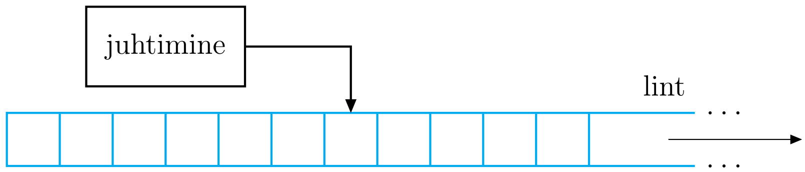
# Sissejuhatus teoreetilisse informaatikasse

Kevad 2019

## 10. Turingi masinad

Lõplikud automaadid on üks võimalik teoreetiline arvutimudel, kuid võrdlemisi piiratud, sest lõplik automaat suudab meeles pidada ainult konstantset hulka informatsiooni. Seetõttu oleme huvitatud niisugustest arvutimudelitest, mis rohkem sarnanevad kaasaja arvutitega. Lisame lõplikule automaadile juurde „lõpmatu mälu“, mida esitab lõpmatu lahtriteks jaotatud lint:




Sellist arvutimudelit nimetatakse Turingi masinaks.

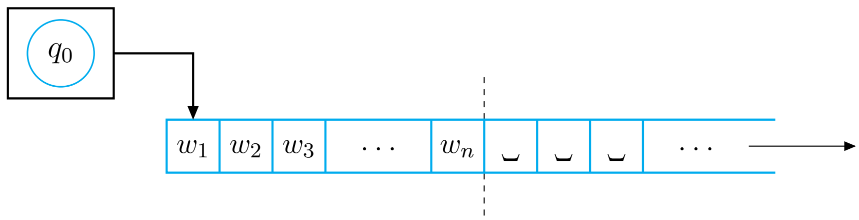
Lint on paremalt lõpmatu ja iga lahter võib sisaldada ühte sümbolit. Turingi masin saab kirjutada lindile informatsiooni ja seda sealt lugeda. Lugevkirjutav pea võib liikuda lahtrikaupa mööda linti vasakule ja paremale. Masinal on kaks spetsiaalset olekut: aktsepteeriv ja tagasilükkav. Kui masin jõuab emba-kumba olekusse, siis ta lõpetab töö vastavalt kas aktsepteerimise või tagasilükkamisega.

**Definitsioon.** *Turingi masin on seitsmik* $(Q, \Sigma, \Gamma, \delta, q_0, q_{akts}, q_{tag})$, *kus*

- $Q$ *on lõplik olekute hulk;*
- $\Sigma$ *on lõplik sisendtähestik, ei sisalda tühikut;*
- $\Gamma \supseteq \Sigma$ *on lõplik linditähestik, sisaldab tühikut;*
- $\delta\colon Q \times \Gamma \to Q \times \Gamma \times \{\mathsf{L}, \mathsf{R}\}$ *on üleminekufunktsioon;*
- $q_0 \in Q$ *on algolek;*
- $q_{akts} \in Q$ *on aktsepteeriv olek;*
- $q_{tag} \in Q$ *on tagasilükkav olek,* $q_{tag} \neq q_{akts}$.

Sisend antakse Turingi masinale ette sõnena lindil, kus sõne sümbolid on kirjutatud järjestikustesse lahtritesse. Sisendi lõppu tähistab tühik. Üleminekufunktsioon $\delta$ näitab sõltuvalt masina senisest olekust ja lindil pea all olevast sümbolist: 1) millisesse järgmisse olekusse masin läheb, 2) millise sümboli pea all olevasse lahtrisse kirjutab ja 3) kummale poole, vasakule või paremale, seejärel pea sammu teeb. Kui masin püüab liigutada pead lindi kõige vasakpoolsemast lahtrist vasakule, siis jääb pea samasse kohta.

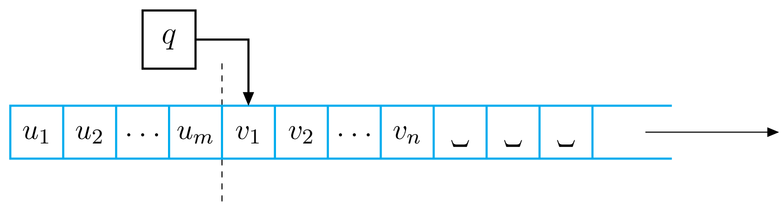
Turingi masin alustab tööd järgmises *algkonfiguratsioonis*:




See tähendab,

- masin on olekus $q_0$;
- lindi lahtritesse alates kõige vasakpoolsemast on kirjutatud sisendsõne ning kõik järgmised lahtrid kuni lindi lõpuni on täidetud tühikutega;
- pea asub kõige vasakpoolsema lahtri kohal.

Üldiselt, igal arvutussammul on masina seis määratud *masina konfiguratsiooniga*, mille moodustavad üheskoos jooksev olek, lindi sisu ja pea asukoht. Konfiguratsiooni märgime tähisega $uqv$, kus $q$ on olek, $u$ peast vasakule jääv sõne ja $v$ peast paremale jääv sõne, kusjuures pea asub sõne $v$ esimese sümboli kohal. Näiteks konfiguratsiooni




tähis on $u_1u_2 \dots u_m q v_1 v_2 \dots v_n$.

Ütleme, et konfiguratsioonist $C_1$ *tuleneb* konfiguratsioon $C_2$, kui Turingi masin saab reegleid järgides minna ühe sammuga konfiguratsioonist $C_1$ konfiguratsiooni $C_2$. Formaalselt, kui masin teeb sammu vasakule, siis konfiguratsioonist $uaq_ibv$ tuleneb konfiguratsioon $uq_jacv$ parajasti siis, kui

$$\delta(q_i, b) = (q_j, c, \mathsf{L});$$

kui aga masin teeb sammu paremale, siis konfiguratsioonist $uaq_ibv$ tuleneb konfiguratsioon $uacq_jv$ parajasti siis, kui

$$\delta(q_i, b) = (q_j, c, \mathsf{R}).$$

*Aktsepteerivaks konfiguratsiooniks* nimetame sellist konfiguratsiooni, milles olek on $q_{\text{akts}}$ ning *tagasilükkavaks konfiguratsiooniks* sellist konfiguratsiooni, milles olek on $q_{\text{tag}}$. Aktsepteeriv konfiguratsioon ja tagasilükkav konfiguratsioon on peatuvad konfiguratsioonid, nendesse jõudes masina töö peatub.

Sõnede hulka, mida Turingi masin $M$ aktsepteerib, nimetatakse *masina $M$ keeleks* ehk *masina $M$ poolt äratuntavaks keeleks*. Need on parajasti kõik sõned, mille puhul masin peatub olekus $q_{\text{akts}}$. Ülejäänud sõnede puhul võib masin peatuda olekus $q_{\text{tag}}$ või jääda lõpmatuseni tööle.

**Definitsioon.** *Keelt nimetatakse Turingi mõttes äratuntavaks, kui leidub Turingi masin, mis seda keelt ära tunneb.*

Turingi masin *lahendab* keelt, kui ta alati peatub olekus $q_{\text{akts}}$ või $q_{\text{tag}}$ ja tunneb ära seda keelt. See tähendab, iga keelde kuuluva sõne puhul lõpetab masin töö olekus $q_{\text{akts}}$ ja iga keelde mittekuuluva sõne puhul olekus $q_{\text{tag}}$.

**Definitsioon.** *Keelt nimetatakse Turingi mõttes lahenduvaks, kui leidub Turingi masin, mis seda keelt lahendab.*

Näide 1. Olgu

$$L = \{w\#w \mid w \in \Sigma^*,\ \Sigma = 0, 1\}.$$

Kirjeldame Turingi masinat $M$, mis tunneb ära keelt $L$.

Ettantud sisendsõne $s$ puhul tegutseb masin $M$ järgmiselt.

1. Liigub edasi-tagasi mööda linti ja kontrollib, kas vastavatel positsioonidel kummalgi pool sümbolit # asuvad alati samad sümbolid. Kui ei asu või kui masin sümbolit # sisendist ei leiagi, siis läheb tagasilükkavasse olekusse. Märgistab kõik kontrollitud sümbolid, et pidada järge vastavate sümbolite üle.

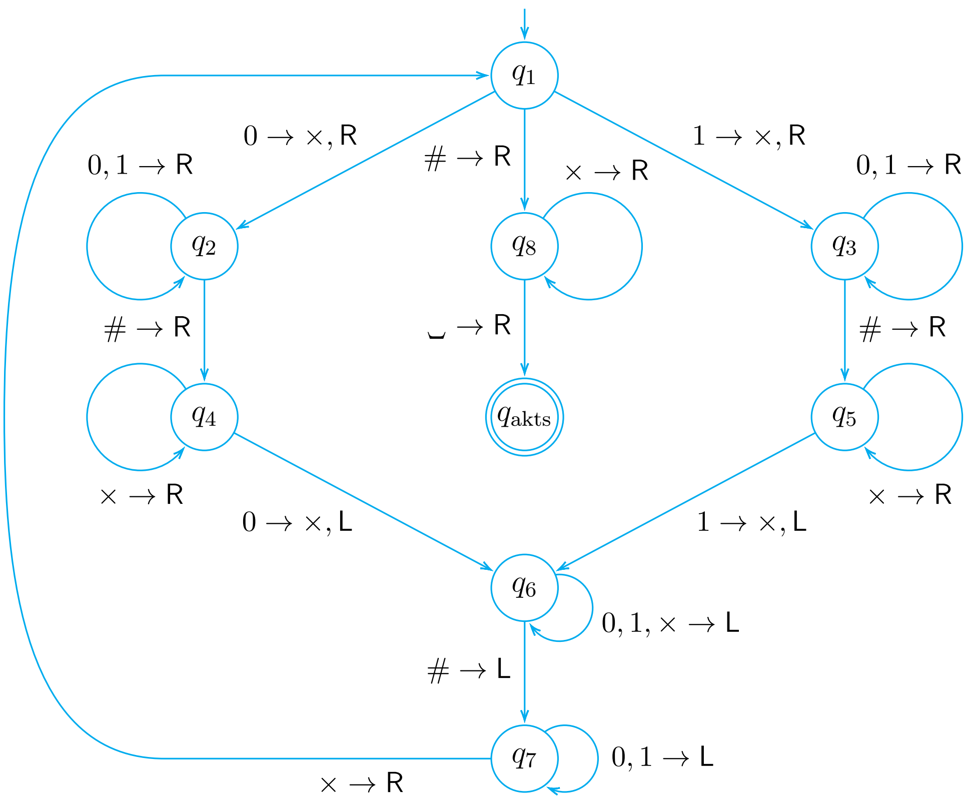
2. Kui kõik sümbolist # vasakul asuvad sümbolid on märgistatud, siis kontrollib, kas sümbolist # paremale on jäänud veel mõni sümbol. Kui on, siis läheb tagasilükkavasse olekusse. Vastasel juhul läheb aktsepteerivasse olekusse.

Vaatleme näiteks masina töötamist sisendil 0110#0110.

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 1 | 1 | 0 | # | 0 | 1 | 1 | 0 | ␣ | … |
| × | 1 | 1 | 0 | # | 0 | 1 | 1 | 0 | ␣ | … |
| × | 1 | 1 | 0 | # | × | 1 | 1 | 0 | ␣ | … |
| × | × | 1 | 0 | # | × | 1 | 1 | 0 | ␣ | … |
| × | × | 1 | 0 | # | × | × | 1 | 0 | ␣ | … |
| | | | · | · | · | · | · | · | | |
| × | × | × | × | # | × | × | × | 0 | ␣ | … |
| × | × | × | × | # | × | × | × | × | ␣ | … |
| × | × | × | × | # | × | × | × | × | ␣ | … |
| × | × | × | × | # | × | × | × | × | ␣ | … |

aktsepteerib

Me võime selle Turingi masina täielikult kirja panna, esitades iga oleku ja sümboli jaoks üleminekufunktsiooni $\delta$ väärtuse järgmise olekudiagrammiga:




Siin loeme, et kui masin kohtab sisendis sümbolit, mis pole diagrammiga kooskõlas, siis siirdub ta olekusse $q_{\text{tag}}$.

## Praktikumiülesanded

**1.** Tõestada, et keel $L = \{1^{n^2} \mid n \geq 0\}$ ei ole regulaarne.

**2.** Olgu $L = \{0^n 1^m \mid n > m\}$. Tõestada, et $L$ ei ole regulaarne.

**3.** Tõestada, et keel $L = \{1^{2^n} \mid n \geq 0\}$ ei ole regulaarne. (See keel koosneb kõigist ühtede järjenditest pikkusega $2^n$, kus $n \geq 0$.)

**4.** Olgu $C_5 = \{x \mid x \text{ on 5-ga jaguv kahendarv}\}$. Tõestada, et $C_5$ on regulaarne.

**5.** Kas lause on tõene või väär?

(a) Kui $L_1 \cup L_2$ on regulaarne ja $L_1$ on lõplik, siis $L_2$ on regulaarne.

(b) Kui $L_1 \cup L_2$ on regulaarne ja $L_1$ on regulaarne, siis $L_2$ on regulaarne.

(c) Kui $L_1 L_2$ on regulaarne ja $L_1$ on regulaarne, siis $L_2$ on regulaarne.

(d) Kui $L^*$ on regulaarne, siis $L$ on regulaarne.

### Lahendused

**1.** *Lahendus.* Oletame, et $L$ on regulaarne. Olgu $p$ pumpamislemmast saadud „pumpamispikkus“. Valime $w = 1^{p^2} \in L$. Siis $|w| = p^2 \geq p$. Järelikult $w = xyz$, kus iga $i \geq 0$ korral $xy^iz \in L$.

Vaatleme sõnet $xy^2z$. Et $|xy| \leq p$, siis $|xy^2z| \leq p^2 + p$. Et $|y| > 0$, siis $|xy^2z| > p^2$. Kuna

$$p^2 < |xy^2z| \leq p^2 + p = p(p+1) < (p+1)^2,$$

siis on sõne $xy^2z$ pikkus rangelt kahe järjestikuse täisruudu vahel. See tähendab, et $xy^2z \notin L$. Saime vastuolu. Järelikult $L$ ei ole regulaarne.

**2.** *Lahendus.* Oletame, et $L$ on regulaarne. Olgu $p$ pumpamislemmast saadav „pumpamispikkus“. Valime $w = 0^{p+1}1^p \in L$. Ilmselt $|w| \geq p$. Järelikult saame selle sõne esitada kujul $w = xyz$. Kuna $|xy| \leq p$, siis koosneb $y$ ainult nullidest. Nüüd ühelt poolt pumpamislemma põhjal $xz \in L$. Teiselt poolt aga on sõnes $xz$ nullide arv rangelt väiksem kui nullide arv sõnes $w = xyz$. Seega on nullide arv sõnes $xz$ ülimalt $p$. See tähendab, $xz \notin L$, vastuolu. Seega $L$ ei ole regulaarne.

**3.** *Lahendus.* Eeldame, et $L$ on regulaarne. Olgu $p$ pumpamislemmast saadav pikkus. Valime $w = 1^{2^p}$. Siis $w \in L$ ja $|w| \geq p$. Järelikult $w = xyz$, kus $|xy| \leq p$, $|y| > 0$ ja iga $i \geq 0$ puhul $xy^iz \in L$.

Kuna alati $p < 2^p$, siis $|y| < 2^p$. Seega $|xyyz| = |xyz| + |y| < 2^p + 2^p = 2^{p+1}$. Järelikult

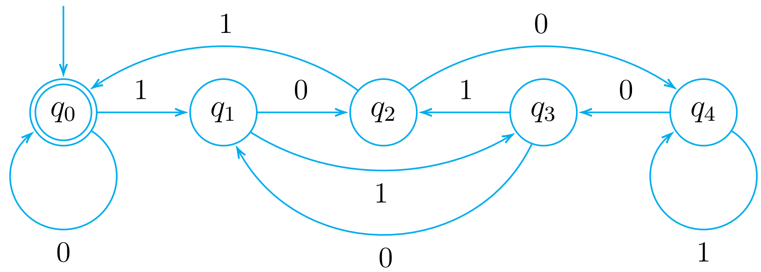
$$2^p < |xyyz| < 2^{p+1}$$

ehk sõne $xyyz$ pikkus ei ole 2 aste. Seetõttu $xyyz \notin L$. Vastuolu. Järelikult $L$ ei ole regulaarne.

**4.** *Lahendus.* Antud keele regulaarsuse tõestamiseks konstrueerime determineeritud lõpliku automaadi $M = (Q, \Sigma, \delta, q_0, F)$, mis tunneb ära keelt $C_5$. Olgu $Q = \{q_0, q_1, q_2, q_3, q_4\}$, $\Sigma = \{0, 1\}$, $F = \{q_0\}$, algolek $q_0$ ja üleminekufunktsioon

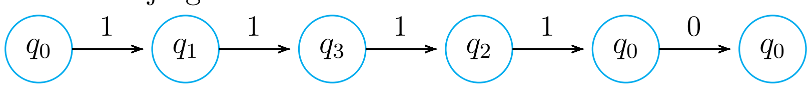
| | 0 | 1 |
|---|---|---|
| $q_0$ | $q_0$ | $q_1$ |
| $q_1$ | $q_2$ | $q_3$ |
| $q_2$ | $q_4$ | $q_0$ |
| $q_3$ | $q_1$ | $q_2$ |
| $q_4$ | $q_3$ | $q_4$ |

Selle automaadi olekudiagramm on




Automaat $M$ peab jooksva oleku kaudu meeles seni loetud arvuosa jääki jagamisel 5-ga: olekud $q_0, \ldots, q_4$ vastavad jääkidele $0, \ldots, 4$. Kui seni sisendist loetud sümbolid moodustavad arvu $x$ ja järgmisena loetakse sümbol 0, siis asendub arv $x$ arvuga $2x$; kui aga loetakse sümbol 1, siis asendub arv $x$ arvuga $2x + 1$. Arvu $2x$ või $2x + 1$ jääk jagamisel 5-ga on üheselt määratud arvu $x$ jäägiga jagamisel 5-ga ning automaat saab sümboli lugemisel minna uuele jäägile vastavasse olekusse.

Näiteks kui anda automaadile sisendiks arv 11110 (kümnendsüsteemis 30) siis töötab automaat järgmiselt:



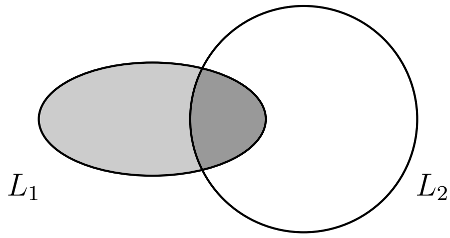

| | | | | | |
|---|---|---|---|---|---|
| sisendist | 1 | 11 | 111 | 1111 | 11110 |
| seni loetud | | $= 3$ | $= 7$ | $= 15$ | $= 30$ |

**5.** *Lahendus.*

(a) Tõene. Esitame keele $L_2$ kujul

$$L_2 = (L_1 \cup L_2) \cap (L_1 \setminus L_2)^c,$$

kus $^c$ tähistab täiendit. Siis $L_1 \cup L_2$ on regulaarne eelduse põhjal, $L_1 \setminus L_2$ on regulaarne, sest on lõplik, ning $(L_1 \setminus L_2)^c$ on regulaarne kui regulaarse keele täiend. Järelikult on $L_2$ kahe regulaarse keele ühisosa ehk ka ise regulaarne.




(b) Väär. Olgu $L_1 = \Sigma^*$ ja $L_2$ suvaline mitteregulaarne keel. Siis $L_1 \cup L_2 = \Sigma^*$ on regulaarne, samuti $L_1$, kuid mitte $L_2$.

(c) Väär. Olgu $L_1 = \{\varepsilon, 0\}$. See on lõplik keel ja seetõttu regulaarne. Olgu $L_2 = (00)^* \cup \{0^{n^2} \mid n \geq 0,\ n \text{ on algarv}\}$. See keel sisaldab kõiki paarisarvulise pikkusega nullijärjendeid ning paaritu pikkusega nullijärjenditest parajasti neid, mille pikkus on algarvu ruut. Keel $L_2$ ei ole regulaarne. Kui oleks, siis saaksime keelest võtta sõne $w = 0^{p^2}$, kus $p$ on paaritu algarv, mille puhul $p^2$ on vähemalt sama suur kui pumpamislemmast saadav pikkus. Esitades sõne kujul $w = xyz$, saame, et $xy^3z \notin L$, sest sõne $xy^3z$ pikkus on paaritu ja $p^2 < |xy^3z| < p^2 + 2p < (p+1)^2$, vastuolu (vt ka 8. nädala praktikumiülesannet 3). Samas aga keel $L_1L_2 = 0^*$ on regulaarne.

(d) Väär. Olgu $L = \{0^n1^n \mid n \geq 0\} \cup \{0, 1\}$. See keel ei ole regulaarne, mida saab tõestada analoogiliselt 8. loengu näitega 1. Kuid $L^* = \Sigma^*$ on regulaarne.