

2.6 SÄÄNNÖLLISET LAUSEKKEET

Automaattimalleista poikkeava tapa kuvata yksinkertaisia kieliä.

Olkoot A ja B aakkoston Σ kieliä. Perusoperaatioita:

- ▶ **Yhdiste:** $A \cup B = \{x \in \Sigma^* \mid x \in A \text{ tai } x \in B\}$;
- ▶ **Katenaatio:** $AB = \{xy \in \Sigma^* \mid x \in A, y \in B\}$;
- ▶ **Potenssit:**

$$\begin{cases} A^0 = \{\varepsilon\}, \\ A^k = AA^{k-1} = \{x_1 \dots x_k \mid x_i \in A \ \forall i = 1, \dots, k\} \quad (k \geq 1); \end{cases}$$

- ▶ **Sulkeuma t. "Kleenen tähti":**

$$\begin{aligned} A^* &= \bigcup_{k \geq 0} A^k \\ &= \{x_1 \dots x_k \mid k \geq 0, x_i \in A \ \forall i = 1, \dots, k\}. \end{aligned}$$

Määritelmä 2.3 Aakkoston Σ säännölliset lausekkeet määritellään induktiivisesti säännöillä:

1. \emptyset ja ε ovat Σ :n säännöllisiä lausekkeita;
2. a on Σ :n säännöllinen lauseke kaikilla $a \in \Sigma$;
3. jos r ja s ovat Σ :n säännöllisiä lausekkeita, niin $(r \cup s)$, (rs) ja r^* ovat Σ :n säännöllisiä lausekkeita;
4. muita Σ :n säännöllisiä lausekkeita ei ole.

Aakkoston $\{a, b\}$ säännöllisiä lausekkeita:

$$\begin{aligned} r_1 &= ((ab)b), \quad r_2 = (ab)^*, \\ r_3 &= (ab^*), \quad r_4 = (a(b \cup (bb)))^*. \end{aligned}$$

Lausekkeiden kuvaamat kielet:

$$\begin{aligned} L(r_1) &= (\{a\}\{b\})\{b\} = \{ab\}\{b\} = \{abb\}; \\ L(r_2) &= \{ab\}^* = \{\varepsilon, ab, abab, ababab, \dots\} \\ &= \{(ab)^i \mid i \geq 0\}; \\ L(r_3) &= \{a\}(\{b\})^* = \{a, ab, abb, abbb, \dots\} \\ &= \{ab^i \mid i \geq 0\}; \\ L(r_4) &= (\{a\}\{b, bb\})^* = \{ab, abb\}^* \\ &= \{\varepsilon, ab, abb, abab, ababb, \dots\} \\ &= \{x \in \{a, b\}^* \mid \text{kutakin } a\text{-kirjainta } x\text{:ssä seuraavaa } 1 \text{ tai } 2 \text{ } b\text{-kirjainta } \}. \end{aligned}$$

Sulkumerkkien vähentämissääntöjä:

Operaattoreiden prioriteetti:

$$* \quad \cdot \quad \cup \quad \cap$$

Yhdiste- ja tulo-operaattorien assosiativisuus:

$$\begin{aligned} L(((rs) \cup t)) &= L((r \cup (s \cup t))) \\ L(((rs) \cap t)) &= L((r \cap (s \cap t))) \end{aligned}$$

\Rightarrow peräkkäisiä yhdisteitä ja tuloja ei tarvitse suluttaa.

Käytetään tavallisia kirjaimia, mikäli sekaannuksen vaaraa merkijonoihin ei ole.

Yksinkertaisemmin siis:

$$r_1 = abb, \quad r_2 = (ab)^*, \quad r_3 = ab^*, \quad r_4 = (a(b \cup bb))^*.$$

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

Säännöllisten lausekkeiden sieventäminen

Säännöllisillä kielillä on yleensä useita vaihtoehtoisia kuvauksia, esim.:

$$\begin{aligned} \Sigma^* &= L((a \cup b)^*) \\ &= L((a^* b^*)^*) \\ &= L(a^* b^* \cup (a \cup b)^* b a (a \cup b)^*). \end{aligned}$$

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

Lausekkeen sieventäminen = "yksinkertaisimman" ekvivalentin lausekkeen määrittäminen.

Säännöllisten lausekkeiden ekvivalenssitestaus on epätriviaali, mutta periaatteessa mekaanisesti ratkeava ongelma.

Sievennyssääntöjä:

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

Mikä tahansa säännöllisten lausekkeiden tosi ekvivalenssi voidaan johtaa näistä laskulaeista, kun lisätään päätelysääntö: jos $r = rs \cup t$, niin $r = ts^*$, edellyttäen että $\varepsilon \notin L(s)$.

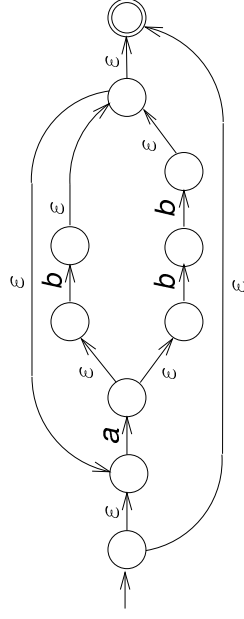
2.7 ÄÄRELLISET AUTOMAATIT JA SÄÄNNÖLLISET KIELET

Lause 2.3 Jokainen säännöllinen kieli voidaan tunnistaa äärellisellä automaatilla.

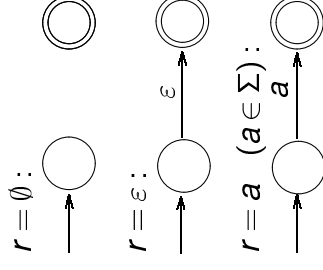
Todistus. Seuraavan kalvon induktiivisen konstruktion avulla voidaan mielivaltaisen säännöllisen lausekkeen r rakennetta seuraten muodostaa ε -automaatti M_r , jolla $L(M_r) = L(r)$. Tästä automaattista voidaan poistaa ε -siirtymät Lemman 2.4 mukaisesti, ja tarvittaessa voidaan syntyvä epädeterministinen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

Esitettävästä konstruktiosta on syytä huomata, että muodostettaviin ε -automaatteihin tulee aina yksikäsitteiset alku- ja lopputila, eikä minkään osa-automaatin lopputilasta lähde eikä alkutilaan tule yhtään ko. osa-automaatin sisäistä siirtymää. \square

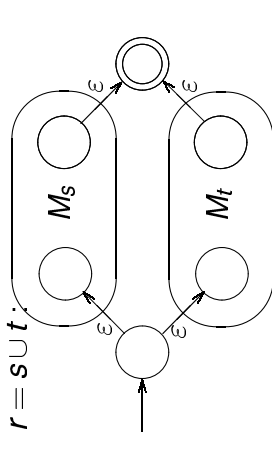
Esimerkiksi lausekkeesta $r = (a(b \cup bb))^*$ saadaan näiden sääntöjen mukaan seuraava ε -automaatti:



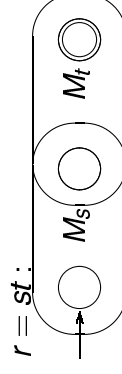
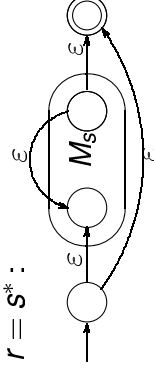
Automaatti on selvästi hyvin redundantti. Käsien automaatteja suunniteltaessa ne kannattaakin usein muodostaa suoraan.



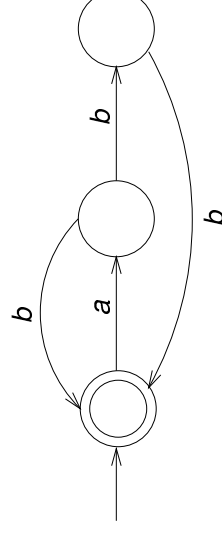
$$r = s \cup t:$$



$$r = s^*:$$



Esim. lausekkeen $r = (a(b \cup bb))^*$ perusteella on helppo muodostaa seuraava yksinkertainen epädeterministinen tunnistaja-automaatti:



Lause 2.4 Jokainen äärellisellä automaatilla tunnistettava kieli on säännöllinen.

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

Formalisointi: Merk. $RE_{\Sigma} =$ aakkoston Σ säännöllisten lausekkeiden joukko. *Lausekeautomaatti* on viisikko

$$M = (Q, \Sigma, \delta, q_0, F),$$

missä siirtymäfunktio δ on äärellinen kuvaus

$$\delta : Q \times RE_{\Sigma} \rightarrow \mathcal{P}(Q)$$

(so. $\delta(q, r) \neq \emptyset$ vain äärellisen monella parilla $(q, r) \in Q \times RE_{\Sigma}$).

Yhden askelen tilannejohto määritellään:

$$(q, w) \vdash_M (q', w')$$

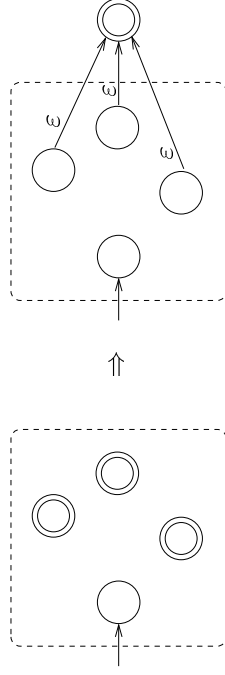
jos on $q' \in \delta(q, r)$ jollakin sellaisella $r \in RE_{\Sigma}$, että $w = zw'$, $z \in L(r)$. Muut määritelmät samat kuin aiemmin.

Todistetaan: jokainen lausekeautomaatilla tunnistettava kieli on säännöllinen.

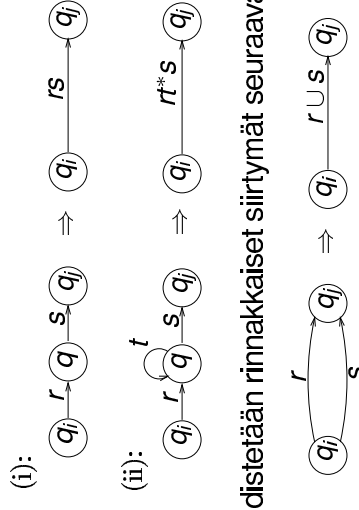
Olkoon M jokin lausekeautomaatti. Säännöllinen lauseke, joka kuvaa M :n tunnistaman kielen, muodostetaan kahdessa vaiheessa:

1. Tiivistetään M ekvivalentiksi enintään 2-tilaiseksi lausekeautomaatiksi seuraavilla muunnoksilla:

(i) Jos M :llä on useita lopputiloja, yhdistetään ne seuraavasti.



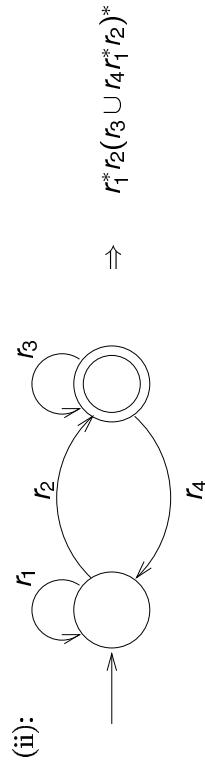
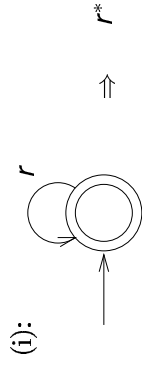
(ii) Poistetaan M :n muut kuin alkuperäiset lopputilat yksi kerrallaan seuraavasti. Olk. q jokin M :n tila, joka ei ole alkuperäinen lopputila; tarkastellaan kaikkia "reittejä", jotka M :ssä kulkevat q :n kautta. Olk. q_i ja q_j q :n välittömät edeltäjä- ja seuraajatilat jollakin tällaisella reitillä. Poistetaan q reitiltä $q_i \rightarrow q_j$ oheisen kuvan (i) muunnoksella, jos tilasta q ei ole siirtymää itseensä, ja kuvan (ii) muunnoksella, jos tilasta q on siirtymä itseensä:



Samalla yhdistetään rinnakkaiset siirtymät seuraavasti:



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



Esimerkki:

