cup b)a \cup (\emptyset \cup a))(aa \cup bab)^* \cup (f(aa) \cup f(bab))(aa \cup bab)^* \\
&= (\emptyset \cup a \cup ba)(aa \cup bab)^* \cup ((f(a)a \cup f(a)) \cup (f(b)ab \cup f(a)b \cup f(b)))(aa \cup bab)^* \\
&= (\emptyset \cup a \cup ba)(aa \cup bab)^* \cup (\emptyset \cup a \cup aa \cup b \cup ab \cup bab)(aa \cup bab)^* \\
&= (\emptyset \cup a \cup aa \cup b \cup ab \cup bab \cup ba)(aa \cup bab)^* \\
&= (\emptyset \cup a \cup b \cup ab \cup ba)(aa \cup bab)^*
\end{aligned}$$