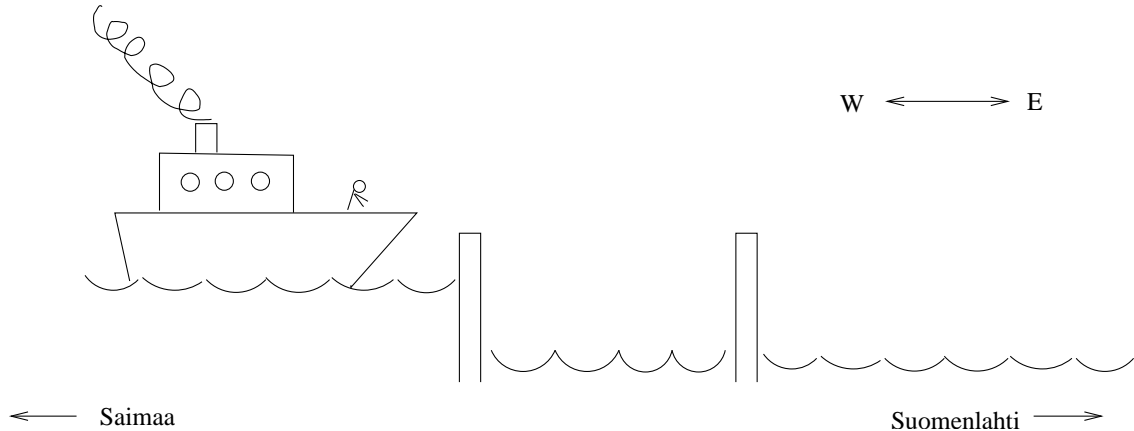


4.9 Harjoituksia säännöllisistä kielistä

1. Simuloikaa annetun äärellisen automaatin toimintaa leikkinä! Kukin opiskelija vastaa yhtä automaatin tilaa, joka lukee seuraavan syötemerkin ja antaa kontrollin seuraajatilalle. Hyväksyvä tai hylkäävä lopputila raportoi tunnistuksen tuloksesta. (Yksi piste jokaiselle leikkiin osallistujalle!)



2. Olkoon aakkosto $\Sigma = \{a, b\}$. Muodosta automaatti, joka hyväksyy seuraavan kielen:
 - a) $L(M) = L(\emptyset)$
 - b) $L(M) = L(\emptyset^*)$
 - c) $L(M) = L(\epsilon)$
 - d) $L(M) = L(\epsilon^*)$
 - e) $L(M) = L(\Sigma^*)$
3. Olkoon aakkosto $\Sigma = \{0, 1\}$. Laadi deterministinen äärellinen automaatti M , joka tunnistaa seuraavan kielen
 - a) $L(M) = \{w \mid w\text{:n pituus on pariton}\}$
 - b) $L(M) = \{w \mid 1\text{:ien lukumäärä } w\text{:ssä on kolmella jaollinen}\}$
 - c) $L(M) = \{w \mid w\text{:ssä on parillinen määrä } 1\text{:iä ja } 0\text{:ia}\}$
4. Laadi kanavan sulkuautomaatti, joka avaa ja sulkee sulut sekä huolehtii veden pinnan laskusta ja nostamisesta automaattisesti. Länsisulun saa avata vain kun vesi on ylhäällä ja itäsulun vain kun vesi on alhaalla. Vettä saa nostaa ja

laskea vain, kun sulkuportit ovat kiinni. Automaatti saa valvontajärjestelmästä seuraavia syötetietoja:

VL = vesi laskettu

VN = vesi nostettu

LL = laiva lännestä

LI = laiva idästä

SK = sulut kiinni

LS = laiva suluissa

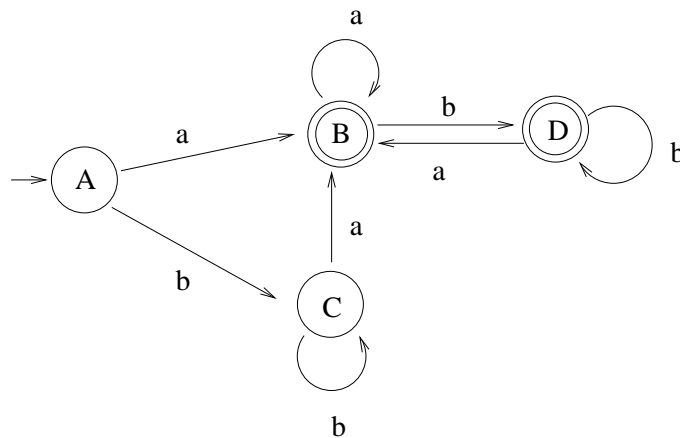
Voit olettaa, että laivojen välillä automaatti odottaa toinen sulkua auki, kunnes saa tiedon uuden laivan saapumisesta.

(Huom! Automaatilla ei siis ole erityisiä alku- ja lopputiloja.)

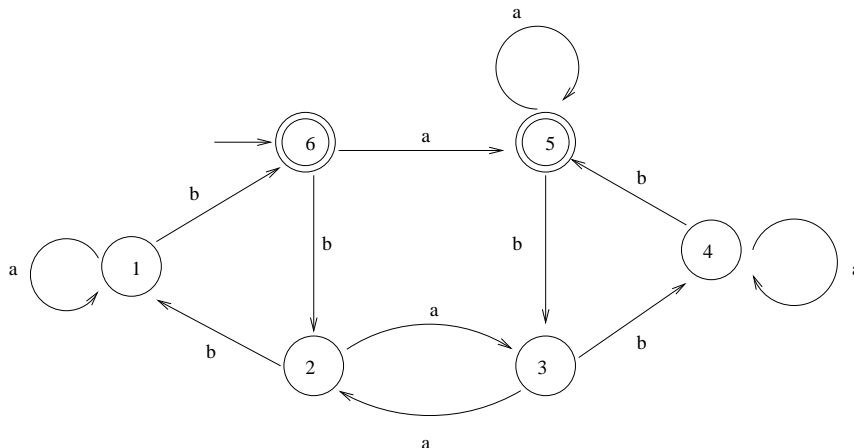
5. Laadi äärellinen automaatti, joka kuvaa kahden kerroksen väliä kulkevan hissinkin toimintaa. Hissi voi olla joko ylhäällä tai alhaalla. Kummastakin kerroksessa on yksinkertainen “tänne”-nappi ja hissinkin sisällä “ylös”- ja “alas”-nappit. Hississä on lisäksi ovi, jonka voi avata tai sulkea; hissi liikkuu vain oven ollessa kiinni. (Automaatilla ei tarvitse olla erityisiä lopputiloja.)

Syöte: napin painallukset, oven avaus ja sulkeminen. Tilat: ylhäällä/alhaalla; ovi auki/kiinni. Merkitään Y = ylhäällä, A = alhaalla, a = ovi auki, k = ovi kiinni.

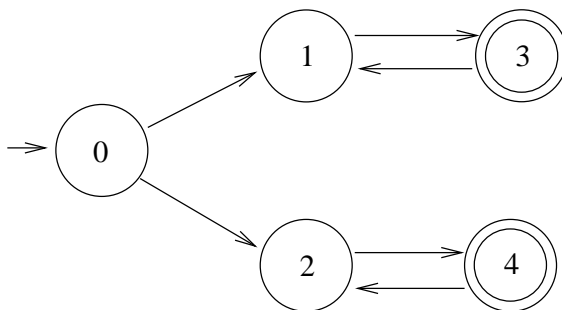
6. Muodosta seuraavaa determinististä äärellistä automaattia vastaava minimi-automaatti:



7. Muodosta seuraavaa determinististä äärellistä automaattia vastaava minimi-automaatti:



8. Tarkastele luentomateriaalista löytyvää sarjakuvaa epädeterministisestä automaatista tohtorin vastaanotolla. Mistä välivaiheista operaatio koostuu? Meneekö kaikki niin kuin pitääkin?
9. Keksi ainakin yksi arkielämän ilmiö, jota voit parhaiten kuvata epädeterministisellä automaatilla! Piirrä automaattisi siirtymäkaavio. Osaatko determinisoida sen?
10. Laadi epädeterministinen äärellinen automaatti, joka testaa, sisältääkö aakkoston $\{a, b\}$ merkkijonoa "abaa" ja determinisoi se.
11. Muunna seuraava automaatti deterministiseksi!



12. Muodosta epädeterministinen äärellinen automaatti, joka hyväksyy seuraavan kielen. Yritä hyödyntää epädeterminismiiä niin paljon kuin mahdollista!
 - a) Aakkoston $\{0, 1, \dots, 9\}$ ne merkkijonot, joiden viimeinen merkki on esiintynyt aiemmin.
 - b) Aakkoston $\{0, 1, \dots, 9\}$ ne merkkijonot, joiden viimeinen merkki *ei* ole esiintynyt aiemmin.

c) Aakkoston $\{0, 1\}$ ne merkkijonot, joissa on kaksi 0:aa, joiden välissä on neljällä jaollinen määrä ykkösiä. Huom! Myös 0 on neljällä jaollinen.

13. UNIX:in egrep-komennolla (extended grep) voi etsiä tekstistä hahmoja, jotka on määritelty säännöllisinä lausekkeina. egrepin perussyntaksi on seuraava: `egrep <lauseke> <tiedosto>`, missä lauseke voi olla

- (a) hakasuluissa lista merkkejä, esim. `[abcd]`: mikä tahansa merkeistä a, b, c, d
- (b) `(lauseke)(lauseke)`: kahden lausekkeen katenaatio
- (c) `(lauseke1)|(lauseke2)`: joko lauseke1 tai lauseke2
- (d) `(lauseke)*`: lauseke toistuu 0 kertaa tai useampia kertoja (sulkeuma)
- (e) `\b`: tyhjä merkki sanan reunassa `\B`: tyhjä merkki sanan keskellä

Huom! Lauseke kannattaa laittaa hipsuihin ('lauseke'). Lisää tietoa komennolla `man egrep`.

Testaa egrep-komentoa tiedostolla

<http://www.cs.joensuu.fi/pages/whamalai/tepe04/esim.c!>

Millaisia hahmoja löydät seuraavalla komennolla?

`egrep '(uu)|(sata)|(issa)|(oita)|(alla)'` esim.c

Keksitkö yksinkertaisempaa komentoa, joka tulostaisi täsmälleen runon rivit?

14. Millaisella egrep-komennolla löydät seuraavat rivit:

- a) Rivit, joilla on numeroita
- b) Rivit, joilla esiintyy sana *while* tai *for*
- c) Rivit, joilla esiintyy numero 10
- d) Rivit, joilla esiintyy kokonaislukuja. (Huom! Komentosi ei siis saa hyväksyä desimaalilukuja.)

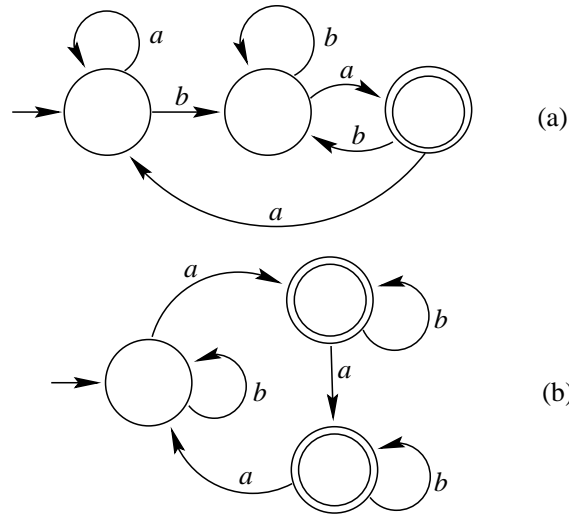
15. Laadi egrep-komento, jolla löydät html-tiedostosta sähköposti- ja katuosoitteet sekä puhelinnumerot. Käytä datana Kissastanian Logiikkakoulun sivuja (<http://cs.joensuu.fi/pages/whamalai/tepe04/kissastania/kissastania.htm!>)!

16. Tarkastellaan seuraavia aakkoston $\Sigma = \{a, b\}$ kieliä. Anna kustakin kielestä kaksi merkkijonoa, jotka kuuluvat kieleen, ja kaksi, jotka eivät kuulu kieleen!

- a) a^*b^*
- b) $a(ba)^*b$
- c) $a^* \cup b^*$

- d) $(aaa)^*$
- e) $(\epsilon \cup a)b$
- f) $\Sigma^*a\Sigma^*a\Sigma^*a\Sigma^*$

17. Mitä merkkijonoja kuuluu seuraavaan lausekkeen kuvaamaan kieleen?
 $(c \cup h \cup m \cup r)at((c \cup t)a \cup (s \cup t)o)ught(m \cup l \cup tw \cup r)ice$
18. Mitä merkkijonoja kuuluu kieleen $L(\emptyset^*)$? Entä $L(\epsilon^*)$?
19. Etsi lyhin merkkijono, joka kuuluu seuraavan lausekkeen kuvaamaan kieleen!
- a) $a^*(b \cup abb)b^*b$
 - b) $a^*b^*b(a \cup (ab)^*)^*b^*$
 - c) $(a \cup ab)(a^* \cup ab)^*b$
20. Muodosta seuraavia kieliä vastaavat säännölliset lausekkeet:
- a) $\{w \in \{a, b\}^* | w:n \text{ kolmanneksi viimeinen merkki on } a\}$
 - b) $\{w \in \{a, b\}^* | w \text{ sisältää joko merkkijonon } ab \text{ tai } ba\}$
 - c) $\{w \in \{a, b\}^* | w \text{ ei sisällä merkkijonoa } bab\}$
21. Muodosta seuraavia kieliä vastaavat säännölliset lausekkeet:
- a) $\{w \in \{a, b\}^* | w \text{ sisältää parillisen määrän merkkiä } a\}$
 - b) $\{w \in \{a, b\}^* | w:n \text{ pituus on pariton}\}$
 - c) $\{w \in \{a, b\}^* | w:n \text{ sisältämien } b\text{-merkkien lukumäärä on kolmella jaollinen}\}$
22. Esitä yksikertaisemmassa muodossa seuraavat lausekkeet! (s.e. ne yhä generoivat saman kielen!)
- a) $(0 \cup 1 \cup 01 \cup 11)^*$
 - b) $(0^* \cup 10^*)^*$
 - c) $1^*(011^*)^* \cup 1^*(011^*)^*0$
23. Poista ϵ -siirtymät Peikkojen Piparkakkuautomaatista ja determinisoi se!
(<http://cs.joensuu.fi/pages/whamalai/tepe04/redhood.pdf>)
24. Millaisen kielen Peikkojen Piparkakkuautomaatti (Kuva 4.5.1) hyväksyy? Esi-
tä kieli säännöllisenä lausekkeena!
25. Muodosta seuraavia äärellisiä automaatteja vastaavat säännölliset lausekkeet:



26. Muodosta seuraavia säännöllisiä lausekkeitä vastaavat deterministiset äärelliset automaattit!

- a) $(ab)^*(ba)^* \cup aa^*$
 b) $((ab \cup abb)^* a^*)^*$

27. Säännöllisen kielen generointi: ”Runoautomaatti”

Toteuta ohjelma, joka tuottaa säännöllistä lauseketta

$(ATTR)^* \text{SUBJ PRED } (ATTRa)^* \text{OBJ} (ATTRIB \cup \epsilon)$

vastaavia rivejä. Kielet ATTR, SUBJ, OBJ, PRED ja ATTRIB koostuvat esim. seuraavista sanoista:

ATTR = {paksu, musta}
 SUBJ = {kissa, kuu, kala}
 OBJ = {kissaa, kuuta, kalaa}
 PRED = {paistaa, katselee, ui}
 ATTRIB = {taivaalla, järvessä}

Ohjelmasi voi siis tuottaa esimerkiksi seuraavanlaisia ”runonsäkeitä”:

Musta kissa syö paksua kalaa järvessä

Kuu katselee taivaalla

Huom! Määrittele, ettei objekti voi seurata kaikkia predikaatteja (esim. verbi ”uida”).

Täydennä esimerkkiä mielestäsi (runoon) sopivilla sanoilla! Lisää huudahduslauseet

”Mikä (ATTR) (SUBJ)!”

sekä kysymykset

”Oletko (ATTR) (SUBJ)?”

28. Osoita, että säännöllisten kielten luokka on suljettu leikkauksen ja katenaation suhteen! Ts. jos L_1 ja L_2 ovat säännöllisiä kieliä, niin myös $L_1 \cap L_2$ ja $L_1 L_2$ ovat säännöllisiä. (Vihje: automaatit.)
29. Edellä havaitsimme, ettei kielen säännöllisyydestä seuraa, että se olisi kahden säännöllisen kielen yhdiste. Pätevätkö muut sulkeumaominaisuudet toiseen suuntaan? Perustele!
30. Ovatko seuraavat kielet säännöllisiä? Perustele!
- $\{w|w \text{ on } a\text{:sta ja } b\text{:stä koostuva merkkijono, jonka pituus on } 3\}$
 - $\{ww|w \in \{a, b\}^*\}$
 - $\{w^*|w\text{:ssä yhtä monta } 1\text{:ä ja } 0\text{:aa}\}$
 - $\{w|w\text{:ssä kaksi kertaa niin monta } 0\text{:aa kuin } 1\text{:ä}\}$
31. Mitä tapahtuu, jos yrität todistaa Pumpsauslemmalla epäsäännölliseksi kielen, joka onkin oikeasti säännöllinen? Tarkastele esimerkiksi kieliä \emptyset , $\{aa, bb\}$, $\{ab^*a^*b\}$.
32. Mitä vikaa on seuraavissa Pumpsauslemmatodistuksissa?
- Olkoon $L = \{(aa)^i(bb)^j|i, j \geq 0\}$.
Väite: L on epäsäännöllinen.
Todistus: Olkoon $x = (aa)^n(bb)^k = a^{2n}b^{2n}$, $|x| = 4n$. Nyt x voidaan jakaa osiin uvw vain yhdellä tavalla: $u = a^i$, $v = a^j$, $j \geq 1$, $w = a^{2n-i-j}b^{2n}$. Nyt $uw^0w = a^{2n-j}b^{2n} \notin L$, joten L on epäsäännöllinen.
 - Olkoon $L = \{c^r a^k b^k|r \geq 1, k \geq 0\} \cup \{a^k b^l|k, l \geq 0\}$.
Väite: L on säännöllinen.
Todistus: $L = L_1 \cup L_2$, missä $L_1 = \{c^r a^k b^k|r \geq 1, k \geq 0\}$ ja $L_2 = \{a^k b^l|k, l \geq 0\}$. Jokaiselle $x \in L$ pätee $x \in L_1$ tai $x \in L_2$. Tutkitaan erikseen kumpaakin tapausta:
1) Jos $x \in L_1$, valitaan $x = c^n a^k b^k$. $|x| = n + 2k > n$. x voidaan jakaa osiin uvw vain yhdellä tavalla:

$u = c^i, v = c^j, j \geq 1, w = c^{n-i-j}a^k b^k$. Nyt $uv^k w \in L$ kaikilla $k = 0, 1, 2, \dots$, eli voimme aina pumpata x :ää.

2) Jos $x \in L_2$, valitaan $x = a^n b^l, |x| = n+l$. x voidaan jakaa osiin vain yhdellä tavalla: $u = a^i, v = a^j, j \geq 1, w = a^{n-i-j} b^l$. $uv^k w \in L$ kaikilla $k = 0, 1, 2, \dots$
 $\Rightarrow L$ on säännöllinen.

33. Tiedetään, että yleisessä tapauksessa ongelma $REG(L)$ (ts. onko annettu kieli L säännöllinen vai ei) on ratkeamaton. Mistä tämä johtuu? Voisitko silti laatia ohjelman, joka auttaisi ihmistä osoittamaan Pumpsauslemmalla, että kieli on epäsäännöllinen?

34. Vahvempi Pumpsauslemman variaatio antaa sekä riittävät että välttämättömät ehdot kielen säännöllisyydelle:

Pumpsauslemma 2: Kieli $A \in \Sigma^*$ on säännöllinen, jos ja vain jos on olemassa vakio $n \geq 1$ s.e. kaikilla $x \in \Sigma^*$, jos $|x| \geq n$ on olemassa u, v, w s.e. $x = uvw$ ja $|v| \geq 1$ ja kaikilla $i \geq 0$ ja kaikilla $y \in \Sigma^*$ $xy \in A$ jos ja vain jos $uv^i wy \in A$.

Voiko tämän pohjalta laatia ohjelman, joka ratkaisee minkä tahansa kielen A osalta, onko A säännöllinen vai ei?

35. Osoita pumpsauslemmalla, että seuraavat kielet eivät ole säännöllisiä:

(a) $\{a^n b^n c^k | n, k = 0, 1, \dots\}$

(b) $\{a^n b^k c^k | n, k = 0, 1, \dots\}$

(c) $\{a^n b^n a^m b^m | n, k = 0, 1, \dots\}$

36. Merkitään w^R :llä merkkijonoa w takaperin kirjoitettuna (so. jos $w = a_1 a_2 \dots a_n$, niin $w^R = a_n \dots a_2 a_1$). Merkkijono on palindromi, jos $w = w^R$ (esimerkiksi "isorikassikakissakiroso"). Tarkastellaan aakkoston $\{a, b\}$ palindromien muodostamaa kieltä $L_{pal} = \{ww^R | w \in \{a, b\}^*\}$

Osoita Pumpsauslemmalla, että palindromikieli L_{pal} on epäsäännöllinen!

37. Laadi ohjelma, joka muuttaa yksinkertaisen HTML-tiedoston vastaavaksi latex-tiedostoksi. HTML-tiedostoa `<html><body> teksti </body></html>` vastaa latex-tiedosto `\documentclass[a4paper,12pt]{article} \begin{document} teksti \end{document}`. Teksti sisältää tavallisen tekstin lisäksi vain otsikkomäärittelyjä. Otsikkomäärittelyjä `<h1>otsikko</h1>`, `<h2>otsikko</h2>`, `<h3>otsikko</h3>` vastaavat latexissa `\section{otsikko}`, `\subsection{otsikko}` ja `\subsubsection{otsikko}`.

Riittääkö äärelliseen automaattiin perustuva sovellus myös HTML-tiedoston oikeellisuuden tarkistamiseen?

38. Mihin käytännön sovelluksiin sinä voisit käyttää äärellisiä automaatteja tai säännöllisiä lausekkeita? Mainitse ainakin kolme sovellusta!
39. Mitä osaat sanoa säännöllisten kielten luokkaan kuuluvien ongelmien aika-vaativuudesta? Entä tilavaativuudesta? (Vihje: Mieti äärellisen automaatin pohjalta laadittua ohjelmaa.) Onko mitään eroa, onko vastaava automaatti deterministinen vai epädeterministinen?