2.6 SÄÄNNÖLLISET LAUSEKKEET

Automaattimalleista poikkeava tapa kuvata yksinkertaisia kieliä.

Olkooot \(A\) ja \(B\) aakkoston \(\Sigma\) kieliä. Perusope
raatioita:

(i) Yhdiste:
\[
A \cup B = \{ x \in \Sigma^* \mid x \in A \text{ tai } x \in B \};
\]
(ii) Kätenaatio:
\[
AB = \{ xy \in \Sigma^* \mid x \in A, y \in B \};
\]
(iii) Potenssit:
\[
\begin{align*}
A^0 &= \{ \varepsilon \},
A^k &= AA^{k-1} = \{ x_1 \ldots x_k \mid x_i \in A \quad \forall i = 1, \ldots, k \}
\end{align*}
\]
(iv) Sulkeuma t. "Kleenem tähtr":
\[
A^* = \bigcup_{k \geq 0} A^k
= \{ x_1 \ldots x_k \mid k \geq 0, x_i \in A \quad \forall i = 1, \ldots, k \}.
\]

Määritelmä 2.3 Aakkoston \(\Sigma\) säännölliset lausekeet määrittävät lausekkeitä:

(i) \(\emptyset\) ja \(\varepsilon\) ovat \(\Sigma\)-säännöllisiä lausekkeita;
(ii) \(a\) on \(\Sigma\)-säännöllinen lauseke kaikilla \(a \in \Sigma\);
(iii) jos \(r\) ja \(s\) ovat \(\Sigma\)-säännöllisiä lausekkeita, niin \((r \cup s)\), \((rs)\) ja \(r^*\) ovat \(\Sigma\)-säännöllisiä lausekkeita;
(iv) muita \(\Sigma\)-säännöllisiä lausekkeita ei ole.

Kukin \(\Sigma\)-n säännöllinen lauseke \(r\) kuvaa kielen \(L(r)\), joka määritellään:

(i) \(L(\emptyset) = \emptyset\);
(ii) \(L(\varepsilon) = \{ \varepsilon \};
(iii) \(L(a) = \{ a \}\) kaikilla \(a \in \Sigma\);
(iv) \(L((r \cup s)) = L(r) \cup L(s)\);
(v) \(L((rs)) = L(r)L(s)\);
(vi) \(L(r^*) = (L(r))^*\).

Aakkoston \(\{a,b\}\) säännöllisiä lausekkeita:
\[
\begin{align*}
r_1 &= ((ab)b), \quad r_2 = (ab)^*,
&\quad r_3 = (ab^*), \quad r_4 = (a(b \cup (bb)))^*.
\end{align*}
\]

Lausekkeiden kuvaamat kielet:
\[
\begin{align*}
L(r_1) &= \{(a\{b\}\{b\} = \{a\}\{b\} = \{a\}\{b\} = \emptyset\};
L(r_2) &= \{a\}^* = \{ \varepsilon, a, ab, abab, ababab, \ldots \}
= \{ (ab)^i \mid i \geq 0 \};
L(r_3) &= \{(a\{b\})^* = \{ a, ab, abb, abbab, \ldots \}
= \{ a^i \mid i \geq 0 \};
L(r_4) &= \{(a\{b\}\{b\})^* = \{a, abb\}^*
= \{ \varepsilon, ab, abb, abab, ababab, \ldots \}
= \{ x \in \{a, b\}^* \mid \text{kutakin } a\text{-kirjainta } x:ssä
\text{ seuraa } 1 \text{ tai } 2 \text{ } b\text{-kirjainta } \}.
\end{align*}
\]
Sulkumerkkien vähentämissääntöjä:

Operaattoreiden prioriteetti:

\[
* \succ \cdot \succ \cup
\]

Yhdiste- ja tulo-operaatioiden associatiivisuus:

\[
L(((rU_s)Ut)) = L((rU(sUt))) \\
L(((rs)t)) = L((r(st)))
\]

→ peräkkäisiä yhdisteitä ja tuloja ei tarvitse suluttaa.

Käytetään tavallisia kirjasimia, mikäli sekaannuksen vaaraa merkkijonoihin ei ole.

Yksinkertaisemmin siis:

\[
\begin{align*}
    r_1 &= a\,b\,b, \quad r_2 = (ab)^*, \quad r_3 = ab^*, \quad r_4 = (a(b\cup b))^*.
\end{align*}
\]

### 2.4 Määritelmä
Kieli on säännöllinen, jos se voidaan kuvata säännöllisellä lausekkeella.

### Säännöllisten lausekkeiden sieventäminen

Säännöllisellä kielillä on yleensä useita vaihtoehtoisia kuvaauksia, esim.:

\[
\begin{align*}
    \Sigma^* &= L((a \cup b)^*) \\
             &= L((a^*b^*)^*) \\
             &= L(a^*b^* \cup (a \cup b)^*ba(a \cup b)^*)
\end{align*}
\]

Määritelmä. Säännölliset lausekkeet \( r \) ja \( s \) ovat **ekvivalentit**, merk. \( r = s \), jos \( L(r) = L(s) \).

Lausekkeen sieventäminen on ”yksinkertaisimman” ekvivalentin lausekkeen määrittäminen.

Säännöllisten lausekkeiden ekvivalentisuudesta on epätriviaali, mutta periaatteessa mukaan sesti ratkeava ongelma.

Sievennyssääntöjä:

\[
\begin{align*}
    r \cup (s \cup t) &= (r \cup s) \cup t \\
    r(st) &= (rs)t \\
    r \cup s &= s \cup r \\
    r(s \cup t) &= rs \cup rt \\
    (r \cup s)t &= rt \cup st \\
    r \cup r &= r \\
    s &= r \\
    \emptyset &= \emptyset \\
    r^* &= \varepsilon \cup r^* \\
    r^* &= (\varepsilon \cup r)^*
\end{align*}
\]

Mikä tahansa säännöllisten lausekkeiden tosi ekvivalenti voidaan johtaa näistä laskulaitoista, kun lisätään päättelysääntö:

jos \( r = rs \cup t \), niin \( r = ts^* \), edellytäen että \( \varepsilon \notin L(s) \).
2.7 ÄÄRELLISET AUTOMAATIT JA SÄÄNNÖLLISTEN KIELET

Lause 2.3 Jokainen säännöllinen kieli voidaan tunnistaa äärellisellä automaattilla.

Todistus. Seuraavan kalvon induktiivisen konstruktion avulla voidaan mielivaltaisen säännöllisen lausekkeen \( r \) rakennetta seuraten muodostaa \( \varepsilon \)-automaatti \( M_r \), jolla \( L(M_r) = L(r) \). Tästä automaattista voidaan poistaa \( \varepsilon \)-siirtymät Lemman 2.4 mukaisesti, ja tarvittaessa voidaan syntyvä epädeterministinen automaatti determinoida Lauseen 2.2 konstruktiolla.

Esitettävästä konstruktiosta on syytä huomata, että muodostettavien \( \varepsilon \)-automaatteihin tulee aina yksikäsitteliset alku- ja lopputila.

Esimerkiksi lausekkeesta \( r = (abab)\)\(^*\) saadaan näiden sääntöjen mukaan seuraava \( \varepsilon \)-automaatti:

![Diagram](image)

Automaatti on selvästi hyvin redundantti. Käsin automaatteja suunniteltaessa ne kannattaa usein muodostaa suoraan. Esim. lausekkeen \( r = (a(bab)\)\(^*\) perusteella on helppo muodostaa seuraava yksinkertainen epädeterministinen tunnistaja-automaatti:

![Diagram](image)

Lause 2.4 Jokainen äärellisellä automaattilla tunnistettava kieli on säännöllinen.

Todistus. Tarvitaan vielä yksi äärellisten automaattien laajennus: lausekeautomaattissa voidaan siirtymien ehtoina käyttää mielivaltaisia säännöllisiä lausekkeita.

Formalisointi: Merk. \( \text{RE}_\Sigma = \) aakkoston \( \Sigma \) säännöllisten lausekkeiden joukko. Lausekeautomaatti on viisikko

\[
M = (Q, \Sigma, \delta, q_0, F),
\]

missä siirtymäfunktio \( \delta \) on äärellinen kuvaus

\[
\delta : Q \times \text{RE}_\Sigma \rightarrow \mathcal{P}(Q)
\]

(so. \( \delta(q, r) \neq \emptyset \) vain äärellisen monella parilla \( (q, r) \in Q \times \text{RE}_\Sigma \)).
Yhden askeleen tilannejohto määritellään:

\[(q, w) \mapsto (q', w')\]

jos on \(q' \in \delta(q, r)\) jollakin sellaisella \(r \in \mathbb{RE}_\Sigma\), että \(w = zw'\), \(z \in L(r)\). Muut määritelmat saa mat kuin aiemmin.

Todistetaan: jokainen lausekeautomaatilla tunnistettava kieli on säännöllinen.

Olkoon \(M\) jokin lausekeautomaatti. Säännöllinen lauseke, joka kuvaa \(M\):n tunnistaman kielen, muodostetaan kahdessa vaiheessa:

1. Tiivistetään \(M\) ekvivalentiksi enintään 2-tilaiseksi lausekeautomaatiksi seuraavilla muunnoksilla:
   
   (i) Yhdistetään \(M\):n lopputiet yhdessä seuraavan kuvan mukaisesti:

   
   (ii) Poistetaan \(M\):n muut kuin alku- ja lopputiela yksi kerrallaan seuraavasti. Olk. \(q\) jokin \(M\):n tila, joka ei ole alku- eikä lopputila; tarkastellaan kaikkia "rettejä", jotka \(M\):ssä kulkevat \(q\):n kautta. Olk. \(q_i\) ja \(q_j\) \(q\):n välittömät edeltäjä ja seuraajatila jollakin tällaisella reitillä. Poistetaan \(q\) reitillä \(q_i \rightarrow q_j\) oheisen kuvan (i) muunnoksella, jos tilasta \(q\) ei ole siirtymää itseensä, ja kuvan (ii) muunnoksella, jos tilasta \(q\) on siirtymää itseensä:

   

   Samalla yhdistetään rinnakkaiset siirtymät seuraavasti:

2. Tiivistyksen päättymättä jäljellä olevaa enintään 2-tilaista automaattia vastaava säännöllinen lauseke muodostetaan seuraavan kuvan esittämällä tavalla:

\[\square\]
Esimerkki:

\[
\begin{align*}
A & \Rightarrow B \\
A \cup B & \Rightarrow (A \cup B)^2 \\
\end{align*}
\]