Sissejuhatus teoreetilisse informaatikasse

Kevad 2019

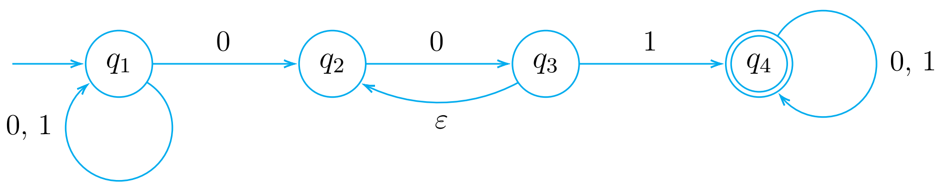
6. Mittedetermineeritud automaadid

Mittedetermineeritus

Eelmisel nädalal vaatlesime automaate, mille puhul järgmine olek oli üheselt määratud jooksva oleku ja sisendsümboliga. Niisuguseid automaate nimetatakse *determineeritud* automaatideks. Sellel nädalal vaatleme *mittedetermineeritud* automaate, mille puhul võib arvutuse igal sammul olla sõltuvalt jooksvast olekust ja sisendsümbolist järgmiseks olekuks mitu valikut.

Inglise keeles kasutatakse sageli lühendeid: DFA — determineeritud lõplik automaat, NFA — mittedetermineeritud lõplik automaat.

Näide 1. Järgmine mittedetermineeritud automaat aktsepteerib kõiki sõnesid, mis sisaldavad alamsõnet 001:



Sellel automaadil on tavalise determineeritud automaadiga võrreldes mõned tähelepandavad iseärasused.

- Olekus q_1 on kaks väljuvat kaart 0. See tähendab, et kui automaat on olekus q_1 ja kohtab sisendsõnes sümbolit 0, siis võib ta minna nii olekusse q_1 kui ka olekusse q_2.

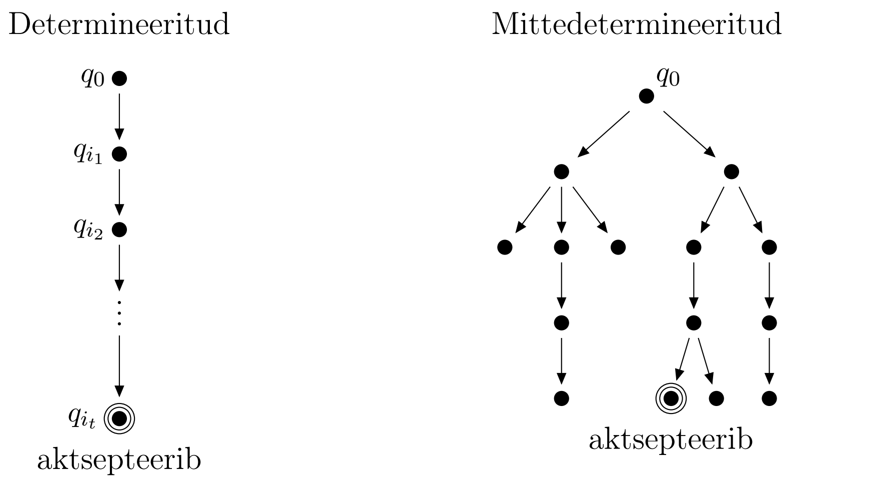
- Olekus q_2 pole väljuvat kaart 1. See tähendab, et kui automaat on olekus q_2 ja kohtab sümbolit 1, siis automaadi töö katkeb. Analoogiliselt, olekus q_3 pole väljuvat kaart 0.

Lisaks sisaldab automaat kaart märgendiga ε olekust q_3 olekusse q_2, mis tähendab, et automaat võib ilma järjekordset sisendsümbolit töötlemta lihtsalt minna olekust q_3 olekusse q_2.

Mittedetermineeritud lõplik automaat töötab järgmiselt.

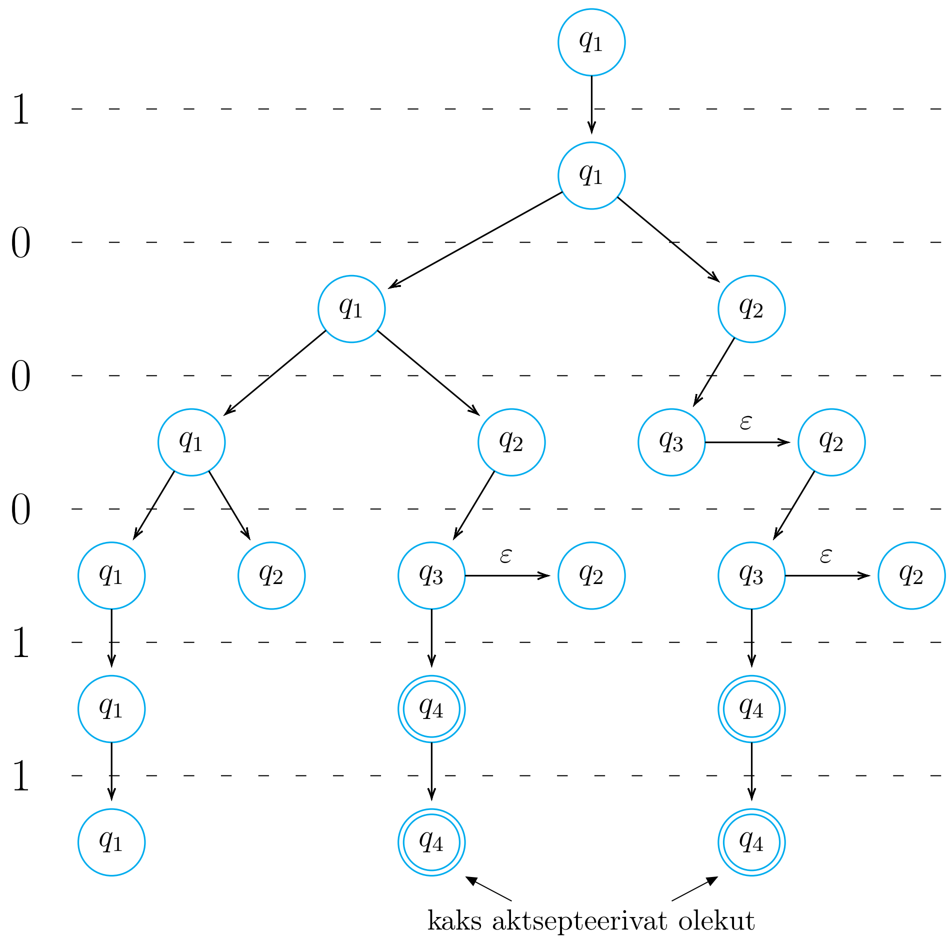
- Kui edasiminemiseks on mitu valikut, siis käivitab automaat paralleelselt mitu arvutuskäiku.
- Kui mingi sümboliga väljuv kaar puudub, siis katkestab automaat vastava arvutuskäigu.
- Kui olekust väljub kaari märgendiga ε, siis käivitab automaat paralleelselt mitu arvutuskäiku, mis algavad jooksvast olekust ja kõigist olekutest, kuhu kaared märgendiga ε suubuvad.
- Mittedetermineeritud lõplik automaat aktsepteerib sisendsõnet, kui vähemalt üks paralleelsetest arvutuskäikudest seda sisendsõnet aktsepteerib.

Kui võrrelda automaadi töötamist etteantud sisendsõnel, siis determineeritud automaadi puhul on olekute järgnevus lineaarne, kuid mittedetermineeritud automaadi puhul võivad paralleelsete arvutuskäikude tõttu tekkida hargnemised.



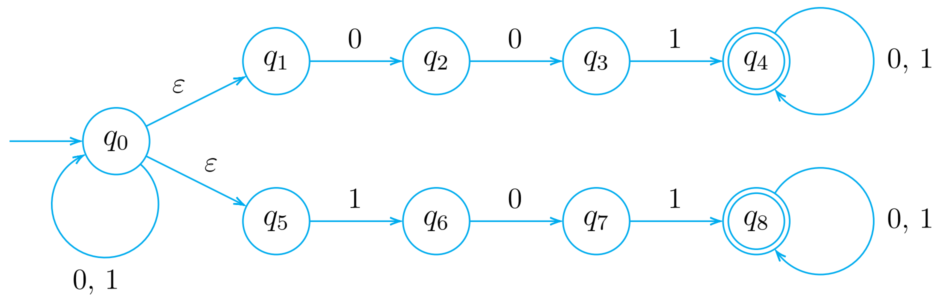
Automaat aktsepteerib sisendsõnet, sest leidub aktsepteeriv arvutustee.

Näide 2. Vaatleme uuesti näidet 1 ja anname sellele automaadile ette sisendsõne 100011. Järgmisel joonisel on kujutatud automaadi töötamisel tekkiv täielik hargnemiste puu.



Et leidub arvutustee, mis lõpeb aktsepteerivas olekus (isegi kaks sellist), siis automaat aktsepteerib seda sõnet.

Näide 3. Järgmine mittedetermineeritud lõplik automaat aktsepteerib kõiki sõnesid, mis sisaldavad alamsõnet 001 või 101.



Pärast nullidest ja ühtedest koosneva prefiksi lugemist olekus q_0 hüppab automaat kas olekusse q_1 või q_5 ning loeb vastava alamsõne. Kui sisendsõnes sellist alamsõnet ei leidu, siis sobivat hüppamiskohta pole ning automaat jääb töö lõpuni olekusse q_0.

Definitsioon. *Mittedetermineeritud lõplik automaat on viisik* $(Q, \Sigma, \delta, q_0, F)$, *kus*

- Q *on lõplik olekute hulk;*
- Σ *on lõplik tähestik;*
- $\delta \colon Q \times \Sigma_\varepsilon \to \mathcal{P}(Q)$ *on üleminekufunktsioon;*
- $q_0 \in Q$ *on algolek;*
- $F \subseteq Q$ *on aktsepteerivate olekute hulk.*

Võrreldes determineeritud lõpliku automaadiga erineb siin üleminekufunktsioon: sihthulk on $\mathcal{P}(Q)$ ehk hulga Q kõigi alamhulkade hulk ning lähtehulgas on teine komponent $\Sigma_\varepsilon = \Sigma \cup \{\varepsilon\}$ ehk lisatud on tühisümbol ε.

Näide 4. Vaatleme uuesti näite 1 automaati.

- Olekute hulk on $Q = \{q_1, q_2, q_3, q_4\}$.
- Tähestik on $\Sigma = \{0, 1\}$.
- Üleminekufunktsioon δ on antud tabeliga

	0	1	ε
q_1	$\{q_1, q_2\}$	$\{q_1\}$	$\emptyset$
q_2	$\{q_3\}$	$\emptyset$	$\emptyset$
q_3	$\emptyset$	$\{q_4\}$	$\{q_2\}$
q_4	$\{q_4\}$	$\{q_4\}$	$\emptyset$

- Algolek on q_1.
- Aktsepteerivate olekute hulk on $F = \{q_4\}$.

Definitsioon. *Olgu* $M = (Q, \Sigma, \delta, q_1, F)$ *mittedetermineeritud lõplik automaat ja* w *sõne tähestikus* Σ. *Ütleme, et automaat* M *aktsepteerib sõnet* w, *kui sõne saab esitada kujul* $w = a_1 a_2 \ldots a_m$, *kus* $a_i \in \Sigma_\varepsilon$, *ja leidub olekute järjend* $r_0, r_1, \ldots, r_m$, *et*

1. $r_0 = q_0$;
2. *iga* $i = 1, \ldots, m$ *puhul* $r_i \in \delta(r_{i-1}, a_i)$;
3. $r_m \in F$.

Paneme tähele, et sõne w esituses võib mõni a_i olla ε ehk tühisümbol.

Definitsioon. *Kaks automaati on ekvivalentsed, kui nad aktsepteerivad (tunnevad ära) sama keelt.*

Triviaalselt on iga determineeritud lõplik automaat ühtlasi ka mittedetermineeritud lõplik automaat. Pöördväite formuleerime teoreemina.

Teoreem. *Iga mittedetermineeritud lõpliku automaadi jaoks leidub temaga ekvivalentne determineeritud lõplik automaat.*

Tõestuse idee. Kui mittedetermineeritud automaadil on k olekut, siis vastaval determineeritud automaadil on 2^k olekut, milleks on mittedetermineeritud automaadi olekute hulga kõik alamhulgad.

Tõestus. Olgu $N = (Q, \Sigma, \delta, q_0, F)$ mittedetermineeritud lõplik automaat ja L keel, mida ta ära tunneb. Konstrueerime determineeritud lõpliku automaadi $M = (Q', \Sigma, \delta', q'_0, F')$, mis samuti tunneb ära keelt L.

1) Vaatleme esmalt lihtsat erijuhtu, kus N ei sisalda kaari märgendiga ε.

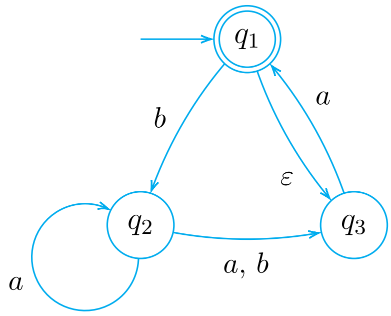
- Olekute hulk on $Q' = \mathcal{P}(Q)$, st hulga Q kõigi alamhulkade hulk.
- Iga $R \in Q'$ ja $a \in \Sigma$ korral olgu
$$\delta'(R, a) = \bigcup_{r \in R} \delta(r, a) = \{q \in Q \mid q \in \delta(r, a) \text{ mingi } r \in R \text{ puhul}\}. \quad (*)$$
  Automaadi M jaoks on R olek. Automaadi N jaoks on R olekute hulk. Kui automaat M loeb olekus R sümbolit a, siis võtame kokku kõik olekud, kuhu automaat N mõnest hulka R kuuluvast olekust sümboli a lugemisel võib minna, ja võtame saadud hulga funktsiooni δ' väärtuseks.
- Algolek on $q'_0 = \{q_0\}$.
- Aktsepteerivate olekute hulk on
$$F' = \{R \in Q' \mid R \text{ sisaldab automaadi } N \text{ aktsepteerivat olekut}\}.$$

2) Eeldame nüüd, et N võib sisaldada ka kaari märgendiga ε. Tähistame
$$E(R) = \{q \mid \text{mõnest hulga } R \text{ olekust pääseb mööda } \varepsilon\text{-kaari olekusse } q\}.$$
(Näites 1 esitatud automaadi puhul $E(\{q_1\}) = \{q_1\}$, $E(\{q_3\}) = \{q_2, q_3\}$). Asendame seose (*) seosega
$$\delta'(R, a) = \bigcup_{r \in R} E(\delta(r, a)) = \{q \in Q \mid q \in E(\delta(r, a)) \text{ mingi } r \in R \text{ puhul}\}$$
ja algoleku $\{q_0\}$ olekuga $E(\{q_0\})$. Sellega on automaat M konstrueeritud.

Automaat M jäljendab automaadi N tööd selles mõttes, et igal arvutussammul, mille automaat M sisendsõnel w sooritab, siirdub ta olekusse, mis võrdub nende olekute hulgaga, milles automaat N saab asuda sellel sammul. Järelikult kui automaat N jõuab sõne w lugedes mingil sammul aktsepteerivasse olekusse, siis jõuab samal sammul aktsepteerivasse olekusse ka automaat M; kui aga N aktsepteerivasse olekusse ei jõua, siis ei tee seda ka M. Seega aktsepteerivad automaadid M ja N sama hulka sõnesid. $\square$

Praktikumiülesanded

1. Antud on mittedetermineeritud lõplik automaat



st automaat $N = (\{q_1, q_2, q_3\}, \{a, b\}, \delta, q_1, \{q_1\})$ üleminekufunktsiooniga

	a	b	ε
q_1	$\emptyset$	$\{q_2\}$	$\{q_3\}$
q_2	$\{q_2, q_3\}$	$\{q_3\}$	$\emptyset$
q_3	$\{q_1\}$	$\emptyset$	$\emptyset$

Konstrueerida selle automaadiga ekvivalentne determineeritud lõplik automaat.

2. Tõestada, et kui L_1 ja L_2 on regulaarsed keeled, siis ka $L_1 \circ L_2$ (konkatenatsioon) on regulaarne keel.

3. Tõestada, et kui L_1 on regulaarne keel, siis ka L_1^* on regulaarne keel.

Lahendused

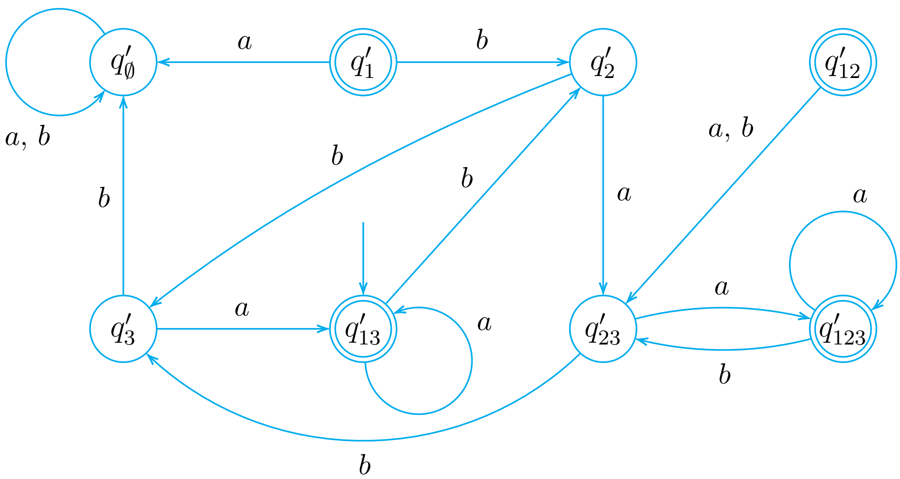
1. *Lahendus.* Konstrueerime automaadi $M = (Q', \{a,b\}, \delta', q'_0, F')$ järgmiselt.

- $Q' = \{q'_\emptyset, q'_1, q'_2, q'_3, q'_{12}, q'_{13}, q'_{23}, q'_{123}\}$
- Üleminekufunktsioon δ':

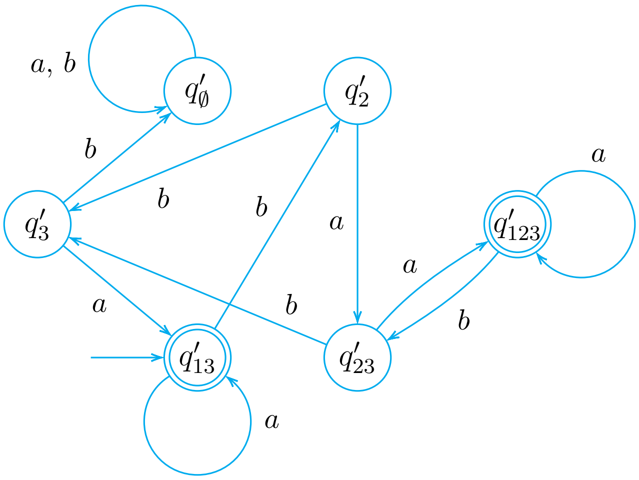
	a	b
$q'_\emptyset$	$q'_\emptyset$	$q'_\emptyset$
q'_1	$q'_\emptyset$	q'_2
q'_2	q'_{23}	q'_3
q'_3	q'_{13}	$q'_\emptyset$
q'_{12}	q'_{23}	q'_{23}
q'_{13}	q'_{13}	q'_2
q'_{23}	q'_{123}	q'_3
q'_{123}	q'_{123}	q'_{23}

- Algolek on q'_{13} (sest automaadi N algolekust q_1 pääseb olekusse q_3 mööda ε-kaart).
- $F = \{q'_1, q'_{12}, q'_{13}, q'_{123}\}$.

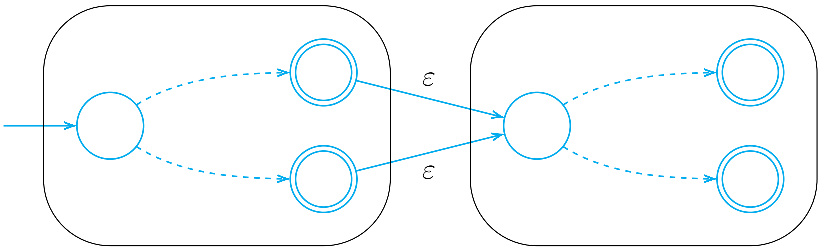
Niimoodi saame automaadi M:



Paneme tähele, et olekutesse q'_1 ja q'_{12} me kunagi jõuda ei saa. Seetõttu võib need olekud kustutada. Järele jääb järgmine determineeritud lõplik automaat:



2. *Lahenduse idee.* Olgu N_1 ja N_2 mittedetermineeritud lõplikud automaadid, mis vastavalt tunnevad ära keeli L_1 ja L_2. Konstrueerime automaadi N nii, et ühendame kõik N_1 aktsepteerivad olekud N_2 algolekusse ε-noolte abil:



Sisendsõne w töödeldes automaat N mittedetermineeritult „arvab ära“, kus keelde L_1 kuuluv osa lõpeb ja keelde L_2 kuuluv osa algab.

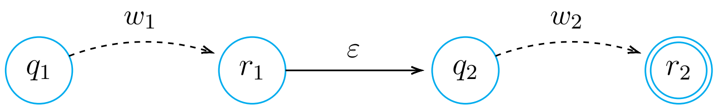
Lahendus. Olgu $N_1 = (Q_1, \Sigma, \delta_1, q_1, F_1)$ ja $N_2 = (Q_2, \Sigma, \delta_2, q_2, F_2)$ mittedetermineeritud lõplikud automaadid, mis vastavalt tunnevad ära keeli L_1 ja L_2. Konstrueerime mittedetermineeritud lõpliku automaadi $N = (Q, \Sigma, \delta, q_1, F_2)$, mis tunneb ära keelt $L_1 \circ L_2$.

- Olekute hulk on $Q = Q_1 \cup Q_2$
- Algolek on q_1, st sama mis automaadi N_1 algolek.
- Aktsepteerivate olekute hulk on F_2, st sama mis automaadi N_2 aktsepteerivate olekute hulk.

- Üleminekufunktsiooni δ määrame järgmiselt:

$$\delta(q,a) = \begin{cases} \delta_1(q,a) & \text{kui } q \in Q_1,\ q \notin F_1, \\ \delta_1(q,a) & \text{kui } q \in F_1,\ a \neq \varepsilon, \\ \delta_1(q,a) \cup \{q_2\} & \text{kui } q \in F_1,\ a = \varepsilon, \\ \delta_2(q,a) & \text{kui } q \in Q_2. \end{cases}$$

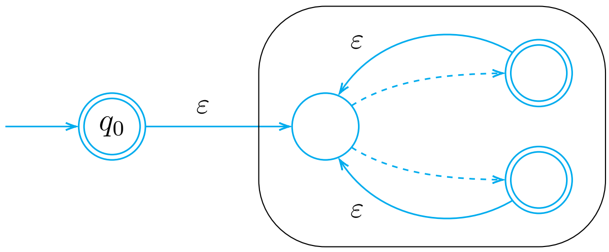
Automaat N aktsepteerib sisendsõnet w parajasti siis, kui leidub arvutustee, mis algab olekust q_1 ja lõpeb mingis hulga F_2 olekus. See tähendab, leidub arvutustee



$\in F_1$ $\in F_2$

kus $w = w_1 w_2$. Teiste sõnadega, leiduvad $w_1 \in L_1$ ja $w_2 \in L_2$, mille puhul $w = w_1 w_2$. See aga on samaväärne sellega, et $w \in L_1 \circ L_2$.

3. *Lahenduse idee.* Olgu N_1 keelt L_1 aktsepteeriv mittedetermineeritud automaat. Konstrueerime uue automaadi N nii, et tõmbame automaadi N_1 igast aktsepteerivast olekust algolekusse ε-kaare. Nii aktsepteerib uus automaat ükskõik mitme keele L_1 sõne järjestkirjutust. Et automaat töötaks õigesti ka tühisõne puhul, lisame kõige ette uue alg- ja ühtlasi aktsepteeriva oleku q_0:



N_1

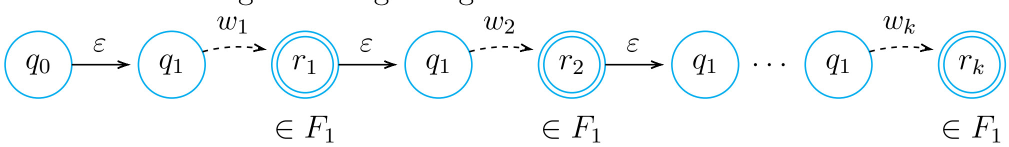
Lahendus. Olgu $N_1 = (Q_1, \Sigma, \delta_1, q_1, F_1)$ mittedetermineeritud lõplik automaat, mis tunneb ära keelt L_1. Konstrueerime mittedetermineeritud lõpliku automaadi $N = (Q, \Sigma, \delta, q_0, F)$, mis tunneb ära keelt L_1^*.

- Olekute hulk on $Q = Q_1 \cup \{q_0\}$, st senistele olekutele lisame oleku q_0.
- Uus algolek on q_0.
- Aktsepteerivate olekute hulk on $F = F_1 \cup \{q_0\}$.

- Üleminekufunktsioon on

$$\delta(q,a) = \begin{cases} \delta_1(q,a) & \text{kui } q \in Q_1,\ q \notin F_1, \\ \delta_1(q,a) & \text{kui } q \in F_1,\ a \neq \varepsilon, \\ \delta_1(q,a) \cup \{q_1\} & \text{kui } q \in F_1,\ a = \varepsilon, \\ \{q_1\} & \text{kui } q = q_0,\ a = \varepsilon, \\ \emptyset & \text{kui } q = q_0,\ a \neq \varepsilon. \end{cases}$$

Automaat N aktsepteerib sisendsõnet w parajasti siis, kui $w = \varepsilon$ või kui automaadi töö kulgeb nii nagu diagrammil



kus w_1, w_2, $\ldots$, w_k on mingid sõned, mille puhul $w = w_1 w_2 \ldots w_k$. See on samaväärne sellega, et $w = \varepsilon$ või leiduvad sõned $w_1 \in L_1$, $w_2 \in L_1$, $\ldots$, $w_k \in L_1$, et $w = w_1 w_2 \ldots w_k$. See aga omakorda tähendab, et $w \in L_1^*$.