Formaalsete keelte teooria

Mati Pentus
http://lpcs.math.msu.su/~pentus/ftp/fkt/

2009

Peatükk 1. Keeled ja grammatikad

Definitsioon 1.1. *Naturaalarvudeks* nimetame mittenegatiivseid täisarve. Naturaalarvude hulka $\{0, 1, 2, \ldots\}$ tähistame sümboliga $\mathbb{N}$.

Definitsioon 1.2. *Tähestik* (*alfabeet*) on suvaline mittetühi lõplik hulk. Tähestiku elemente nimetatakse *sümboliteks*.

Definitsioon 1.3. *String* (*sõna*) tähestikus Σ on lõplik sümbolite jada, kus iga sümbol kuulub hulka Σ.

Definitsioon 1.4. *Tühi string* on string, mis ei sisalda ühtegi sümbolit (tähistus ε).

Definitsioon 1.5. Tähestiku Σ kõikide stringide hulka tähistatakse Σ^*.

Definitsioon 1.6. Tähestiku Σ kõikide mittetühjade stringide hulka tähistatakse Σ^+.

Definitsioon 1.7. Kui $L \subseteq \Sigma^*$, siis L on *keel* (ehk *formaalne keel*) tähestikus Σ.

Kuna keeled on hulgad, siis on defineeritud kahe keele ühend, ühisosa ja hulgateoreetiline vahe (kasutame tähistusi $L_1 \cup L_2$, $L_1 \cap L_2$, $L_1 - L_2 = \{w \in L_1 \mid w \notin L_2\}$).

Definitsioon 1.8. Olgu $L \subseteq \Sigma^*$. Siis keelt $\Sigma^* - L$ nimetatakse keele L *täiendiks* (*complement*) tähestiku Σ suhtes. Kui on selge, millist tähestikku vaadeldakse, siis öeldakse lihtsalt, et $\Sigma^* - L$ on keele L täiend.

Definitsioon 1.9. Kui x ja y on stringid tähestikus Σ, siis stringi xy (mis on saadud antud stringide järjest kirjutamisel) nimetatakse stringide x ja y *konkatenatsiooniks* (*korrutiseks*). Mõnikord on stringide x ja y konkatenatsiooni tähistuseks $x \cdot y$.

Definitsioon 1.10. Kui x on string ja $n \in \mathbb{N}$, siis x^n tähistab stringi $\underbrace{x \cdot x \cdot \ldots \cdot x}_{n \text{ korda}}$. Seejuures $x^0 \leftrightharpoons \varepsilon$ (märki $\leftrightharpoons$ loetakse „võrdub definitsiooni põhjal“).

Näide 1.11. Kokkuleppe kohaselt $ba^3 = baaa$ ja $(ba)^3 = bababa$.

Näide 1.12. Hulk $\{a^k b a^l \mid k \leqslant l\}$ on keel tähestikus $\{a, b\}$. See keel sisaldab stringid b, ba, aba, baa, $abaa$, $baaa$, $aabaa$, $abaaa$, $baaaa$ jne.

Definitsioon 1.13. Stringi w *pikkus* $|w|$ on sümbolite arv stringis w.

Näide 1.14. Näiteks $|abba| = 4$ ja $|\varepsilon| = 0$.

Definitsioon 1.15. Sümboli a esinemiste arvu stringis w tähistatakse $|w|_a$.

Näide 1.16. Olgu $\Sigma = \{a, b, c\}$. Siis $|baaa|_a = 3$, $|baaa|_b = 1$ ja $|baaa|_c = 0$.

Ülesanne 1.17. Olgu $L_1 = \{(ab)^n \mid n \geqslant 0\}$ ja $L_2 = \{a^m b^m \mid m \geqslant 0\}$. Loetleda kõik stringid keeles $L_1 \cap L_2$.

Ülesanne 1.18. Olgu $\Sigma = \{a, b, c\}$. Kas keeled $L_1 = \{(abc)^n a \mid n \geqslant 2\}$ ja $L_2 = \{ab(cab)^n ca \mid n \geqslant 1\}$ on võrdsed?

Ülesanne 1.19. Olgu $\Sigma = \{a, b, c\}$, $L_1 = \{w \in \Sigma^* \mid |w| = 4\}$ ja $L_2 = \{w \in \Sigma^* \mid |w|_c = 1\}$. Mitu stringi on keeles $L_1 - L_2$?

Definitsioon 1.20. Kui $L_1, L_2 \subseteq \Sigma^*$, siis $L_1 \cdot L_2 \leftrightharpoons \{xy \mid x \in L_1,\ y \in L_2\}$. Keelt $L_1 \cdot L_2$ nimetatakse keelte L_1 ja L_2 *konkatenatsiooniks.*

Näide 1.21. Kui $L_1 = \{a, abb\}$ ja $L_2 = \{bbc, c\}$, siis $L_1 \cdot L_2 = \{ac, abbc, abbbbc\}$.

Definitsioon 1.22. Olgu $L \subseteq \Sigma^*$. Siis

$$L^0 \leftrightharpoons \{\varepsilon\},$$
$$L^n \leftrightharpoons \underbrace{L \cdot \ldots \cdot L}_{n \text{ korda}}, \text{ kui } n > 0.$$

Näide 1.23. Kui $L = \{a^k b a^l \mid 0 < k < l\}$, siis $L^2 = \{a^k b a^l b a^m \mid 0 < k < l-1,\ m > 1\}$.

Ülesanne 1.24. Olgu $\Sigma = \{a, b\}$ ja $L = \{aa, ab\}$. Leida L^3.

Definitsioon 1.25. Keele L *iteratsiooniks* (*Kleene'i katteks*, *Kleene closure, Kleene star, star operation*) nimetatkse keelt $L^* \leftrightharpoons \bigcup\limits_{n \in \mathbb{N}} L^n$.

Näide 1.26. Kui $\Sigma = \{a, b\}$ ja $L = \{aa, ab, ba, bb\}$, siis $L^* = \{w \in \Sigma^* \mid |w| \,\vdots\, 2\}$.

Ülesanne 1.27. Olgu $\Sigma = \{a, b, c, d\}$ ja $L = \{w \in \Sigma^* \mid |w|_a = 1,\ |w|_c = 1\}$. Kas $abcdcacdcabbacba \in L^*$?

Definitsioon 1.28. Stringi w *peegelpildiks* (*reversal*) (tähistus w^{R}) nimetatakse stringi, mis on saadud stringist w sümbolite järjekorra vastupidiseks muutmise teel.

Näide 1.29. Kui $w = baaca$, siis $w^{\mathrm{R}} = acaab$.

Definitsioon 1.30. Kui $L \subseteq \Sigma^*$, siis $L^{\mathrm{R}} \leftrightharpoons \{w^{\mathrm{R}} \mid w \in L\}$.

Definitsioon 1.31. String x on stringi y *prefiks* (tähistame $x \sqsubset y$), kui leidub string u, mille korral $y = xu$.

Näide 1.32. Definitsioonist ilmneb, et $\varepsilon \sqsubset baa$, $b \sqsubset baa$, $ba \sqsubset baa$ ja $baa \sqsubset baa$.

Definitsioon 1.33. String x on stringi y *sufiks* (tähistame $x \sqsupset y$), kui leidub string u, mille korral $y = ux$.

Definitsioon 1.34. *Generatiivne grammatika* (*0-tüüpi grammatika* ehk lihtsalt *grammatika*, *generative grammar, rewrite grammar*) on nelik $G = \langle N, \Sigma, P, S\rangle$, kus N ja Σ on lõplikud hulgad, $N \cap \Sigma = \varnothing$, $P \subset (N \cup \Sigma)^+ \times (N \cup \Sigma)^*$, P on lõplik ja $S \in N$. Siin Σ on *terminalide tähestik*, N on *mitteterminalide tähestik*, mille elemente nimetatakse ka *muutujateks* (*nonterminal, variable*), S on *lähtesümbol* (ehk *stardisümbol*). Stringipaare $(\alpha, \beta) \in P$ nimetatakse *tuletusreegliteks* (*produktsioonireegliteks*, *produktsioonideks*) ja kirjutatkse kujul $\alpha \to \beta$.

Näide 1.35. Olgu $N = \{S\}$, $\Sigma = \{a, b, c\}$ ja $P = \{S \to acSbcS,\ cS \to \varepsilon\}$. Siis $\langle N, \Sigma, P, S\rangle$ on grammatika.

Märkus 1.36. Hulga Σ elemente tähistame väiketähtedega ladina tähestiku algusest ja hulga N elemente tähistame suurte ladina tähtedega. Näidetes ja ülesannetes esitatakse grammatika tavaliselt tuletusreeglite loeteluna. Seejuures mitteterminalideks on kõik tuletusreeglites esinevad suurtähed ja terminalideks kõik väiketähed ning tuletusreeglid loetletakse sellises järjekorras, et esimese tuletusreegli esimene sümbol on grammatika algsümbol.

Märkus 1.37. Kui grammatikas on n reeglit ühesuguste vasakute pooltega $\alpha \to \beta_1, \ldots, \alpha \to \beta_n$, siis võib kasutada lühendatud kirjutusviisi $\alpha \to \beta_1 \mid \ldots \mid \beta_n$.

Definitsioon 1.38. Olgu antud grammatika G. Kirjutame $\phi \underset{G}{\Rightarrow} \psi$ (või lihtsalt $\phi \Rightarrow \psi$), kui leiduvad stringid α, β, η, θ tähestikus $N \cup \Sigma$, mille korral $\phi = \eta\alpha\theta$, $\psi = \eta\beta\theta$ ja $(\alpha \to \beta) \in P$. Kui $\phi \underset{G}{\Rightarrow} \psi$, siis öeldakse, et string ψ on *otseselt tuletatud* sõnast ϕ.

Näide 1.39. Olgu $G = \langle \{S\}, \{a, b, c\}, \{S \to acSbcS,\ cS \to \varepsilon\}, S\rangle$. Siis $cSacS \underset{G}{\Rightarrow} cSa$.

Definitsioon 1.40. Kui $\omega_0 \Rightarrow \omega_1 \Rightarrow \ldots \Rightarrow \omega_n$, kus $n \geqslant 0$, siis kirjutatakse $\omega_0 \overset{*}{\Rightarrow} \omega_n$ ja öeldakse, et string ω_n on *tuletatav* stringist ω_0. Teisiti öeldes seos $\overset{*}{\Rightarrow}$ on hulgal $(N \cup \Sigma)^*$ defineeritud seose $\Rightarrow$ refleksiivne transitiivne kate.

Seejuures stringide jada $\omega_0, \omega_1, \ldots, \omega_n$ nimetatakse stringi ω_n *tuletuseks* (ehk *derivatsiooniks*, *derivation*) stringist ω_0 grammatikas G. Arvu n nimetatakse selle tuletuse *pikkuseks*. (ehk *sammude arvuks*).

Märkus 1.41. Definitsioonist ilmneb, et iga stringi $\omega \in (N \cup \Sigma)^*$ korral $\omega \overset{*}{\Rightarrow} \omega$ (kuna on lubatud tuletus, mille pikkus on 0).

Näide 1.42. Olgu $G = \langle \{S\}, \{a, b\}, \{S \to aSa,\ S \to b\}, S\rangle$. Siis $aSa \overset{*}{\Rightarrow} aaaaSaaaa$. Selle tuletuse pikkus on 3.

Definitsioon 1.43. Grammatika G poolt *genereeritav* keel on

$$L(G) \leftrightharpoons \{\omega \in \Sigma^* \mid S \underset{G}{\overset{*}{\Rightarrow}} \omega\}.$$

Näide 1.44. Kui $G = \langle \{S\}, \{a, b\}, \{S \to aSa,\ S \to bb\}, S\rangle$, siis

$$L(G) = \{a^n bba^n \mid n \geqslant 0\}.$$

Definitsioon 1.45. Grammatikad G_1 ja G_2 on *ekvivalentsed*, kui $L(G_1) = L(G_2)$.

Näide 1.46. Grammatika

$$S \to abS,$$
$$S \to a$$

ja grammatika

$$T \to aU,$$
$$U \to baU,$$
$$U \to \varepsilon$$

on ekvivalentsed.

Ülesanne 1.47. Kirjeldada järgmise grammatika poolt genereeritavat keelt:

$$\begin{aligned} S &\to FF, \\ F &\to aFb, \\ F &\to ab. \end{aligned}$$

Ülesanne 1.48. Kirjeldada järgmise grammatika poolt genereeritavat keelt:

$$\begin{aligned} C &\to Cc, \\ C &\to A, \\ A &\to aAb, \\ A &\to \varepsilon. \end{aligned}$$

Ülesanne 1.49. Kirjeldada järgmise grammatika poolt genereeritavat keelt:

$$\begin{aligned} K &\to \varepsilon, \\ K &\to a, \\ K &\to b, \\ K &\to aKa, \\ K &\to bKb. \end{aligned}$$

Ülesanne 1.50. Kirjeldada järgmise grammatika poolt genereeritavat keelt:

$$\begin{aligned} D &\to DA, \\ DAA &\to ADb, \\ ADA &\to b, \\ A &\to a. \end{aligned}$$

Ülesanne 1.51. Kirjeldada järgmise grammatika poolt genereeritavat keelt:

$$\begin{aligned} F &\to aFH, \\ F &\to abc, \\ bH &\to bbc, \\ cH &\to Hc. \end{aligned}$$

Ülesanne 1.52. Kirjeldada järgmise grammatika poolt genereeritavat keelt:

$$\begin{aligned} S &\to aTS, \\ S &\to U, \\ Ta &\to aaT, \\ TU &\to U, \\ U &\to a. \end{aligned}$$

Ülesanne 1.53. Kas grammatika

$$\begin{aligned} S &\to ab, \\ S &\to aKSb, \\ K &\to bSb, \\ KS &\to b, \\ K &\to \varepsilon \end{aligned}$$

ja grammatika

$$\begin{aligned} S &\to aTb, \\ T &\to \varepsilon, \\ T &\to b, \\ T &\to S, \\ T &\to bSbS \end{aligned}$$

on ekvivalentsed?

Ülesanne 1.54. Kas grammatika

$$\begin{aligned} S &\to aD, \\ D &\to bba, \\ D &\to baDa, \\ D &\to aDaDa \end{aligned}$$

ja grammatika

$$\begin{aligned} S &\to aaE, \\ S &\to abJ, \\ E &\to bJJ, \\ J &\to aaEa, \\ J &\to abJa, \\ J &\to ba \end{aligned}$$

on ekvivalentsed?

Definitsioon 1.55. *Kontekstitundlik grammatika* (*1-tüüpi grammatika*, *kontekstist sõltuv grammatika*, *context-sensitive grammar*, *phrase-structure grammar*) on grammatika, mille iga tuletusreegel on kujul $\xi A\zeta \to \xi\alpha\zeta$, kus $A \in N$, $\xi \in (N \cup \Sigma)^*$, $\zeta \in (N \cup \Sigma)^*$, $\alpha \in (N \cup \Sigma)^+$.

Näide 1.56. Grammatika

$$\begin{aligned} S &\to TS, & A &\to a, \\ S &\to US, & TA &\to AAT, \\ S &\to b, & UAb &\to b, \\ Tb &\to Ab, & UAAA &\to AAU \end{aligned}$$

ei ole kontekstitundlik.

Definitsioon 1.57. *Kontekstivaba grammatika* (*2-tüüpi grammatika*, context-free grammar) on grammatika, mille iga tuletusreegel on kujul $A \to \alpha$, kus $A \in N$, $\alpha \in (N \cup \Sigma)^*$.

Näide 1.58. Grammatika

$$\begin{array}{ll} S \to ASTA, & AT \to UT, \\ S \to AbA, & UT \to UV, \\ A \to a, & UV \to TV, \\ bT \to bb, & TV \to TA \end{array}$$

on kontekstitundlik, kuid ei ole kontekstivaba.

Definitsioon 1.59. *Lineaarne grammatika* (linear grammar) on grammatika, mille iga tuletusreegel on kujul $A \to u$ või $A \to uBv$, kus $A \in N$, $u \in \Sigma^*$, $v \in \Sigma^*$, $B \in N$.

Näide 1.60. Grammatika

$$\begin{array}{l} S \to TT, \\ T \to cTT, \\ T \to bT, \\ T \to a \end{array}$$

on kontekstivaba, kuid ei ole lineaarne.

Definitsioon 1.61. *Paremlineaarne grammatika* (*3-tüüpi grammatika*, right-linear grammar) on grammatika, mille iga tuletusreegel on kujul $A \to u$ või $A \to uB$, kus $A \in N$, $u \in \Sigma^*$, $B \in N$.

Näide 1.62. Grammatika

$$\begin{array}{l} S \to aSa, \\ S \to T, \\ T \to bT, \\ T \to \varepsilon \end{array}$$

on lineaarne, kuid ei ole paremlineaarne.

Näide 1.63. Grammatika

$$\begin{array}{l} S \to T, \\ U \to abba \end{array}$$

on paremlineaarne. See grammatika genereerib keele $\varnothing$.

Näide 1.64. Grammatika

$$\begin{array}{ll} S \to aS, & T \to aT, \\ S \to bS, & T \to bT, \\ S \to aaaT, & T \to \varepsilon \\ S \to aabaT, & \\ S \to abaaT, & \\ S \to aabbaT, & \\ S \to ababaT, & \\ S \to abbaaT, & \end{array}$$

on paremlineaarne.

Definitsioon 1.65. Tuletusreeglit kujul $\alpha \to \varepsilon$ nimetatakse *ε-tuletusreegliks*.

Lemma 1.66. *Iga paremlineaarne grammatika on lineaarne. Iga lineaarne grammatika on kontekstivaba. Iga ε-tuletusreegliteta kontekstivaba grammatika on kontekstitundlik.*

Definitsioon 1.67. 0-tüüpi, 1-tüüpi, 2-tüüpi ja 3-tüüpi grammatikad moodustavad *Chomsky hierarhia* (*Chomsky hierarchy*).

Definitsioon 1.68. Keelt nimetatakse *0-tüüpi keeleks* (*kontekstitundlikuks*, *kontekstivabaks*, *lineaarseks*, *paremlineaarseks*), kui seda keelt saab genereerida 0-tüüpi (vastavalt kontekstitundliku, kontekstivaba, lineaarse, paremlineaarse) grammatikaga.

Näide 1.69. Olgu $\Sigma = \{a_1, \ldots, a_n\}$. Keele Σ^* genereerib paremlineaarne grammatika

$$\begin{aligned} S &\to \varepsilon, \\ S &\to a_1 S, \\ &\vdots \\ S &\to a_n S. \end{aligned}$$

Teoreem 1.70. *Kontekstitundlike keelte klass on kinnine ühendi, ühisosa ja täiendi suhtes.*

Peatükk 2. Lõplikud automaadid

Definitsioon 2.1. *Lõplik automaat* (finite automaton, finite-state machine) on viisik $M = \langle Q, \Sigma, \Delta, I, F\rangle$, kus Σ on lõplik *sisendtähestik*, Q ja Δ on lõplikud hulgad, $\Delta \subseteq Q \times \Sigma^* \times Q$, $I \subseteq Q$, $F \subseteq Q$. Hulga Q elemente nimetatakse *olekuteks* (state), hulga I elemente nimetatakse *algolekuteks*(*lähteolekuteks*, initial state), hulga F elemente nimetatakse *lõppolekuteks* (final state, accepting state). Kui $\langle p, x, q\rangle \in \Delta$, siis kolmikut $\langle p, x, q\rangle$ nimetatakse *üleminekuks* (transition) olekust p olekusse q ja string x on selle ülemineku *märgend* (label).

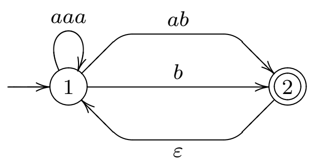
Näide 2.2. Olgu $Q = \{1, 2\}$, $\Sigma = \{a, b\}$, $I = \{1\}$, $F = \{2\}$,

$$\Delta = \{\langle 1, aaa, 1\rangle,\ \langle 1, ab, 2\rangle,\ \langle 1, b, 2\rangle,\ \langle 2, \varepsilon, 1\rangle\}.$$

Siis $\langle Q, \Sigma, \Delta, I, F\rangle$ on lõplik automaat.

Märkus 2.3. Lõplikku automaati võib esitada *seisundidiagrammi* (state diagram, transition diagram) abil. Seisundidiagrammil vastab igale seisundile ring (graafi tipp) ja igale üleminekule nool (graafi serv). Stringiga x märgendatud nool ringist p ringi q näitab, et automaadis on üleminek $\langle p, x, q\rangle$. Iga algolek märgitakse siseneva noolekesega ja iga lõppolek märgitakse kahekordse ringiga.

Näide 2.4. Järgmisel diagrammil on kujutatud lõplik automaat näitest 2.2.



Definitsioon 2.5. Lõpliku automaadi $\langle Q, \Sigma, \Delta, I, F\rangle$ *konfiguratsiooniks* nimetatakse järjestatud paari $\langle q, w\rangle$, kus $q \in Q$ ja $w \in \Sigma^*$.

Definitsioon 2.6. Lõpliku automaadi $M = \langle Q, \Sigma, \Delta, I, F\rangle$ konfiguratsioonide hulgal defineerime binaarse seose $\underset{M}{\vdash}$ (ehk lihtsalt $\vdash$) järgnevalt. Kui $\langle p, x, q\rangle \in \Delta$ ja $w \in \Sigma^*$, siis $\langle p, xw\rangle \vdash \langle q, w\rangle$. Seost $\vdash$ nimetatakse automaadi *töötaktiks* või *sammuks* (step). Binaarne seos $\overset{*}{\vdash}$ on seose $\vdash$ refleksiivne transitiivne kate.

Näide 2.7. Vaatleme lõplikku automaati näitest 2.2. Definitsiooni kohaselt $\langle 1, abba\rangle \vdash \langle 2, ba\rangle$, $\langle 1, aaaab\rangle \overset{*}{\vdash} \langle 1, aaaab\rangle$ ja $\langle 1, aaaab\rangle \overset{*}{\vdash} \langle 2, \varepsilon\rangle$.

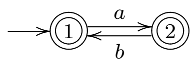
Definitsioon 2.8. Lõplik automaat $M = \langle Q, \Sigma, \Delta, I, F\rangle$ *aktsepteerib* (accepts, recognizes) stringi $w \in \Sigma^*$ parajasti siis, kui leiduvad olekud $p \in I$ ja $q \in F$, mille korral $\langle p, w\rangle \overset{*}{\vdash} \langle q, \varepsilon\rangle$.

Definitsioon 2.9. Lõpliku automaadi M poolt *aktsepteeritav* keel $L(M)$ koosneb selle automaadi poolt aktsepteeritavatest stringidest.

Märkus 2.10. Kui $M = \langle Q, \Sigma, \Delta, I, F\rangle$ ja $I \cap F \neq \varnothing$, siis $\varepsilon \in L(M)$.

Näide 2.11. Olgu $M = \langle Q, \Sigma, \Delta, I, F\rangle$, kus $Q = \{1, 2\}$, $\Sigma = \{a, b\}$, $I = \{1\}$, $F = \{1, 2\}$,

$$\Delta = \{\langle 1, a, 2\rangle,\ \langle 2, b, 1\rangle\}.$$



Siis $L(M) = \{(ab)^n \mid n \geqslant 0\} \cup \{(ab)^n a \mid n \geqslant 0\}$.

Definitsioon 2.12. Lõplikud automaadid M_1 ja M_2 on *ekvivalentsed*, kui $L(M_1) = L(M_2)$.

Märkus 2.13. Iga lõpliku keele jaoks leidub seda keelt aktsepteeriv lõplik automaat.

Ülesanne 2.14. Leida keelt $\{ucv \mid u \in \{a, b\}^*,\ v \in \{a, b\}^*\}$ aktsepteeriv lõplik automaat.

Ülesanne 2.15. Leida keelt $\{w \in \{a, b, c\}^* \mid |w|_c \neq 1\}$ aktsepteeriv lõplik automaat.

Ülesanne 2.16. Leida keelt $\{a, b\}^* - (\{a^n \mid n \geqslant 0\} \cup \{b^n \mid n \geqslant 0\})$ aktsepteeriv lõplik automaat.

Ülesanne 2.17. Leida keelt $\{aub \mid u \in \{a, b\}^*\} \cup \{bua \mid u \in \{a, b\}^*\}$ aktsepteeriv lõplik automaat.

Ülesanne 2.18. Leida keelt $\{w \in \{a, b\}^* \mid |w|_a \geqslant 3\}$ aktsepteeriv lõplik automaat.

Lemma 2.19. *Iga lõpliku automaadi jaoks leidub ekvivalentne ühe algolekuga lõplik automaat, kus iga ülemineku märgendi pikkus on* 1.

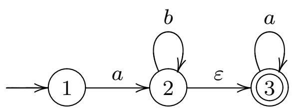
Tõestus. Lihtne on leida ekvivalentne ühe algolekuga lõplik automaat, kus iga ülemineku märgendi pikkus on 0 või 1. Olgu see automaat $M = \langle Q, \Sigma, \Delta, I, F\rangle$. Koostame automaadi $M' = \langle Q, \Sigma, \Delta', I, F'\rangle$, kus

$$\Delta' = \{\langle p, a, r\rangle \mid \exists q \in Q\,(\langle q, a, r\rangle \in \Delta \wedge \langle p, \varepsilon\rangle \underset{M}{\vdash} \langle q, \varepsilon\rangle)\},$$
$$F' = \{p \in Q \mid \exists q \in F\,(\langle p, \varepsilon\rangle \underset{M}{\vdash} \langle q, \varepsilon\rangle)\}.$$

Lihtne on näha, et $L(M') = L(M)$. □

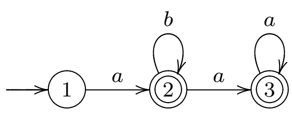
Näide 2.20. Olgu $M = \langle Q, \Sigma, \Delta, I, F\rangle$, kus $Q = \{1, 2, 3\}$, $\Sigma = \{a, b\}$, $I = \{1\}$, $F = \{3\}$,

$$\Delta = \{\langle 1, a, 2\rangle,\ \langle 2, b, 2\rangle,\ \langle 2, \varepsilon, 3\rangle,\ \langle 3, a, 3\rangle\}.$$



Lemmat 2.19 rakendades saame $M' = \langle Q, \Sigma, \Delta', I, F'\rangle$, kus $F' = \{2, 3\}$ ja

$$\Delta' = \{\langle 1, a, 2\rangle,\ \langle 2, b, 2\rangle,\ \langle 2, a, 3\rangle,\ \langle 3, a, 3\rangle\}.$$



Teoreem 2.21. *Keel on paremlineaarne parajasti siis, kui leidub lõplik automaat, mis seda keelt aktsepteerib.*

Näide 2.22. Näites 2.2 esitatud lõpliku automaadi poolt aktsepteeritava keele genereerib paremlineaarne grammatika

$$
\begin{aligned}
K_1 &\to aaaK_1,\\
K_1 &\to abK_2,\\
K_1 &\to bK_2,\\
K_2 &\to K_1,\\
K_2 &\to \varepsilon.
\end{aligned}
$$

Näide 2.23. Olgu $\Sigma = \{a, b\}$. Paremlineaarne grammatika

$$
\begin{aligned}
S &\to aa,\\
S &\to T,\\
T &\to baT,\\
T &\to a.
\end{aligned}
$$

on ekvivalentne paremlineaarse grammatikaga

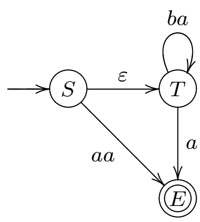
$$
\begin{aligned}
S &\to aaE,\\
S &\to T,\\
T &\to baT,\\
T &\to aE,\\
E &\to \varepsilon.
\end{aligned}
$$

Nenede grammatikate poolt genereeritavat keelt aktsepteerib lõplik automaat

$$\langle Q, \Sigma, \Delta, I, F\rangle,$$

kus $Q = \{S, T, E\}$, $I = \{S\}$, $F = \{E\}$ ja

$$\Delta = \{\langle S, aa, E\rangle,\ \langle S, \varepsilon, T\rangle,\ \langle T, ba, T\rangle,\ \langle T, a, E\rangle\}.$$



Ülesanne 2.24. Koostada paremlineaarne grammatika, mis genereerib keele $\{w \in \{a, b\}^* \mid |w|_a \vdots 3,\ |w|_b \geqslant 3\}$.

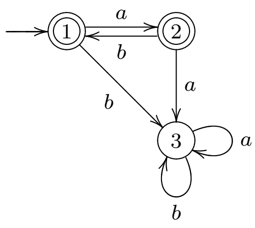
Definitsioon 2.25. Lõplik automaat $\langle Q, \Sigma, \Delta, I, F\rangle$ on *determineeritud* (*deterministlik*, *deterministic*), kui

1) hulgas I on täpselt üks element;

2) iga ülemineku $\langle p, x, q\rangle \in \Delta$ jaoks kehtib võrdus $|x| = 1$;

3) iga sümboli $a \in \Sigma$ ja iga oleku $p \in Q$ jaoks leidub täpselt üks olek $q \in Q$, mille korral $\langle p, a, q\rangle \in \Delta$.

Näide 2.26. Lõplik automaat



on determineeritud.

Teoreem 2.27. *Iga lõpliku automaadi jaoks leidub ekvivalentne determineeritud lõplik automaat.*

Tõestus. Lemma 2.19 alusel võime eeldada, et antud lõplikus automaadis $\langle Q, \Sigma, \Delta, I, F\rangle$ on iga ülemineku märgendi pikkus 1. Iga $a \in \Sigma$ ja $H \subseteq Q$ korral tähistame

$$\Delta_a(H) \coloneqq \{q \in Q \mid \exists p \in H\,(\langle p, a, q\rangle \in \Delta)\}.$$

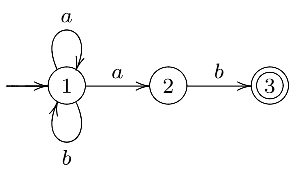
Hulga Q kõigi alamhulkade hulka tähistatakse $\mathcal{P}(Q)$. Koostame determineeritud lõpliku automaadi $M' = \langle Q', \Sigma, \Delta', I', F'\rangle$, kus $Q' = \mathcal{P}(Q)$,

$$\Delta' = \{\langle H, a, \Delta_a(H)\rangle \mid H \subseteq Q,\ a \in \Sigma\},$$

$I' = \{I\}$ ja $F' = \{H \subseteq Q \mid H \cap F \neq \varnothing\}$. □

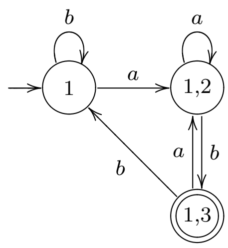
Näide 2.28. Olgu $\Sigma = \{a, b\}$ ja $M = \langle\{1, 2, 3\}, \Sigma, \Delta, \{1\}, \{3\}\rangle$, kus

$$\Delta = \{\langle 1, a, 1\rangle,\ \langle 1, b, 1\rangle,\ \langle 1, a, 2\rangle,\ \langle 2, b, 3\rangle\}.$$



Järgides teoreemi 2.27 töestust saame $M' = \langle\{\{1\}, \{1, 2\}, \{1, 3\}\}, \Sigma, \Delta', \{\{1\}\}, \{\{1, 3\}\}\rangle$, kus

$$\Delta' = \{\langle\{1\}, a, \{1, 2\}\rangle,\ \langle\{1\}, b, \{1\}\rangle,\ \langle\{1, 2\}, a, \{1, 2\}\rangle,\\ \langle\{1, 2\}, b, \{1, 3\}\rangle,\ \langle\{1, 3\}, a, \{1, 2\}\rangle,\ \langle\{1, 3\}, b, \{1\}\rangle\}.$$



Peatükk 3. Lõplike automaatide keelte omadused

Teoreem 3.1. *Lõplike automaatide poolt aktsepteeritavate keelte klass on kinnine iteratsiooni, konkatenatsiooni ja ühendi suhtes.*

Teoreem 3.2. *Lõplike automaatide poolt aktsepteeritavate keelte klass on kinnine täiendi ja ühisosa suhtes.*

Lemma 3.3 (pumpamislemma, kasvatamise lemma, *pumping lemma*)**.** *Olgu L lõpliku automaadi poolt aktsepteeritav keel tähestikus Σ. Siis leidub selline positiivne täisarv p, et iga stringi $w \in L$ korral, kui $|w| \geqslant p$, siis leiduvad sellised stringid $x, y, z \in \Sigma^*$, et $xyz = w$, $y \neq \varepsilon$, $|xy| \leqslant p$ ja $xy^iz \in L$ iga $i \geqslant 0$ korral.*

Näide 3.4. Olgu $\Sigma = \{a, b\}$ ja $L = \{ab^na^n \mid n \geqslant 0\}$. Lemmas 3.3 formuleeritud väited ei ole tõesed ühegi p väärtuse korral. Tõepoolest, kui $w = ab^pa^p$, siis $x = ab^k$, $y = b^m$, $z = b^{p-k-m}a^p$ mingite arvude $k \geqslant 0$ ja $m \geqslant 1$ korral või $x = \varepsilon$, $y = ab^l$, $z = b^{p-l}a^p$ mingi arvu $l \geqslant 0$ korral. Mõlemal juhul $xyyz \notin L$. Järelikult L ei ole lõpliku automaadi poolt aktsepteeritav keel.

Näide 3.5. Olgu $\Sigma = \{a, b\}$ ja $L = \{a^kb^ma^n \mid k = 0 \text{ või } m = n\}$. Olgu $p = 1$. Siis iga stringi $w \in L$ jaoks, kui $|w| \geqslant p$, siis leiduvad stringid $x, y, z \in \Sigma^*$, mille korral Lemmas 3.3 formuleeritud omadused on tõesed. Sellele vaatamata L ei ole lõpliku automaadi poolt aktsepteeritav keel, kuna

$$L \cap \{ab^ma^n \mid m \geqslant 0,\ n \geqslant 0\} = \{ab^na^n \mid n \geqslant 0\}$$

ja teoreemi 3.2 kohaselt oleks koos keelega L ka keel $\{ab^na^n \mid n \geqslant 0\}$ lõpliku automaadi poolt aktsepteeritav.

Peatükk 4. Regulaaravaldised

Definitsioon 4.1. *Regulaaravaldis* (*regular expression*) üle tähestiku Σ defineeritakse rekursiivselt: 0 on regulaaravaldis; 1 on regulaaravaldis; kui $a \in \Sigma$, siis a on regulaaravaldis; kui e ja f on regulaaravaldised, siis ka $(e+f)$, $(e \cdot f)$ ja e^* on regulaaravaldised.

Kõige tugevamalt seob tehe *, järgneb korrutamine, kõige nõrgemalt seob liitmine. Tavaliselt kirjutatakse $e \cdot f$ asemel lihtsalt ef.

Definitsioon 4.2. Iga regulaaravaldis e *esitab* (*denotes, represents*) teatud keele (tähistus $L(e)$), mis defineeritakse rekursiivselt:

$$\begin{aligned} L(a) &\fallingdotseq \{a\}, \text{ kui } a \in \Sigma, \\ L(0) &\fallingdotseq \varnothing, \\ L(1) &\fallingdotseq \{\varepsilon\}, \\ L(e+f) &\fallingdotseq L(e) \cup L(f), \\ L(e \cdot f) &\fallingdotseq L(e) \cdot L(f), \\ L(e^*) &\fallingdotseq L(e)^*. \end{aligned}$$

Tavaliselt kirjutatakse $L(e)$ asemel lihtsalt e.

Näide 4.3. Olgu $\Sigma = \{a, b\}$. Definitsiooni kohaselt

$$L((ab)^* \cdot (1+a)) = \{(ab)^n \mid n \geqslant 0\} \cup \{(ab)^n a \mid n \geqslant 0\}.$$

Definitsioon 4.4. Keel on *regulaarne*, kui leidub seda keelt esitav regulaaravaldis.

Definitsioon 4.5. Kui e on regulaaravaldis, siis $e^+ \fallingdotseq e^* e$.

Ülesanne 4.6. Lihtsustada regulaaravaldis $((a+bc)^*)^*$.

Teoreem 4.7 (Kleene'i teoreem)**.** *Keel on regulaarne parajasti siis, kui leidub lõplik automaat, mis seda keelt aktsepteerib.*

Peatükk 5. Kontekstivabad grammatikad

Definitsioon 5.1. Tuletust kontekstivabas grammatikas nimetatakse *vasaktuletuseks* (*leftmost derivation*), kui tuletuse igal sammul asendatakse vasakpolseim mitteterminal (see tähendab, et iga samm on esitatav kujul $uA\theta \Rightarrow u\beta\theta$, kus $(A \to \beta) \in P$, $u \in \Sigma^*$ ja $\theta \in (N \cup \Sigma)^*$).

Definitsioon 5.2. Kontekstivaba grammatika on *mitmene* (*ambiguous*), kui leidub string, millel on vähemalt kaks vasaktuletust. Vastasel korral on kontekstivaba grammatika *ühene* (*unambiguous*).

Definitsioon 5.3. *Dyck'i keel* üle $2n$ sümboli on kontekstivaba grammatika

$$\begin{aligned} &S \to \varepsilon, \\ &S \to a_1 S b_1 S, \\ &\quad\vdots \\ &S \to a_n S b_n S \end{aligned}$$

poolt genereeritav keel tähestikus $\{a_1, b_1, a_2, b_2, \ldots, a_n, b_n\}$.

Teoreem 5.4. *String w tähestikus $\{a, b\}$ on grammatikas*

$$\begin{aligned} &S \to \varepsilon, \\ &S \to aSbS \end{aligned}$$

tuletatav parajasti siis, kui $|w|_a = |w|_b$ ja $\forall x \sqsubset w\ (|x|_a \geqslant |x|_b)$.

Definitsioon 5.5. *Łukasiewicz'i keel* üle $n+1$ sümboli on kontekstivaba grammatika

$$\begin{aligned} &S \to a_0, \\ &S \to a_1 S, \\ &S \to a_2 SS, \\ &\quad\vdots \\ &S \to a_n S^n \end{aligned}$$

poolt genereeritav keel tähestikus $\{a_0, a_1, \ldots, a_n\}$.

Teoreem 5.6. *String w tähestikus $\{a_0, a_1 \ldots, a_n\}$ kuulub Łukasiewicz'i keelde üle $n+1$ sümboli parajasti siis, kui $\sum_{i=0}^{n}(i-1)|w|_{a_i} = -1$ ja $\forall x \underset{\neq}{\sqsubset} w\ (\sum_{i=0}^{n}(i-1)|x|_{a_i} \geqslant 0)$. (Siin $x \underset{\neq}{\sqsubset} w$ tähendab, et $x \sqsubset w$ ja $x \neq w$.)*

Teoreem 5.7. *Kui keel L on kontekstivaba, siis leidub ε-tuletusreegilteta kontekstivaba grammatika, mis genereerib keele $L - \{\varepsilon\}$.*

Definitsioon 5.8. Kontekstivaba grammatika $\langle N, \Sigma, P, S\rangle$ on *Chomsky normaalkujul* (*grammar in Chomsky normal form*), kui iga tuletusreegel on ühel järgmistest kujudest: $S \to \varepsilon$, $A \to a$, $A \to BC$, kus $A \in N$, $B \in N - \{S\}$, $C \in N - \{S\}$, $a \in \Sigma$.

Teoreem 5.9. *Iga kontekstivaba grammatika jaoks leidub ekvivalentne kontekstivaba grammatika Chomsky normaalkujul.*

Lemma 5.10 (pumpamislemma, kasvatamise lemma, *pumping lemma*)**.** *Olgu L kontekstivaba keel tähestikus Σ. Siis leidub selline positiivne täisarv p, et iga stringi $w \in L$ jaoks, kui $|w| \geqslant p$, siis leiduvad sellised stringid $u, v, x, y, z \in \Sigma^*$, et $uvxyz = w$, $vy \neq \varepsilon$, $|vxy| \leqslant p$ ja $uv^ixy^iz \in L$ iga $i \geqslant 0$ korral.*

Näide 5.11. Olgu $\Sigma = \{a, b, c\}$ ja $L = \{a^nb^nc^n \mid n \geqslant 0\}$. Lemmas 5.10 formuleeritud väited ei ole tõesed ühegi p väärtuse korral. Tõepoolest, kui $uvxyz = a^pb^pc^p$, $|vy| > 0$ ja $|vxy| \leqslant p$, siis $|vy|_a = 0$ või $|vy|_c = 0$. Järelikult, $|uvvxyyz|_a = p$ või $|uvvxyyz|_c = p$. Kuna $|uvvxyyz| > 3p$, siis $uvvxyyz \notin L$. Järelikult L ei ole kontekstivaba keel.

Teoreem 5.12. *Kui $|\Sigma| = 1$ ja keel $L \subseteq \Sigma^*$ on kontekstivaba, siis L on regulaarne keel.*

Teoreem 5.13. *Kontekstivabade keelte klass on kinnine ühendi, konkatenatsiooni ja iteratsiooni suhtes.*

Teoreem 5.14. *Leiduvad kontekstivabad keeled L_1 ja L_2, mille korral $L_1 \cap L_2$ ei ole kontekstivaba.*

Tõestus. Olgu $L_1 = \{a^mb^mc^n \mid m \geqslant 0,\ n \geqslant 0\}$ ja $L_2 = \{a^mb^nc^n \mid m \geqslant 0,\ n \geqslant 0\}$. Näitest 5.11 järeldub, et keel $L_1 \cap L_2$ ei ole kontekstivaba. □

Teoreem 5.15. *Leidub kontekstivaba keel $L \subseteq \Sigma^*$, mille korral $\Sigma^* - L$ ei ole kontekstivaba.*

Teoreem 5.16. *Kui keel L_1 on kontekstivaba ja keel L_2 on regulaarne, siis keel $L_1 \cap L_2$ on kontekstivaba.*

Tõestus. Olgu $G = \langle N, \Sigma, P, S\rangle$ keelt L_1 genereeriv kontekstivaba grammatika. Võime eeldada, et selles grammatikas iga tuletusreegel on kujul $A \to a$ või $A \to \alpha$, kus $A \in N$, $a \in \Sigma$ ja $\alpha \in N^*$. Olgu $M = \langle Q, \Sigma, \Delta, I, F\rangle$ keelt L_2 aktsepteeriv lõplik automaat. Lemma 2.19 alusel võime eeldada, et iga ülemineku $\langle p, x, q\rangle \in \Delta$ jaoks kehtib võrdus $|x| = 1$.

Koostame kontekstibvaba grammatika $\overline{G} = \langle \overline{N}, \Sigma, \overline{P}, \overline{S}\rangle$, mis genereerib keele $L_1 \cap L_2$:

$$\begin{aligned}
\overline{N} &= \{\overline{S}\} \cup (Q \times N \times Q),\\
\overline{P} &= \{\overline{S} \to \langle p, S, q\rangle \mid p \in I,\ q \in F\} \cup\\
&\quad \cup \{\langle p, A, q\rangle \to a \mid \langle p, a, q\rangle \in \Delta,\ (A \to a) \in P\} \cup\\
&\quad \cup \{\langle p_0, A, p_n\rangle \to \langle p_0, B_1, p_1\rangle \ldots \langle p_{n-1}, B_n, p_n\rangle \mid\\
&\qquad (A \to B_1 \ldots B_n) \in P,\ p_0 \in Q, \ldots,\ p_n \in Q\},
\end{aligned}$$

kus $\overline{S}$ on uus sümbol, mis ei kuulu hulka $Q \times N \times Q$.

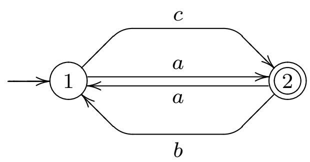
Induktsiooniga stringi $w \in \Sigma^*$ pikkuse järgi võib tõestada, et $\langle p, A, q\rangle \overset{*}{\underset{\overline{G}}{\Rightarrow}} w$ parajasti siis, kui $A \overset{*}{\underset{G}{\Rightarrow}} w$ ja $\langle p, w\rangle \overset{*}{\underset{M}{\vdash}} \langle q, \varepsilon\rangle$. Järelikult $L(\overline{G}) = L_1 \cap L_2$. □

Näide 5.17. Olgu $\Sigma = \{a, b, c\}$. Vaatleme kontekstivaba keelt L_1, mille genereerib grammatika

$$\begin{aligned} S &\to a, \\ S &\to BS, \\ S &\to CSS, \\ B &\to b, \\ C &\to c, \end{aligned}$$

ja regulaarset keelt L_2, mille aktsepteerib lõplik automaat $M = \langle Q, \Sigma, \Delta, I, F \rangle$, kus $Q = \{1, 2\}$, $I = \{1\}$, $F = \{2\}$,

$$\Delta = \{\langle 1, a, 2\rangle,\ \langle 1, c, 2\rangle,\ \langle 2, a, 1\rangle,\ \langle 2, b, 1\rangle\}.$$



Siis keele $L_1 \cap L_2$ genereerib kontekstivaba grammatika

$$\begin{aligned} \overline{S} &\to S_{12}, & S_{11} &\to cS_{21}S_{11}, \\ S_{12} &\to a, & S_{11} &\to cS_{22}S_{21}, \\ S_{21} &\to a, & S_{12} &\to cS_{21}S_{12}, \\ S_{21} &\to bS_{11}, & S_{12} &\to cS_{22}S_{22}. \\ S_{22} &\to bS_{12}, \end{aligned}$$

Siin S_{11}, S_{12}, S_{21} ja S_{22} vastavad sümbolitele $\langle 1, S, 1\rangle$, $\langle 1, S, 2\rangle$, $\langle 2, S, 1\rangle$ ja $\langle 2, S, 2\rangle$ teoreemi 5.16 tõestusest.

Peatükk 6. Magasinmäluga automaadid

Definitsioon 6.1. *Magasinmäluga automaat* (*magasiniga automaat*, *pinuautomaat*, *pushdown automaton*) on kuuik $M = \langle Q, \Sigma, \Gamma, \Delta, I, F \rangle$, kus Q, Σ, Δ ja Γ on lõplikud hulgad, $I \subseteq Q$, $F \subseteq Q$ ja

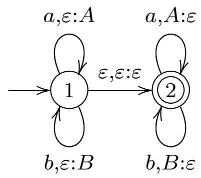
$$\Delta \subseteq (Q \times \Sigma^* \times \Gamma^*) \times (Q \times \Gamma^*).$$

Hulga Q elemente nimetatakse *olekuteks*, hulga I elemente nimetatakse *algolekuteks*, hulga F elemente nimetatakse *lõppolekuteks*, Σ on *sisendtähestik*, Γ on magasinitähestik (*stack alphabet*) ja Δ on *üleminekute hulk* (*transition relation*).

Näide 6.2. Olgu $Q = \{1, 2\}$, $\Sigma = \{a, b\}$, $\Gamma = \{A, B\}$,

$$\Delta = \{\langle\langle 1, a, \varepsilon\rangle, \langle 1, A\rangle\rangle,\ \langle\langle 1, b, \varepsilon\rangle, \langle 1, B\rangle\rangle,\ \langle\langle 1, \varepsilon, \varepsilon\rangle, \langle 2, \varepsilon\rangle\rangle,$$
$$\langle\langle 2, a, A\rangle, \langle 2, \varepsilon\rangle\rangle,\ \langle\langle 2, b, B\rangle, \langle 2, \varepsilon\rangle\rangle\},$$

$I = \{1\}$, $F = \{2\}$. Siis $\langle Q, \Sigma, \Gamma, \Delta, I, F \rangle$ on magasinmäluga automaat. See automaat on kujutatud järgmisel diagrammil.



Definitsioon 6.3. Magasinmäluga automaadi $\langle Q, \Sigma, \Delta, I, F \rangle$ *konfiguratsiooniks* nimetatakse kolmikut $\langle q, w, \alpha \rangle$, kus $q \in Q$, $w \in \Sigma^*$, $\alpha \in \Gamma^*$.

Definitsioon 6.4. Magasinmäluga automaadi $M = \langle Q, \Sigma, \Gamma, \Delta, I, F \rangle$ konfiguratsioonide hulgal defineerime binaarse seose $\underset{M}{\vdash}$ (ehk lihtsalt $\vdash$) järgnevalt. Kui $\langle\langle p, x, \beta\rangle, \langle q, \gamma\rangle\rangle \in \Delta$, $w \in \Sigma^*$ ja $\alpha \in \Gamma^*$, siis $\langle p, xw, \beta\alpha\rangle \underset{M}{\vdash} \langle q, w, \gamma\alpha\rangle$. Binaarne seos $\overset{*}{\vdash}$ on seose $\vdash$ refleksiivne transitiivne kate.

Näide 6.5. Näite 6.2 magasinmäluga automaadi jaoks kehtib $\langle 1, abba, \varepsilon\rangle \overset{*}{\vdash} \langle 1, ba, BA\rangle$ ja $\langle 1, abba, \varepsilon\rangle \overset{*}{\vdash} \langle 2, \varepsilon, \varepsilon\rangle$.

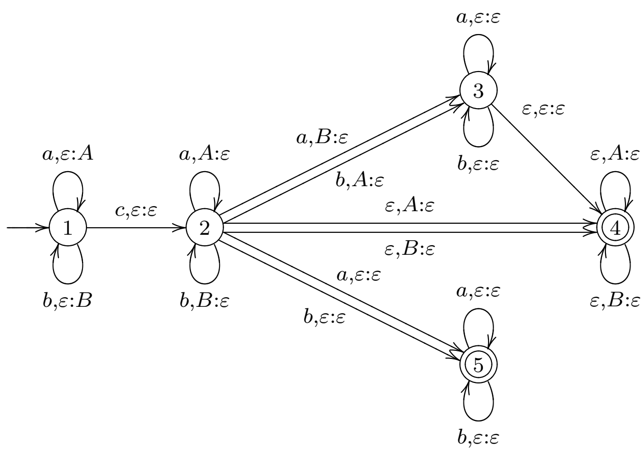
Definitsioon 6.6. Magasinmäluga automaat $M = \langle Q, \Sigma, \Gamma, \Delta, I, F \rangle$ *aktsepteerib* stringi $w \in \Sigma^*$ parajasti siis, kui leiduvad olekud $p \in I$ ja $q \in F$, mille korral $\langle p, w, \varepsilon\rangle \overset{*}{\vdash} \langle q, \varepsilon, \varepsilon\rangle$.

Definitsioon 6.7. Magasinmäluga automaadi M poolt *aktsepteeritav* keel $L(M)$ koosneb selle automaadi poolt aktsepteeritavatest stringidest.

Näide 6.8. Vaatleme magasinmäluga automaati M näitest 6.2. On lihtne näha, et $L(M) = \{uu^{\mathrm{R}} \mid u \in \{a, b\}^*\}$.

Näide 6.9. Olgu $M = \langle Q, \Sigma, \Gamma, \Delta, I, F \rangle$, kus $Q = \{1,2,3,4,5\}$, $\Sigma = \{a,b,c\}$, $\Gamma = \{A,B\}$, $I = \{1\}$, $F = \{4,5\}$ ja

$$\begin{aligned}\Delta = \{&\langle\langle 1,a,\varepsilon\rangle,\langle 1,A\rangle\rangle,\ \langle\langle 1,b,\varepsilon\rangle,\langle 1,B\rangle\rangle,\\ &\langle\langle 1,c,\varepsilon\rangle,\langle 2,\varepsilon\rangle\rangle,\\ &\langle\langle 2,a,A\rangle,\langle 2,\varepsilon\rangle\rangle,\ \langle\langle 2,b,B\rangle,\langle 2,\varepsilon\rangle\rangle,\\ &\langle\langle 2,a,B\rangle,\langle 3,\varepsilon\rangle\rangle,\ \langle\langle 2,b,A\rangle,\langle 3,\varepsilon\rangle\rangle,\\ &\langle\langle 2,\varepsilon,A\rangle,\langle 4,\varepsilon\rangle\rangle,\ \langle\langle 2,\varepsilon,B\rangle,\langle 4,\varepsilon\rangle\rangle,\\ &\langle\langle 2,b,\varepsilon\rangle,\langle 5,\varepsilon\rangle\rangle,\ \langle\langle 2,a,\varepsilon\rangle,\langle 5,\varepsilon\rangle\rangle,\\ &\langle\langle 3,a,\varepsilon\rangle,\langle 3,\varepsilon\rangle\rangle,\ \langle\langle 3,b,\varepsilon\rangle,\langle 3,\varepsilon\rangle\rangle,\\ &\langle\langle 3,\varepsilon,\varepsilon\rangle,\langle 4,\varepsilon\rangle\rangle,\\ &\langle\langle 4,\varepsilon,A\rangle,\langle 4,\varepsilon\rangle\rangle,\ \langle\langle 4,\varepsilon,B\rangle,\langle 4,\varepsilon\rangle\rangle,\\ &\langle\langle 5,a,\varepsilon\rangle,\langle 5,\varepsilon\rangle\rangle,\ \langle\langle 5,b,\varepsilon\rangle,\langle 5,\varepsilon\rangle\rangle\}.\end{aligned}$$



Siis $L(M) = \{ucv \mid u \in \{a,b\}^*,\ v \in \{a,b\}^*,\ u \neq v^{\mathrm{R}}\}$.

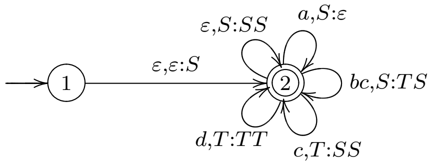
Teoreem 6.10. *Kui keel on kontekstivaba, siis leidub seda keelt aktsepteeriv magasinmäluga automaat.*

Näide 6.11. Olgu $\Sigma = \{a,b,c,d\}$. Kontekstivaba grammatika

$$\begin{aligned}&S \to SS,\\ &S \to a,\\ &S \to bcTS,\\ &T \to cSS,\\ &T \to dTT\end{aligned}$$

ja magasinmäluga automaat $\langle\{1,2\},\Sigma,\{S,T\},\Delta,\{1\},\{2\}\rangle$, kus

$$\begin{aligned}\Delta = \{&\langle\langle 1,\varepsilon,\varepsilon\rangle,\langle 2,S\rangle\rangle,\ \langle\langle 2,\varepsilon,S\rangle,\langle 2,SS\rangle\rangle,\ \langle\langle 2,a,S\rangle,\langle 2,\varepsilon\rangle\rangle,\\ &\langle\langle 2,bc,S\rangle,\langle 2,TS\rangle\rangle,\ \langle\langle 2,c,T\rangle,\langle 2,SS\rangle\rangle,\ \langle\langle 2,d,T\rangle,\langle 2,TT\rangle\rangle\},\end{aligned}$$



esitavad ühe ja sama keele.

Lemma 6.12. *Iga magasinmäluga automaadi jaoks leidub ekvivalentne magasinmäluga automaat,* $\langle Q, \Sigma, \Gamma, \Delta, I, F\rangle$*, kus* $|I| = 1$*,* $|F| = 1$ *ja iga ülemineku* $\langle\langle p, x, \beta\rangle, \langle q, \gamma\rangle\rangle \in \Delta$ *korral kehtib* $|x| \leqslant 1$ *ja* $|\beta| + |\gamma| = 1$*.*

Teoreem 6.13. *Iga magasinmäluga automaadi jaoks leidub sama keelt genereeriv kontekstivaba grammatika.*

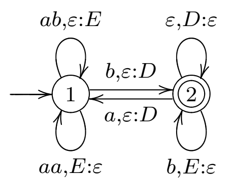
Tõestus. Lemma 6.12 alusel võime eeldada, et on antud magasinmäluga automaat $\langle Q, \Sigma, \Gamma, \Delta, I, F\rangle$, kus $I = \{q_{\mathrm{s}}\}$, $F = \{q_{\mathrm{a}}\}$ ja iga ülemineku $\langle\langle p, x, \beta\rangle, \langle q, \gamma\rangle\rangle \in \Delta$ korral kehtib võrdus $|\beta| + |\gamma| = 1$. Koostame kontekstivaba grammatika $\langle N, \Sigma, P, S\rangle$, kus $N = \{A_{p,q} \mid p \in Q,\ q \in Q\}$, $S = A_{q_{\mathrm{s}},q_{\mathrm{a}}}$ ja

$$\begin{aligned} P = {} & \{A_{p,p} \to \varepsilon \mid p \in Q\} \cup \\ & \cup \{A_{p,t} \to xA_{q,r}yA_{s,t} \mid \langle\langle p, x, \varepsilon\rangle, \langle q, C\rangle\rangle \in \Delta,\ \langle\langle r, y, C\rangle, \langle s, \varepsilon\rangle\rangle \in \Delta,\ C \in \Gamma,\ t \in Q\}. \end{aligned}$$

Induktsiooni abil võib tõestada, et $\langle p, x, \varepsilon\rangle \overset{*}{\vdash} \langle q, \varepsilon, \varepsilon\rangle$ parajasti siis, kui $A_{p,q} \overset{*}{\Rightarrow} x$ (siin $x \in \Sigma^*$). □

Näide 6.14. Magasinmäluga automaat $M = \langle\{1, 2\}, \Sigma, \Gamma, \Delta, \{1\}, \{2\}\rangle$, kus $\Sigma = \{a, b\}$, $\Gamma = \{D, E\}$,

$$\begin{aligned} \Delta = \{ & \langle\langle 1, ab, \varepsilon\rangle, \langle 1, E\rangle\rangle,\ \langle\langle 1, aa, E\rangle, \langle 1, \varepsilon\rangle\rangle,\ \langle\langle 1, b, \varepsilon\rangle, \langle 2, D\rangle\rangle, \\ & \langle\langle 2, a, \varepsilon\rangle, \langle 1, D\rangle\rangle,\ \langle\langle 2, \varepsilon, D\rangle, \langle 2, \varepsilon\rangle\rangle,\ \langle\langle 2, b, E\rangle, \langle 2, \varepsilon\rangle\rangle\}, \end{aligned}$$



ja kontekstivaba grammatika

$$\begin{aligned} S &\to abTaaS, & T &\to abTaaT, \\ S &\to abSbU, & T &\to \varepsilon, \\ S &\to bUU, & U &\to aSU, \\ & & U &\to \varepsilon \end{aligned}$$

esitavad ühe ja sama keele. Sümbolid S, T ja U vastavad sümbolitele $A_{1,2}$, $A_{1,1}$ ja $A_{2,2}$ teoreemi 6.13 tõestusest.

Ülesanne 6.15. Koostada kontekstivaba grammatika, mis genereerib keele $\{w \in \{a, b\}^* \mid |w|_a = |w|_b\}$.

Ülesanne 6.16. Koostada kontekstivaba grammatika, mis genereerib keele $\{w \in \{a, b\}^* \mid |w|_a \geqslant |w|_b\}$.

Ülesanne 6.17. Koostada kontekstivaba grammatika, mis genereerib keele $\{w \in \{a, b\}^* \mid |w|_a = 2|w|_b\}$.

Peatükk 7. Determineeritud kontekstivabad keeled

Definitsioon 7.1. Magasinmäluga automaadi üleminekuid $\langle\langle p_1, x_1, \beta_1\rangle, \langle q_1, \gamma_1\rangle\rangle$ ja $\langle\langle p_2, x_2, \beta_2\rangle, \langle q_2, \gamma_2\rangle\rangle$ nimetatakse *konfliktseteks*, kui

1) $p_1 = p_2$;

2) $x_1 \sqsubset x_2$ või $x_2 \sqsubset x_1$;

3) $\beta_1 \sqsubset \beta_2$ või $\beta_2 \sqsubset \beta_1$.

Definitsioon 7.2. Magasinmäluga automaat on *determineeritud* (*deterministic*), kui algolekuid on täpselt üks ja kõik üleminekud on omavahel mittekonfliktsed.

Definitsioon 7.3. Keel L tähestikus Σ on *determineeritud kontekstivaba keel*, kui leidub determineeritud magasinmäluga automaat, mis aktsepteerib keele $L \cdot \{\$\}$ tähestikus $\Sigma \cup \{\$\}$, kus **\$** on uus sümbol, mis ei kuulu tähestikku Σ. Sümbolit **\$** nimetatakse *stringi lõpu markeriks*.

Näide 7.4. Olgu $\Sigma = \{a, b\}$. Keel $L = \{a^m b^n \mid m = n \text{ või } n = 0\}$ on determineeritud kontekstivaba keel.

Teoreem 7.5. *Iga regulaarne keel on determineeritud kontekstivaba keel.*

Teoreem 7.6. *Iga determineeritud kontekstivaba keele jaoks leidub seda keelt genereeriv ühene kontekstivaba grammatika.*

Teoreem 7.7. *Iga determineeritud kontekstivaba keele täiend on determineeritud kontekstivaba keel.*

Teoreem 7.8. *Leiduvad determineeritud kontekstivabad keeled L_1 ja L_2, mille korral $L_1 \cap L_2$ ei ole determineeritud kontekstivaba keel.*

Teoreem 7.9. *Leiduvad determineeritud kontekstivabad keeled L_1 ja L_2, mille korral $L_1 \cup L_2$ ei ole determineeritud kontekstivaba keel.*

Peatükk 8. Algoritmilised probleemid formaalsete keelte teoorias

Teoreem 8.1. *Keel on rekursiivselt loenduv parajasti siis, kui leidub seda keelt genereeriv 0-tüüpi grammatika.*

Ülesanne 8.2. Kas grammatika

$$
\begin{array}{ll}
S \to Sb, & Ta \to aaT, \\
S \to ST, & Tb \to ab, \\
S \to SU, & Uaaa \to aaU, \\
S \to \varepsilon, & Uab \to b
\end{array}
$$

aktsepteerib stringid aab ja $aaaaaaaab$?

Ülesanne 8.3. Kas grammatika

$$
\begin{array}{lll}
S \to cS, & ac \to ca, & cdc \to cdcR, \\
S \to dS, & ca \to ac, & Rcca \to ccR, \\
S \to aa, & bc \to cb, & Rddb \to ddR, \\
 & cb \to bc, & Rcdd \to cddT, \\
 & ad \to da, & Tcc \to ccaT, \\
 & da \to ad, & Tdd \to ddbT, \\
 & bd \to db, & Tcdc \to cdc \\
 & db \to bd, &
\end{array}
$$

poolt genereeritav keel on lahenduv?

Lemma 8.4. *Grammatika*

$$
\begin{array}{ll}
R \to aaa, & ce \to eca, \\
ac \to ca, & eca \to ce, \\
ca \to ac, & de \to edb, \\
bc \to cb, & eda \to de, \\
cb \to bc, & aaa \to caaa, \\
ad \to da, & caaa \to aaa, \\
da \to ad, & aaa \to daaa, \\
bd \to db, & daaa \to aaa, \\
db \to bd, & cdca \to cdcae, \\
 & cdcae \to cdca
\end{array}
$$

genereerib mittelahenduva keele.

Lemma 8.5. *Grammatika*

$$\begin{aligned} R &\to aRe, & fgggff &\to gfffgg,\\ R &\to bRf, & gfffgg &\to fgggff,\\ R &\to cRg, & ffggffffgg &\to ffggffffgge,\\ R &\to d, & ffggffffgge &\to ffggffffgg,\\ ggff &\to ffgg, & eggff &\to ffe,\\ ffgg &\to ggff, & ffe &\to eggff,\\ & & efgggff &\to gffe,\\ & & gffe &\to efgggff \end{aligned}$$

genereerib mittelahenduva keele.

Definitsioon 8.6. *Posti vastavuse süsteem* tähestikus Σ on stringide korteežide paar $((x_1, \ldots, x_n), (y_1, \ldots, y_n))$, kus iga i korral $x_i \in \Sigma^*$ ja $y_i \in \Sigma^*$.

Märkus 8.7. Posti vastavuse süsteemi $((x_1, \ldots, x_n), (y_1, \ldots, y_n))$ kujutatakse tavaliselt nii:

$$\left[\frac{x_1}{y_1}\right], \ldots, \left[\frac{x_n}{y_n}\right].$$

Definitsioon 8.8. Posti vastavuse süsteemi $((x_1, \ldots, x_n), (y_1, \ldots, y_n))$ *lahendus* (*match*) on mittetühi indeksite jada $(i_1, \ldots, i_k)$, mille korral

$$x_{i_1} \ldots x_{i_k} = y_{i_1} \ldots y_{i_k}$$

(siin iga j korral kehtib $1 \leqslant i_j \leqslant n$).

Näide 8.9. Olgu $\Sigma = \{a, b, c\}$. Vaatleme Posti vastavuse süsteemi

$$\left[\frac{aab}{a}\right], \left[\frac{a}{aa}\right], \left[\frac{caa}{bc}\right].$$

Jada $(2, 1, 3, 2, 2)$ on selle süsteemi lahenduseks, kuna

$$a \cdot aab \cdot caa \cdot a \cdot a = aa \cdot a \cdot bc \cdot aa \cdot aa.$$

Ülesanne 8.10. Olgu $\Sigma = \{a, b, c, 0, 1, \diamondsuit, \heartsuit, \spadesuit\}$. Kas Posti vastavuse süsteemil

$$\left[\frac{a\spadesuit}{a\spadesuit aba\heartsuit 1aca\spadesuit a}\right], \left[\frac{ab}{ba}\right], \left[\frac{ac}{ca}\right], \left[\frac{a\spadesuit}{\spadesuit a}\right], \left[\frac{a\spadesuit}{ba\spadesuit a}\right], \left[\frac{a\spadesuit}{\spadesuit aba}\right],$$
$$\left[\frac{a\heartsuit 1ab}{\diamondsuit a}\right], \left[\frac{a\heartsuit 1ac}{ca\heartsuit 10a}\right], \left[\frac{aba\heartsuit 10ab}{\heartsuit 10abaca}\right], \left[\frac{aca\heartsuit 10ab}{\heartsuit 10acaca}\right], \left[\frac{a\heartsuit 10ac}{ca\heartsuit 11a}\right], \left[\frac{a\heartsuit 11ac}{ca\heartsuit 1a}\right],$$
$$\left[\frac{a\diamondsuit ab}{\diamondsuit a}\right], \left[\frac{a\diamondsuit ac}{\diamondsuit a}\right], \left[\frac{aba\diamondsuit}{\diamondsuit a}\right], \left[\frac{aca\diamondsuit}{\diamondsuit a}\right], \left[\frac{a\diamondsuit a\spadesuit a\spadesuit}{\spadesuit}\right]$$

on lahendus?

Teoreem 8.11. *Vaatleme Posti vastavuse süsteeme tähestikus Σ, kus $|\Sigma| \geqslant 2$. Pole olemas algoritmi, mis iga sellise Posti vastavuse süsteemi kohta teeks kindlaks, kas süsteemil lahendus on.*

Teoreem 8.12. *On olemas algoritm, mis iga kontekstitundliku grammatika ja iga stringi kohta teeb kindlaks, kas antud string kuulub antud grammatika poolt genereeritavasse keelde.*

Teoreem 8.13. *On olemas algoritm, mis iga kontekstivaba grammatika ja iga stringi kohta teeb kindlaks, kas antud string kuulub antud grammatika poolt genereeritavasse keelde.*

Teoreem 8.14. *On olemas algoritm, mis iga kontekstivaba grammatika G kohta teeb kindlaks, kas $L(G) \neq \varnothing$.*

Teoreem 8.15. *On olemas algoritm, mis iga kontekstivaba grammatika G kohta teeb kindlaks, kas keel $L(G)$ on lõpmatu.*

Teoreem 8.16. *On olemas algoritm, mis suvaliste paremlineaarsete grammatikate G_1 ja G_2 kohta teeb kindlaks, kas $L(G_1) \subseteq L(G_2)$.*

Teoreem 8.17. *On olemas algoritm, mis suvaliste paremlineaarsete grammatikate G_1 ja G_2 kohta teeb kindlaks, kas $L(G_1) = L(G_2)$. Seega paremlineaarsete grammatikate ekvivalentsiprobleem on algoritmiliselt lahenduv.*

Teoreem 8.18. *Iteratsioonita regulaaravaldiste mittevõrdsuse probleem on* NP*-täielik.*

Definitsioon 8.19. Kasutame tähestikku $\Sigma_3 = \{a, b, c\}$. Olgu $\vec{x} = (x_1, \ldots, x_n)$, kus iga i korral $x_i \in \{a, b\}^*$. Tähistagu $G(\vec{x})$ lineaarset grammatikat $\langle \{S\}, \Sigma_3, P, S\rangle$, kus

$$P = \{S \to ba^i S x_i \mid 1 \leqslant i \leqslant n\} \cup \{S \to ba^i c x_i \mid 1 \leqslant i \leqslant n\}.$$

Grammatika $G(\vec{x})$ poolt genereeritavat keelt tähistame $\mathcal{L}_{\vec{x}}$.

Lemma 8.20. *Keel $\mathcal{L}_{\vec{x}} \cap \mathcal{L}_{\vec{y}}$ on mittetühi parajasti siis, kui Posti vastavuse süsteemil $(\vec{x}, \vec{y})$ on lahendus.*

Teoreem 8.21. *Olgu $|\Sigma| \geqslant 2$. Siis pole olemas algoritmi, mis suvaliste tähestikus Σ toimivate kontekstivabade grammatikate G_1 ja G_2 kohta teeks kindlaks, kas $L(G_1) \cap L(G_2) = \varnothing$.*

Tõestus. Kui $|\Sigma| \geqslant 3$, siis kasutame lemmat 8.20 ja teoreemi 8.11.

Kaheelemendilise tähestiku $\Sigma = \{d, e\}$ jaoks saame tõestuse, kui asendame frammatikas $G(\vec{x})$ sümboli a sõnaga ede, sümboli b sõnaga $edde$ ja sümboli c sõnaga $eddde$. □

Lemma 8.22. *Keel $\mathcal{L}_{\vec{x}} \cap \mathcal{L}_{\vec{y}}$ on lõpmatu parajasti siis, kui Posti vastavuse süsteemil $(\vec{x}, \vec{y})$ on lahendus.*

Teoreem 8.23. *Olgu $|\Sigma| \geqslant 2$. Siis pole olemas algoritmi, mis suvaliste tähestikus Σ toimivate kontekstivabade grammatikate G_1 ja G_2 kohta teeks kindlaks, kas keel $L(G_1) \cap L(G_2)$ on lõpmatu.*

Teoreem 8.24. *Olgu $|\Sigma| \geqslant 2$. Siis pole olemas algoritmi, mis suvalise tähestikus Σ toimiva kontekstivaba grammatika G kohta teeks kindlaks, kas G on ühene.*

Lemma 8.25. *Kasutame tähestikku $\Sigma_3 = \{a, b, c\}$. Keel $\Sigma_3^* - \mathcal{L}_{\vec{x}}$ on kontekstivaba iga $\vec{x}$ korral.*

Näide 8.26. Olgu $\vec{x} = (b, abbaa)$. Siis kontekstivaba grammatika

$$\begin{array}{lll}
S \to \varepsilon, & R \to \varepsilon, & T_1 \to U_b, \\
S \to AW, & R \to BW, & T_2 \to U_a bbaa, \\
S \to bR, & R \to aaaW, & T_2 \to U_b baa, \\
A \to a, & R \to a, & T_2 \to U_b aa, \\
A \to c, & R \to aa, & T_2 \to U_a a, \\
B \to b, & R \to abRb, & T_2 \to U_a, \\
B \to c, & R \to aabRabbaa, & U_a \to WB, \\
Z \to a, & R \to acZWb, & U_a \to \varepsilon, \\
Z \to B, & R \to aacZWabbaa, & U_b \to WA, \\
W \to ZW, & R \to aBT_1, & U_b \to \varepsilon \\
W \to \varepsilon, & R \to aaBT_2, &
\end{array}$$

genereerib keele $\Sigma_3^* - \mathcal{L}_{\vec{x}}$.

Definitsioon 8.27. Olgu $\Sigma_3 = \{a, b, c\}$, $\vec{x} = (x_1, \ldots, x_n)$, $\vec{y} = (y_1, \ldots, y_n)$, kus iga i korral $x_i \in \{a, b\}^*$ ja $y_i \in \{a, b\}^*$. Tähistame $\mathcal{M}_{\vec{x},\vec{y}} \leftrightharpoons \Sigma_3^* - (\mathcal{L}_{\vec{x}} \cap \mathcal{L}_{\vec{y}})$.

Lemma 8.28. *Keel $\mathcal{M}_{\vec{x},\vec{y}}$ on kontekstivaba iga $\vec{x}$ ja $\vec{y}$ korral.*

Tõestus. $\mathcal{M}_{\vec{x},\vec{y}} = (\Sigma_3^* - \mathcal{L}_{\vec{x}}) \cup (\Sigma_3^* - \mathcal{L}_{\vec{y}})$. □

Lemma 8.29. *Keele $\mathcal{M}_{\vec{x},\vec{y}}$ täiend tähestiku Σ_3 suhtes on mittetühi parajasti siis, kui Posti vastavuse süsteemil $(\vec{x}, \vec{y})$ on lahendus.*

Teoreem 8.30. *Olgu $|\Sigma| \geqslant 2$. Siis pole olemas algoritmi, mis suvalise tähestikus Σ toimiva kontekstivaba grammatika G kohta teeks kindlaks, kas $L(G) = \Sigma^*$.*

Teoreem 8.31. *Olgu $|\Sigma| \geqslant 2$. Tähestikus Σ toimivate kontekstivabade grammatikate ekvivalentsiprobleem pole algoritmiliselt lahenduv.*

Teoreem 8.32. *Olgu $|\Sigma| \geqslant 2$. Siis pole olemas algoritmi, mis suvaliste tähestikus Σ toimivate kontekstivabade grammatikate G_1 ja G_2 kohta teeks kindlaks, kas $L(G_1) \subseteq L(G_2)$.*

Lemma 8.33. *Keele $\mathcal{M}_{\vec{x},\vec{y}}$ täiend tähestiku Σ suhtes on lõpmatu parajasti siis, kui Posti vastavuse süsteemil $(\vec{x}, \vec{y})$ on lahendus.*

Teoreem 8.34. *Olgu $|\Sigma| \geqslant 2$. Siis pole olemas algoritmi, mis suvalise tähestikus Σ toimiva kontekstivaba grammatika G kohta teeks kindlaks, kas $\Sigma^* - L(G)$ on lõpmatu.*

Lemma 8.35. *Olgu $\vec{x} = (x_1, \ldots, x_n)$ ja $\vec{y} = (y_1, \ldots, y_n)$, kus iga i korral $x_i \in \{a, b\}^*$, $y_i \in \{a, b\}^*$ ja $x_i y_i \neq \varepsilon$. Keel $\mathcal{M}_{\vec{x},\vec{y}}$ on regulaarne parajasti siis, kui Posti vastavuse süsteemil $(\vec{x}, \vec{y})$ pole lahendust.*

Teoreem 8.36. *Olgu $|\Sigma| \geqslant 2$. Siis pole olemas algoritmi, mis suvalise tähestikus Σ toimiva kontekstivaba grammatika G kohta teeks kindlaks, kas keel $L(G)$ on regulaarne.*

Definitsioon 8.37. Olgu $\Sigma_3 = \{a, b, c\}$, $\vec{x} = (x_1, \ldots, x_n)$, $\vec{y} = (y_1, \ldots, y_n)$, kus iga i korral $x_i \in \{a, b\}^*$ ja $y_i \in \{a, b\}^*$. Tähistame $\mathcal{K}_{\vec{x},\vec{y}} \leftrightharpoons \mathcal{L}_{\vec{x}} \cdot \{c\} \cdot \mathcal{L}_{\vec{y}}^{\mathrm{R}}$.

Lemma 8.38. *Keel $\mathcal{K}_{\vec{x},\vec{y}}$ on kontekstivaba iga $\vec{x}$ ja $\vec{y}$ korral.*

Lemma 8.39. *Kasutame tähestikku $\Sigma_3 = \{a, b, c\}$. Olgu $\vec{x} = (x_1, \ldots, x_n)$ ja $\vec{y} = (y_1, \ldots, y_n)$, kus iga i korral $x_i \in \{a, b\}^*$, $y_i \in \{a, b\}^*$ ja $x_i y_i \neq \varepsilon$. Keel $\mathcal{K}_{\vec{x},\vec{y}} \cap \{zcz^{\mathrm{R}} \mid z \in \Sigma_3^*\}$ on kontekstivaba parajasti siis, kui Posti vastavuse süsteemil $(\vec{x}, \vec{y})$ pole lahendust.*

Teoreem 8.40. *Olgu $|\Sigma| \geqslant 2$. Siis pole olemas algoritmi, mis suvaliste tähestikus Σ toimivate kontekstivabade grammatikate G_1 ja G_2 kohta teeks kindlaks, kas keel $L(G_1) \cap L(G_2)$ on kontekstivaba.*

Lemma 8.41. *Kasutame tähestikku $\Sigma_3 = \{a, b, c\}$. Keel $\Sigma_3^* - (\mathcal{K}_{\vec{x},\vec{y}} \cap \{zcz^{\mathrm{R}} \mid z \in \Sigma_3^*\})$ on kontekstivaba iga $\vec{x}$ ja $\vec{y}$ korral.*

Tõestus. Tähistame $L_0 = \{w \in \Sigma_3^* \mid |w|_c = 1\}$. Keele $\Sigma_3^* - (\mathcal{K}_{\vec{x},\vec{y}} \cap \{zcz^{\mathrm{R}} \mid z \in \Sigma_3^*\})$ võib esitada järgmise viie kontekstivaba keele ühendina

$$
\begin{aligned}
L_1 &= \{w \in \Sigma_3^* \mid |w|_c \neq 3\}, \\
L_2 &= \{v_1 c v_2 c v_3 c v_4 \mid v_1, v_2, v_3, v_4 \in \{a, b\}^*,\ v_1 \neq v_4^{\mathrm{R}}\}, \\
L_3 &= \{v_1 c v_2 c v_3 c v_4 \mid v_1, v_2, v_3, v_4 \in \{a, b\}^*,\ v_2 \neq v_3^{\mathrm{R}}\}, \\
L_4 &= ((\Sigma_3^* - \mathcal{L}_{\vec{x}}) \cap L_0) \cdot \{c\} \cdot L_0, \\
L_5 &= L_0 \cdot \{c\} \cdot ((\Sigma_3^* - \mathcal{L}_{\vec{y}})^{\mathrm{R}} \cap L_0).
\end{aligned}
$$

□

Teoreem 8.42. *Olgu $|\Sigma| \geqslant 2$. Siis pole olemas algoritmi, mis suvalise tähestikus Σ toimiva kontekstivaba grammatika G kohta teeks kindlaks, kas keel $\Sigma^* - L(G)$ on kontekstivaba.*

Lemma 8.43. *Kasutame tähestikku $\Sigma_3 = \{a, b, c\}$. Keel $\mathcal{K}_{\vec{x},\vec{y}} \cap \{zcz^{\mathrm{R}} \mid z \in \Sigma_3^*\}$ on kontekstivaba iga $\vec{x}$ ja $\vec{y}$ korral.*

Teoreem 8.44. *Olgu $|\Sigma| \geqslant 2$. Siis pole olemas algoritmi, mis suvalise tähestikus Σ toimiva kontekstitundliku grammatika G kohta teeks kindlaks, kas keel $L(G)$ on kontekstivaba.*

Kirjandus

[1] A. V. Aho, J. D. Ullman. The Theory of Parsing, Translation, and Compiling, Vol. 1, Parsing. Prentice Hall, 1972.

[2] A. V. Aho, J. D. Ullman. The Theory of Parsing, Translation, and Compiling, Vol. 2, Compiling. Prentice-Hall, 1973.

[3] Jaak Henno. Formaalsed keeled, grammatikad ja translaatorid. TTÜ kirjastus, 2006.

[4] Mare Koit. Sissejuhatus arvutuslingvistikasse. 2009.
`http://www.cs.ut.ee/~koit/SAL/`

[5] Merik Meriste. Formaalsed keeled ja kompilaatorite koostamine. 2003.
`http://cs.ttu.ee/materjal/wai4120/`

[6] Peeter Normak. Teoreetiline informaatika. 2008.
`http://www.tlu.ee/~pnormak/TI2008/`

[7] H. R. Lewis, C. H. Papadimitriou. Elements of the Theory of Computation. 2nd ed. Prentice Hall, 1998.

[8] Jaan Penjam. Teoreetiline informaatika. 2001.
`http://www.cs.ioc.ee/~jaan/teorinf/konspekt/ti1a.pdf`

[9] Jaan Penjam. Teoreetiline informaatika. 2008.
`http://www.cs.ioc.ee/~jaan/teorinf/slaidid.html`

[10] M. Sipser. Introduction to the Theory of Computation. PWS Publishing company, 1997.

[11] Varmo Vene. Transleerimismeetodid. 2006.
`http://www.cs.ut.ee/~varmo/TM2006/`

[12] А. Е. Пентус, М. Р. Пентус. Теория формальных языков. Москва: Изд-во ЦПИ при механико-математическом ф-те МГУ, 2004.
`http://lpcs.math.msu.su/~pentus/tfyaw.htm`

[13] А. Е. Пентус, М. Р. Пентус. Математическая теория формальных языков. Москва: ИНТУИТ; БИНОМ. 2006.
`http://lpcs.math.msu.su/~pentus/mtfyaw.htm`