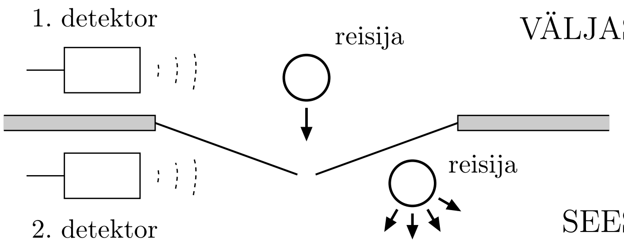
Sissejuhatus teoreetilisse informaatikasse

Kevad 2019

5. Automaaditeooria

Automaaditeooria

Näide 1. Lennujaama sissepääsu juhtimissüsteem koosneb kahest detektorist, millest üks asub väljas ja teine sees. Kumbki detektor suudab tuvastada läheduses olevat inimest. Süsteem peab detektoritelt saadud andmete põhjal uksed avama või sulgema.

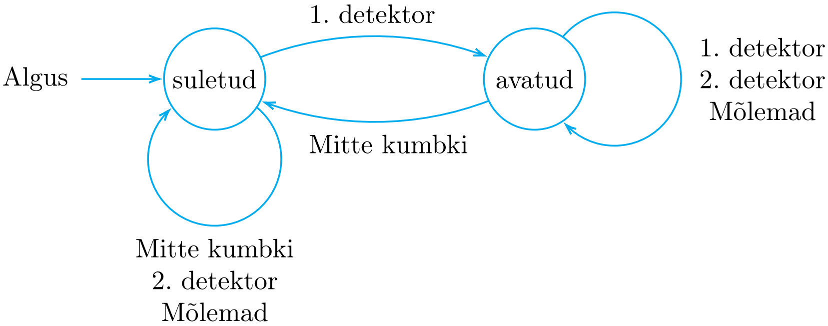


Uste uus olek sõltub kahest asjaolust: milline on uste senine olek ja millised detektorid on aktiveeritud. Järgmises tabelis on esitatud uste uus olek kõigi variantide korral.

Senine olek	Kumbki pole aktiveeritud	1. detektor aktiveeritud	2. detektor aktiveeritud	Mõlemad aktiveeritud
suletud	suletud	avatud	suletud	suletud
avatud	suletud	avatud	avatud	avatud

Kui inimesi läheduses pole (kumbki detektor pole aktiveeritud), siis tuleb uksed sulgeda. Kui 1. detektori vaatevälja ilmub inimene, siis tuleb talle uks avada. Kui inimene asub 2. detektori vaateväljas (sõltumata sellest, kas 1. detektori vaateväljas on veel inimesi või mitte), siis ei tohi uste seisu muuta, sest see võib ohustada sisenenud ja uste lähedal asuvaid inimesi.

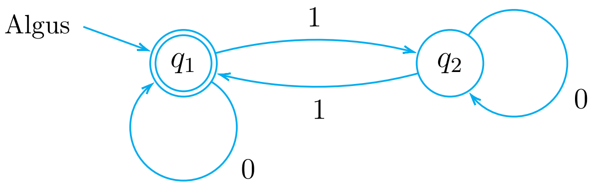
Tabelis kirjeldatud uste seisu muutmise reeglid võib ülevaatlikult kokku võtta järgmise diagrammiga.



Süsteem liigub olekute „suletud“ ja „avatud“ vahel sõltuvalt detektorite aktiveerumisest. Näiteks kui esialgu on süsteem olekus „suletud“ ja 1. (kuid mitte 2.) detektor aktiveerub, siis läheb süsteem üle olekusse „avatud“.

Tavaliselt loeb automaat sisendsümbolite järjendit ehk sisendsõnet.

Näide 2. Järgmisel diagrammil on kujutatud automaat, mis teeb kindlaks, kas sisendiks antud kahendvektor sisaldab paarisarvu ühtesid.



Näiteks anname sellele automaadile ette kahendsõne 010010. Automaat loeb seda sõnet sümbolhaaval vasakult paremale ning igal sammul läheb sõltuvalt oma senisest olekust ja järjekordsest sümbolist uude olekusse.

Samm	Vaatleb sümbolit	Uus olek
	Algus	q_1
1.	0	q_1
2.	1	q_2
3.	0	q_2
4.	0	q_2
5.	1	q_1
6.	0	q_1
	Lõpp	

Diagrammi, mis kirjeldab automaadi olekuid ja nendevahelisi üleminekuid, nimetatakse *olekudiagrammiks*. Olekut, mida diagrammil tähistab kahekordne ring, nimetatakse *aktsepteerivaks olekuks*. Kui automaat lõpetab töö selles olekus, siis võtab ta sisendsõne vastu ehk *aktsepteerib* seda, mis tähendab, et sisendsõnel on kontrollitav omadus. Näiteks eeltoodud automaat aktsepteerib sõnet 010010, nagu peabki, sest selles sõnes on tõepoolest paarisarv ühtesid.

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

- Q *on lõplik olekute hulk;*
- Σ *on lõplik sümbolite hulk (tähestik);*
- $\delta\colon Q \times \Sigma \to Q$ *on üleminekufunktsioon;*
- $q_0 \in Q$ *on algolek;*
- $F \subseteq Q$ *on aktsepteerivate olekute hulk.*

Näide 3. Vaatleme näites 2 esitatud automaati.

- Olekute hulk on $Q = \{q_1, q_2\}$. Automaat on igal hetkel täpselt ühes selle hulga olekus. Töö käigus võib ta siirduda ühest olekust teise.
- Tähestik on $\Sigma = \{0, 1\}$. Sõned, mis automaadile ette antakse, peavad koosnema ainult selle tähestiku sümbolitest.
- Algolek on $q_0 = q_1$. Sellest olekust algab automaadi liikumine olekute vahel, kui ta etteantud sõnet loeb.
- Aktsepteerivate olekute hulk on $F = \{q_1\}$. Kui sõne töötlemine lõpeb sellesse hulka kuuluvas olekus, siis automaat aktsepteerib seda sõnet.
- Üleminekufunktsioon δ on määratud vastavustega

  $$\delta(q_1, 0) = q_1, \quad \delta(q_1, 1) = q_2, \quad \delta(q_2, 0) = q_2, \quad \delta(q_2, 1) = q_1$$

  ehk tabeli kujul

	0	1
q_1	q_1	q_2
q_2	q_2	q_1

  Üleminekufunktsioon näitab, millisesse olekusse automaat läheb, sõltuvalt senisest olekust ja sisendsümbolist.

Definitsioon. *Olgu* $M = (Q, \Sigma, \delta, q_0, F)$ *lõplik automaat ja* $w = a_1 a_2 \ldots a_n$ *sõne, kus iga* $a_i \in \Sigma$*. Ütleme, et automaat* M *aktsepteerib sõnet* w*, kui leidub selline olekute järjend* $r_0, r_1, \ldots, r_n$*, et*

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

Sõnede hulka nimetatakse *keeleks.*

Definitsioon. *Olgu M lõplik automaat ja L keel. Ütleme, et L on automaadi M keel, kui L on parajasti kõigi nende sõnede hulk, mida automaat M aktsepteerib. Sel juhul ütleme ka, et automaat M aktsepteerib (ehk tunneb ära) keelt L.*

Seega automaat M aktsepteerib keelt L parajasti siis, kui

$$L = \{w \mid M \text{ aktsepteerib } w\}.$$

Näide 4. Olgu M näite 2 automaat. Olgu L kõigi kahendsõnede hulk, milles on paarisarv ühtesid, st $L = \{\varepsilon, 0, 00, 11, \ldots, 0101, \ldots, 011110, \ldots\}$, kus ε tähistab tühisõnet. Tõestame, et automaat M aktsepteerib keelt L.

Automaadi üleminekufunktsioonist näeme, et automaadi olek muutub, kui sisendsümbol on 1, ja ei muutu, kui sisendsümbol on 0. Seega kui automaadile etteantavas sõnes on paarisarv ühtesid, siis sooritab automaat paarisarvu üleminekuid olekust q_1 olekusse q_2 või tagasi ja lõpetab aktsepteerivas olekus q_1. Kui aga sõnes on paaritu arv ühtesid, siis sooritab automaat paaritu arvu selliseid üleminekuid ja lõpetab olekus q_2. Järelikult aktsepteerib automaat kõiki sõnesid, milles on paarisarv ühtesid, ja ainult neid. Seega antud keel on tõesti selle automaadi keel.

Definitsioon. *Keelt L nimetatakse regulaarseks, kui leidub lõplik automaat, mis keelt L aktsepteerib.*

Definitsioon. *Olgu L_1 ja L_2 kaks keelt. Defineerime keelte ühendi ja konkatenatsiooni ning tärnoperatsiooni järgmiselt.*

- *Ühend:*
$$L_1 \cup L_2 = \{w \mid w \in L_1 \textit{ või } w \in L_2\}$$
- *Konkatenatsioon:*
$$L_1 \circ L_2 = \{w_1 w_2 \mid w_1 \in L_1 \textit{ ja } w_2 \in L_2\}$$
- *Tärnoperatsioon:*
$$L_1^* = \{w_1 w_2 \ldots w_k \mid k \geq 0 \textit{ ja iga } i = 1, \ldots, k \textit{ puhul } w_i \in L_1\}$$

Näide 5. Olgu $L_1 = \{00, 010\}$, $L_2 = \{11\}$. Siis

$$\begin{aligned} L_1 \cup L_2 &= \{00, 010, 11\} \\ L_1 \circ L_2 &= \{0011, 01011\} \\ L_1^* &= \{\varepsilon, 00, 010, 0000, 00010, 01000, \ldots\} \end{aligned}$$

Teoreem. *Kui L_1 ja L_2 on regulaarsed keeled, siis ka $L_1 \cup L_2$ on regulaarne keel.*

Tõestus. Et L_1 ja L_2 on regulaarsed keeled, siis leiduvad automaadid, mis neid keeli aktsepteerivad. Tähistame neid automaate vastavalt tähistega M_1 ja M_2. Konstrueerime automaadi M, mis aktsepteerib keelt $L_1 \cup L_2$.

Automaadi M saame järgmise ideega: M töötab nii, nagu automaadid M_1 ja M_2 töötaksid paralleelselt. Automaadi M olekuteks võtame paarid (M_1 olek, M_2 olek). Kui emb-kumb automaat M_1 või M_2 lõpetab sisendsõnel töö aktsepeerivas olekus, siis lõpetab ka M töö aktsepteerivas olekus.

Olgu niisiis $M_1 = (Q_1, \Sigma, \delta_1, q_1, F_1)$ ja $M_2 = (Q_2, \Sigma, \delta_2, q_2, F_2)$. Defineerime automaadi $M = (Q, \Sigma, \delta, q_0, F)$ järgmiselt.

- Olekute hulk on
$$Q = \{(r_1, r_2) \mid r_1 \in Q_1, r_2 \in Q_2\}$$
ehk kõigi paaride hulk, kus esimene komponent on M_1 olek ja teine komponent M_2 olek.

- Tähestik Σ on sama nagu lähteautomaatidel.

- Üleminekufunktsiooni δ määrame seosega
$$\delta((r_1, r_2), a) = (\delta_1(r_1, a), \delta_2(r_2, a))$$
ehk tulemuseks on paar, mille komponendid on vastavalt M_1 ja M_2 uued olekud pärast ühte sammu sümbolil a.

- Algolek on $q_0 = (q_1, q_2)$.

- Aktsepteerivate olekute hulk on
$$F = \{(r_1, r_2) \mid r_1 \in F_1 \text{ või } r_2 \in F_2\}.$$
ehk paaride hulk, kus esimene või teine komponent on vastava automaadi aktsepteeriv olek.

Kui $w \in L_1 \cup L_2$, siis lõpetab M_1 või M_2 sõnel w töö aktsepteerivas olekus. Nende automaatide lõppolekutest moodustatud paar on definitsiooni kohaselt aktsepteeriv olek ka automaadi M jaoks. Kui aga $w \notin L_1 \cup L_2$, siis M_1 ja M_2 lõppolekud pole kumbki aktsepteerivad, mistõttu nendest moodustatud paar pole aktsepteeriv ka automaadi M jaoks. Seega automaat M aktsepteerib parajasti keelde $L_1 \cup L_2$ kuuluvaid sõnesid. $\square$

Praktikumiülesanded

1. Vaatleme automaati $M = (Q, \Sigma, \delta, q_0, F)$, kus $Q = \{q_1, q_2\}$, $\Sigma = \{0, 1\}$, $q_0 = q_1$, $F = \{q_2\}$ ja δ on antud tabeliga

	0	1
q_1	q_1	q_2
q_2	q_1	q_2

Millist keelt see automaat ära tunneb? (See automaat sarnaneb loengu automaadiga, kuid δ on erinev.)

2. Mis keelt tunneb ära automaat $M = (Q, \Sigma, \delta, q_0, F)$, kus $Q = \{q_0, q_1, q_2\}$, $\Sigma = \{0, 1, 2, \#\}$, algolek on q_0, $F = \{q_0\}$ ja δ on määratud tabeliga

	0	1	2	#
q_0	q_0	q_1	q_2	q_0
q_1	q_1	q_2	q_0	q_0
q_2	q_2	q_0	q_1	q_0

3. Koostada lõplik automaat M_0, mille äratuntav keel koosneb kõigist kahendsõnedest, mis sisaldavad alamsõnena sõnet 001.

Näiteks aktsepteeriks see automaat sõnesid 001, 000011, 1111001111, kuid mitte sõnet 101.

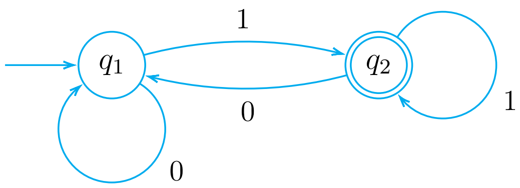
4. Olgu $\Sigma = \{0, 1\}$ ja L regulaarne keel. Tõestada, et keel

$$L' = \{01w \mid w \in L\}$$

on samuti regulaarne.

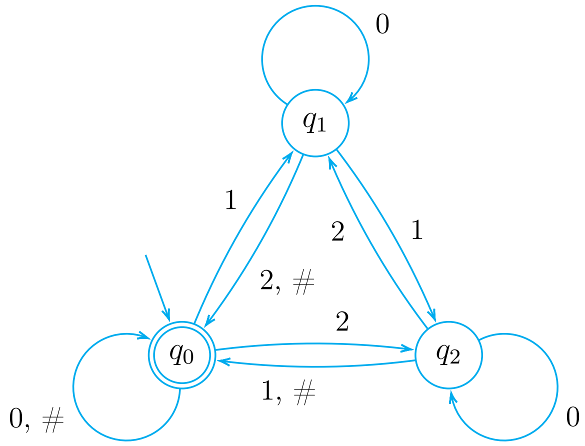
Lahendused

1. *Lahendus.* Selle automaadi olekudiagramm on järgmine:



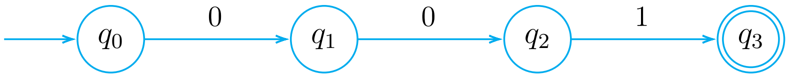
Paneme tähele, et pärast sümboli 1 lugemist on automaat alati aktsepteerivas olekus ja pärast sümboli 0 lugemist ta ei ole kunagi aktsepteerivas olekus. Seega aktsepteerib see automaat parajasti kõiki selliseid sõnesid, mis lõpevad sümboliga 1. Automaadi M äratuntav keel on niisiis kõigi 1-ga lõppevate kahendsõnede hulk.

2. *Lahendus.* Automaadi olekudiagramm on



Automaat M peab olekute abil meeles, mis on sisendsümbolite summa mooduli 3 järgi. Iga kord, kui ta kohtab sümbolit #, lähtestab ta loenduri uuesti nulliks. Seega aktsepteerib automaat M kõiki selliseid sõnesid, milles viimasele #-le järgnevate numbrite summa jagub 3-ga.

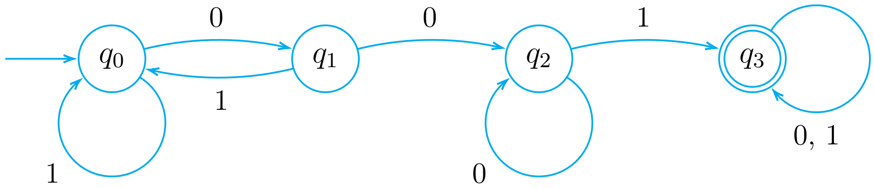
3. *Lahendus.* Lähtume osalisest olekudiagrammist



Niisuguse olekudiagrammiga automaat aktsepteerib sisendit 001. Kuid me peame ette nägema üleminekud ka juhuks, kui automaadile antakse ette teistsugune sisendsõne.

- Kui M_0 on olekus q_0 või q_1 ja kohtab sisendis sümbolit 1, siis peab ta minema tagasi olekusse q_0.
- Kui M_0 on olekus q_2 ja kohtab sümbolit 0, siis on ta selleks hetkeks lugenud sõne 000. Seetõttu peab ta jääma olekusse q_2.
- Kui M_0 on olekus q_3, siis peab ta jääma sellesse olekusse sõltumata järgnevatest sisendsümbolitest.

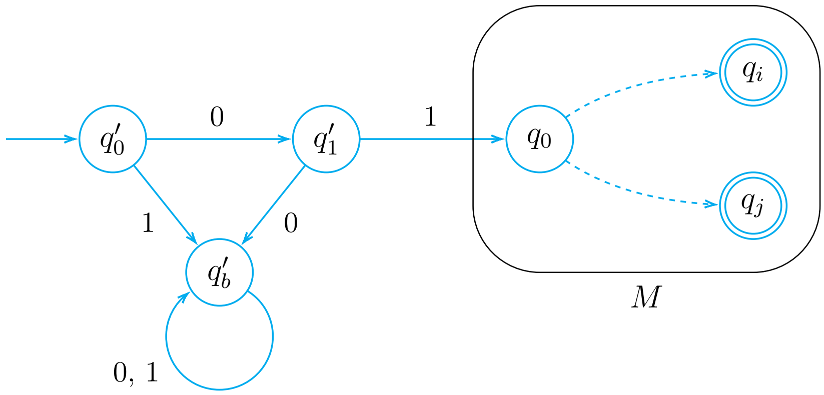
Seega saame järgmise diagrammi:



Algebralisel kujul kirja pannes saame siit järgmiste parameetritega automaadi: $Q = \{q_0, q_1, q_2, q_3\}$, $\Sigma = \{0, 1\}$, algolek on q_0, $F = \{q_3\}$ ja δ on

	0	1
q_0	q_1	q_0
q_1	q_2	q_0
q_2	q_2	q_3
q_3	q_3	q_3

4. *Lahenduse idee.* Et L on regulaarne keel, siis leidub automaat M, mis keelt L aktsepteerib. Meil on vaja konstrueerida automaat, mis aktsepteerib keelt L'. Võtame automaadi M olekudiagrammi ja lisame selle ette osa, mis töötleb sisendsõne prefiksit:



Kui sisendi kaks esimest sümbolit on 01, siis antakse järg üle automaadile M, mis töötleb sõne ülejäänud osa. Kui sisendsõne kaks esimest sümbolit on 00 või esimene sümbol on 1, siis läheb automaat mitteaktsepteerivasse olekusse.

Lahendus. Olgu $M = (Q, \Sigma, \delta, q_0, F)$ automaat, mis aktsepteerib keelt L. Defineerime automaadi $M' = (Q', \Sigma, \delta', q_0', F')$ järgmiselt: $Q' = Q \cup \{q_0', q_1', q_b'\}$, q_0' on uus algolek, $F' = F$, Σ on sama tähestik ning üleminekufunktsioon δ' uute olekute jaoks: $\delta'(q_0', 0) = q_1'$, $\delta'(q_0', 1) = q_b'$, $\delta'(q_1', 0) = q_b'$, $\delta'(q_1', 1) = q_0$, $\delta'(q_b', 0) = \delta'(q_b', 1) = q_b'$ ja vanade olekute jaoks: iga $q \in Q$ ja $a \in \Sigma$ korral $\delta'(q, a) = \delta(q, a)$.

Tõestame, et automaadi M' aktsepteeritavate sõnade hulk on L'.

1) Eeldame, et $w' \in L'$. Siis $w' = 01w$, kus $w \in L$. Automaadi M' konstruktsiooni tõttu on automaat M' pärast sõnest w' sümbolite 01 lugemist olekus q_0. Pärast sisendi ülejäänud osa w lugemist jõuab ta aktsepteerivasse olekusse, sest M aktsepteerib sõnet w. Järelikult M' aktsepteerib sõnet w'.

2) Eeldame, et M' aktsepteerib sõnet w'. Automaadi M' konstruktsiooni tõttu peab ta selleks läbima oleku q_0, sest see on ainuke võimalus pääseda

algolekust q_0' aktsepteerivate olekute hulka F. Üleminekufunktsiooni δ' definitsioonist näeme, et ainuke võimalus pääseda algolekust q_0' olekusse q_0 on lugeda sisendist 01. Seega $w' = 01w$. Kuna M' aktsepteerib sõnet w', siis jõuab M sõnet w lugedes olekust q_0 aktsepteerivasse olekusse. Järelikult M aktsepteerib sõnet w. See tähendab, et $w \in L$. Seega $w' \in L'$.

Harjutusülesanded

5. Mis keelt aktsepteerib automaat $M = (Q, \Sigma, \delta, q_0, F)$, kus $Q = \{q_1, q_2\}$, $\Sigma = \{0, 1\}$, $q_0 = q_1$, $F = \{q_2\}$ ning δ on

	0	1
q_1	q_1	q_2
q_2	q_1	q_1

6. Mis keelt aktsepteerib automaat $M = (Q, \Sigma, \delta, q_0, F)$, kus $Q = \{q_1, q_2, q_3\}$, $\Sigma = \{0, 1\}$, $q_0 = q_1$, $F = \{q_3\}$ ning δ on

	0	1
q_1	q_2	q_1
q_2	q_3	q_1
q_3	q_3	q_1

7. Koostada lõplik automaat, mille äratuntav keel koosneb parajasti kõigist kahendsõnedest, milles esineb ainult üks 1.

8. Koostada lõplik automaat, mille äratuntav keel koosneb parajasti kõigist kahendsõnedest, milles nulle on paaris- ja ühtesid paaritu arv.

9. Koostada lõplik automaat, mille äratuntav keel koosneb parajasti kõigist kahendarvudest, mis jaguvad 5-ga. Näiteks 0, 101, 1010, 1111, 11001 kuuluvad keelde, aga 1, 10, 11, 100, 110, 1110 mitte.

10. Leida kõik keelte paarid L_1, L_2, mille puhul

$$L_1 \circ L_2 = \{10, 11, 1000, 1010, 10111, 101000\}.$$

Juhised. 5., 6. Praktikumiülesanded 1 ja 2. 7., 8. Praktikumiülesanne 3. 9. Vaadata 5 olekuga automaati ning liikuda nende vahel vastavalt seni loetud arvu jäägile 5-ga jagamisel. 10. On rohkem kui üks lahend.