Harjutused 2.

Regulaarsed avaldised.

Meeldetuletus.

Regulaarset avaldist (kasutan edaspidi sageli tekstis RA) võib endale ette kujutada nagu avaldist, mis kirjeldab meile vaadeldava regulaarse keele sõnade üldist kuju. Regulaarses avaldises võivad sisalduda kõik tähed meie keele tähestikust ja 'tehted' (mõtle analoogiale matemaatiliste tehetega aritmeetilises avaldises). Teheteks regulaarses avaldises on:

konkatenatsioon ehk järjest kirjutamine	nt. *aab*
ühend (tähistatakse tavaliselt +)	nt. *a(a+b)*
Kleene'e sulund *	nt. *a*bb*

- tähendus: avaldise osa, mille kohta * käib, võib esineda suvaline arv kordi järjest (0 või enam). Nt a* tähendab seega igasuguse pikkusega järjendit, mis koosneb a-dest (sealhulgas ka järjendit pikkusega 0 ehk tühja sõna). Seega *a*bb* kirjeldab sellist lõpmatut sõnade hulka (keelt) {*bb, abb, aabb, aaabb, aaaabb*, …}
- Kasutan harjutustunnis ka nn.Kleene'e plussi (märgitakse $^+$) – see tähendab, et vastavat avaldise osa võib esineda ÜKS või enam korda järjest (sisuliselt on välja jäetud tühi sõna). Nt a^+bb kirjeldab lõpmatut sõnade hulka (keelt) {*abb, aabb, aaabb, aaaabb*, …}
- Kasutan ka Σ*: kui Σ tähistab meie keele tähestikku, siis Σ* on kõik võimalikud sõnad selles tähestikus. See on tegelikult lühend - kui nt tähestik Σ= {*a,b,c*}, siis Σ* = (*a*+*b*+*c*)* .

Oluline on aru saada sulgude kasutamisest:

a(a+b)a ja *aa+ba*

Mõlemad avaldised kirjeldavad keelt, mis koosneb kahest sõnast. Esimeses on kaks 3tähelist sõna: esimene täht on a, teine täht on kas a või b ja kolmas täht on a. Teine avaldis ütleb, et meie keelde kuuluv sõna on kas aa või ba, ehk siis need kaks sõna ongi kogu meie keel.

a+b** ja *(a+b)**

a+b** tähendab, et iga meie keelde kuuluv sõna koosneb kas suvalisest arvust a-dest või suvalisest arvust b-dest (aga need ei esine läbisegi). Nt kuuluvad keelde *ε, a, aa, aaaaaa, b, bbb, bbbbbbbb* jne, aga mitte ühtegi sõna, kus leiduks nii a-sid kui b-sid.

*(a+b)** tähendab, iga meie keelde kuuluv sõna koosneb suvalisest arvust suvalises järjekorras olevatest a-dest ja b-dest. Miks? Sest * ütleb, et sobivad igasuguse pikkusega sõnad, ja (*a+b*) ütleb, et igale kohale järjendis võime valida kas a või b. Seega kuuluvad keelde *ε, a, aa, aaaaaa, b, bbb, bbbbbbbb*, aga ka *ab, abb, bba, bbbbbba, aaaabba, bbababababababa* jne ehk siis tegelikult kõik sõnad, mida saab a-de ja b-de abil kirja panna.

Viimane märkus. Sama keelt võib sageli kirjeldada mitme erineva RA-ga, seega võib juhtuda, et õigeid vastuseid on mitu.

Ülesanded.

Esiteks vaatame mõningaid samu keeli, millele eelmises tunnis tegime LDAd. Tasub tähele panna vastavaid lahendusi.

1. Olgu Σ = {a,b}. Leida RA, mis vastab keelele

1.1. L = {w| w sisaldab vähemalt 2 a-d}

Lahendus: kui sõna peab sisaldama vähemalt 2 a-d, siis mõtleme, mis võib seal sõnas lisaks neile veel olla.

Esimese a ees võib olla mingi arv(ka 0) b-sid, b*

esimese ja teise a vahel võib olla mingi arv(ka 0) b-sid, b*

ja kui me oleme sõna vasakult paremale lugedes ära näinud, et meie sõnas on tõepoolest 2 a-d, siis teame juba, et see sõna sobib meile, ükskõik kui palju ja millises järjekorras a-sid ja b-sid seal veel tuleb Σ*

Oluline on see, et me tõepoolest need kaks a-d leiame. Seega, kokku

b*ab*aΣ*

Märkus. Tegelikult sobib ka avaldis Σ*aΣ*aΣ*. Mõelge ise, mis on nende kahe vahe.

1.2. L = {w| w sisaldab täpselt 2 a-d}

Lahendus: sama nagu eelmises, ainult et nüüd on olukord selline, niipea kui oleme leidnud sõnas 2 a-d, ei tohi neid enam rohkem olla. Seega

b*ab*ab*

1.3. L = {w| w sisaldab ülimalt 2 a-d}

Lahendus: ‘ülimalt 2’ tähendab ‘0,1 või 2’, seega

b*+b*ab*+b*ab*ab*

1.4. L = {w| w sisaldab paarisarvu a-sid}

Lahendus: meil on vaja RA, kus a-d on kahekaupa. Seejuures ei pea need järjest olema, nende vahel (ja järel) võib olla mingi hulk b-sid. (ab*ab*)

Ja et saada kõikvõimalikke paarisarve, lubame selliseid sõnaosi, kus on 2 a-d, järjest kirjutada kuitahes palju arv kordi. (ab*ab*)*

Lisaks võib selliste sõnade alguses olla suvaline arv b-sid b*

Seega kokku

b*(ab*ab*)*

1.5. L = {w| igale a-le järgneb täpselt 1 või 2 b-d}

Lahendus: sõnaosad, kus igale a-le järgneb täpselt 1 või 2 b-d (ab+abb)

Neid lubame korrata kuitahes palju kordi järjest (ab+abb)*

Lisaks võib sõna alguses, enne kui tuleb esimene a, olla b-sid b*

Seega kokku

b*(ab+abb)*

1.6. L = {w| w algab a-ga ja sisaldab vähemalt 1 b-d}

Lahendus: sõna alguses on a, mis juhtub edasi? Oluline on, et mingil hetkel peab tulema üks b. Esimese a ja esimese b vahel võib olla veel mingi hulk a-sid, aga niipea, kui tuleb b, võib edasi sõna lõpus olla ükskõik mis, ehk siis kokku

$a^+b\Sigma^*$

1.7. L = {w| w ei sisalda alamsõna ‘ab’}

Vastus: b*a* - see hõlmab endas ka kõik sõnad, mis koosnevad ainult a-dest, ja kõik sõnad, mis koosnevad ainult b-dest.

1.8. L = {w| w ei sisalda alamsõna ‘ab’ ega ‘ba’}

Vastus: a*+b*

NB. Nagu ikka, on vigu võimalik teha siin kahel viisil. Saab leida

- avaldise, mis kirjeldab küll õigeid sõnu, aga mitte kõiki õigeid sõnu (s.t. osad õiged sõnad ei mahu avaldise alla)
- avaldise, mis kirjeldab küll õigeid sõnu, aga lisaks ka midagi rohkemat (s.t. sõnu, mis ei kuulu tegelikult keelde).

Sellepärast soovitan enesekontrolliks teha konkreetsete näidetega proovi, kas ikka keel ja avaldis sobivad omavahel.

2. Kas kehtivad võrdused ja miks ($\Sigma = \{a,b\}$):

$$\Sigma^*ab\Sigma^* = b^*a^+b\Sigma^*$$

$$\Sigma^*ab\Sigma^* = b^*ab^+\Sigma^*$$

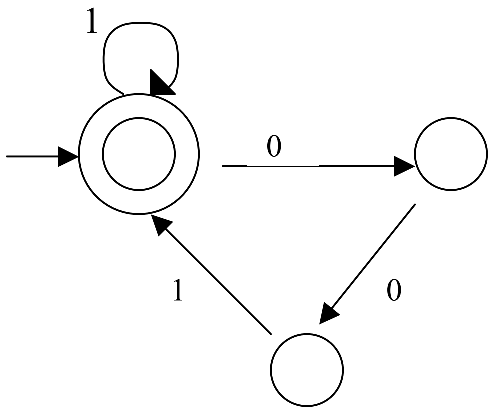
Lahendus. Esimene võrdus kehtib. Vasak pool võrdusest ütleb lihtsalt , et sõnas kuskil leidub alamsõna ab, parem pool võrdusest aga ‘otsib üles esimese koha, kus leidub alamsõna ab’, piiramata seejuures muudmoodi sõnade üldkuju.

Teine võrdus ei kehti. Vasak pool on sama mis esimesel võrdusel. Parem pool nagu otsiks ka üles esimese koha, kus leidub alamsõna ab, aga seejuures on sõnad nii kirjeldatud, et esimene a peab olema kindlasti üksik a. Nt sõna ‘baab’ kuulub võrduse vasakusse poolde, aga mitte võrduse paremasse poolde.

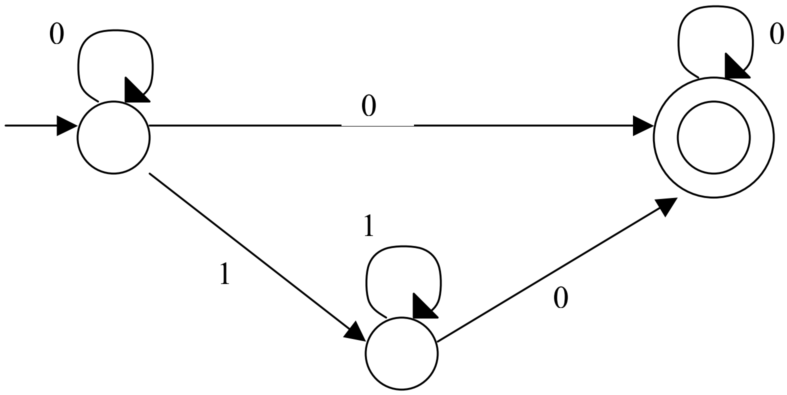
3. Leia LMA, mis vastavad regulaarsetele avaldistele ($\Sigma = \{0,1\}$)

3.1. $1^*(001^+)^*$

Reeglina vastab * mingile tsüklile või silmusele ja mingeid konkreetseid sõnaosi saab fikseerida uutesse seisunditesse minekuga. Seega

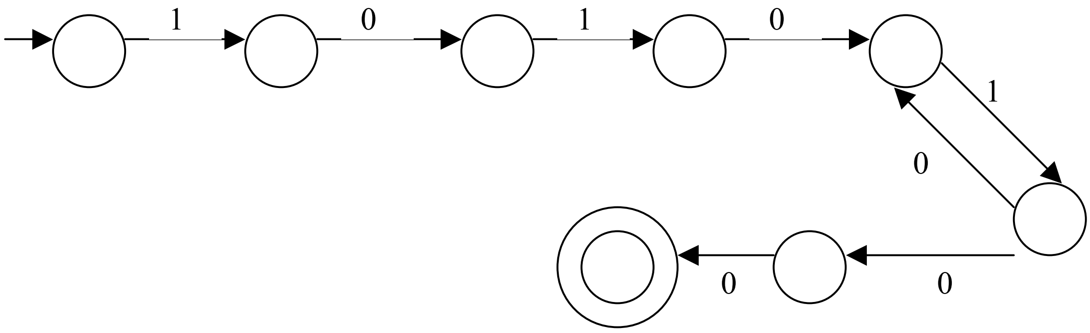


3.2. $0^*1^*0^+$



Lahendus: kui sõnas leidub ühtesid, saame valida alumise tee, kui ainult nullid, siis ülemise tee. Lühim sõna, mis kuulub keelde: 0

3.3. $101(01)^+00$

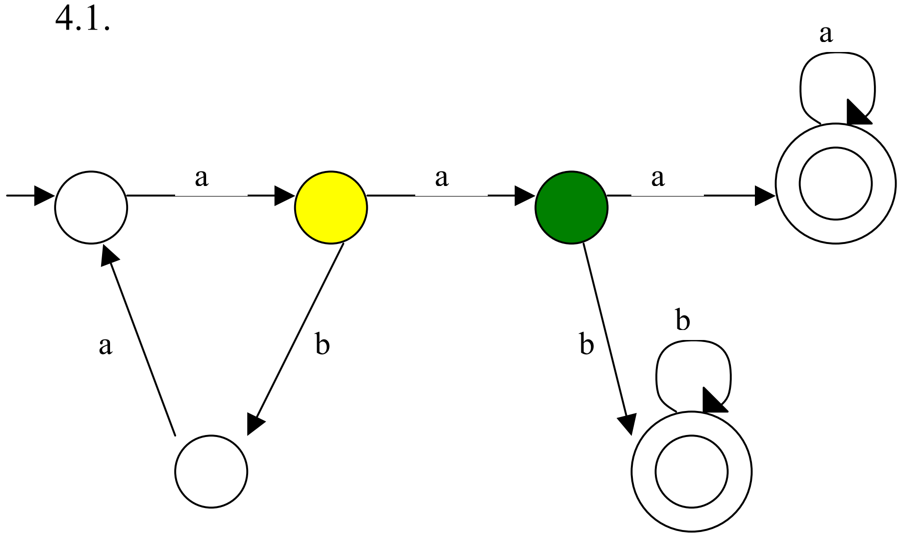


Märkus. Proovi leida sama keelt aktsepteeriv teistsugune lõplik automaat.

NB. Alati tasub kontrollida, kas avaldis ja automaat ikka vastavad üksteisele. Selleks tasub valida erinevaid sõnu, mida RA esindab, ja vaadata, kas need tõepoolest viivad automaadi algseisundist lõppseisundisse. Ja teisipidi, proovida järele erinevaid teid automaadi algseisundist lõppseisundisse ja kontrollida, kas need on ikka hõlmatud RAga.

4. Leia lõplikele automaatidele vastavad regulaaravaldised.

4.1.

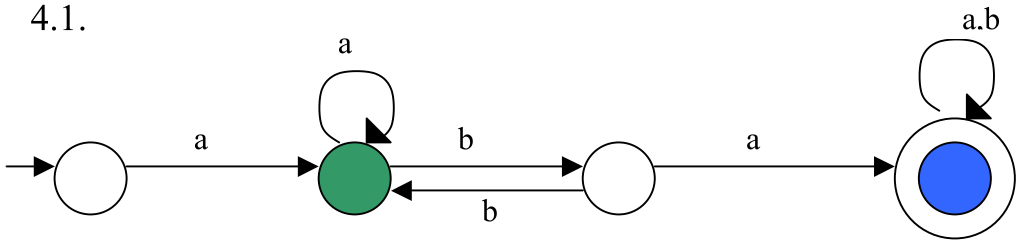


Lahendus: Põhimõtteliselt on meil vaja kirjeldada ära kõik teed algseisundist lõppseisundisse. Kui jõuame kollasesse tippu, on meil 2 võimalust: liikuda mööda tsüklit *baa*, või minna edasi rohelisse tippu.

Seega, selle hetkeni, kui liigume rohelisse tippu, on meie teed kirjeldav RA $a(baa)^*a$ (või samaväärselt $(aba)^*aa$.)

Rohelisest tipust edasi viib kaks teed, $aa^*=a^+$ ja $bb^*=b^+$. Seega kokku saame $a(baa)^*a(a^++b^+)$.

4.1.



Lahendus.Selle automaadi puhul on sõlmkohaks roheline tipp. Nimelt algab sealt 2 tsüklit, a ja bb, mis võivad esineda suvaline arv kordi ja ka vaheldumisi. Nende tsüklite läbimist kirjeldab RA $(a+bb)^*$. Kui tsküklitega on lõpetatud, liigutakse edasi ba ja siis juba, olles sinises tipus, ükskõik kuidas a-de ja b-dega. Seega kokku $a(a+bb)^*ba\Sigma^*$