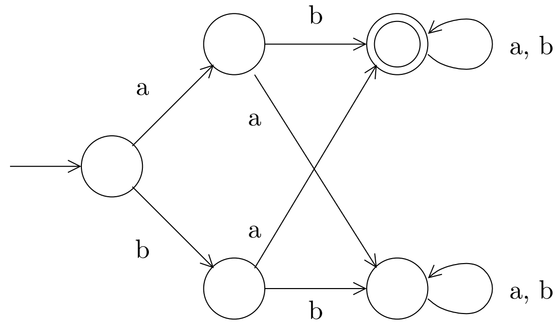
582206 Laskennan mallit (syksy 2010)

1. kurssikoe, ratkaisuja

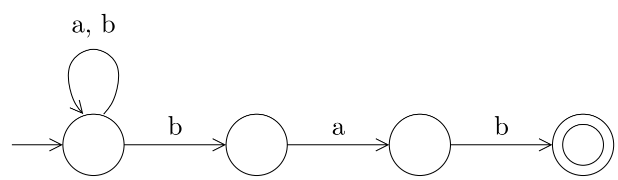
1. [*2+2+2 pistettä*] Säännöllisissä lausekkeissa on käytetty tuttua lyhennysmerkintää $\Sigma = (\mathrm{a} \cup \mathrm{b})$.

   (a) merkkijonot, joiden kaksi ensimmäistä merkkiä ovat joko ”ab” tai ”ba”: deterministinen äärellinen automaatti

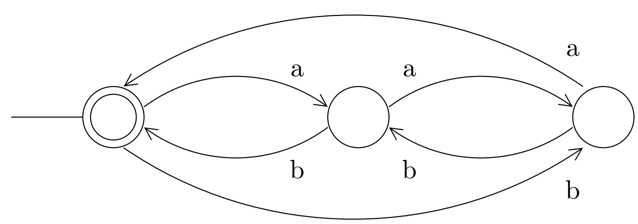
   säännöllinen lauseke $(\mathrm{ab} \cup \mathrm{ba})\Sigma^*$

   (b) merkkijonot, jotka päättyvät ”bab”: epädeterministinen äärellinen automaatti

   säännöllinen lauseke $\Sigma^*\mathrm{bab}$

   (c) merkkijonot, joissa a- ja b-merkkien lukumäärien erotus on tasan kolmella jaollinen.

   Säännöllinen lauseke on ehkä helpoin muodostaa tarkastelemalla automaatin polkuja alkutilasta takaisin alkutilaan käymättä välillä alkutilassa. Hyväksyvässä laskennassa on peräkkäin nolla tai enemmän tällaisia polkuja. Jakamalla ne neljään eri tapaukseen ensimmäisen ja viimeisen merkin perusteella saa lausekkeen

   $$\mathrm{a}(\mathrm{ab})^*\mathrm{b} \cup \mathrm{b}(\mathrm{ba})^*\mathrm{a} \cup \mathrm{a}(\mathrm{ab})^*\mathrm{aa} \cup \mathrm{b}(\mathrm{ba})^*\mathrm{bb}.$$

   Jos ei halua käyttää automaattia apuna, säännöllisen lausekkeen voi muodostaa esim. seuraavasti: Jos w on kieleen kuuluva merkkijono, jaetaan se osajonoihin $w = w_1 \dots w_m$, missä kaikilla i merkkijonossa w_i a-merkkien ja b-merkkien lukumäärien erotus on kolmella jaollinen, mutta millään merkkijonon w_i alkuosalla näin ei ole. Siis merkkijonossa w_i a-merkkien ja b-merkkien lukumäärien erotus on -3, 0 tai 3.

   Tarkastellaan tapausta, että w_i alkaa merkillä a; merkillä b alkava tapaus on symmetrinen.

   - Jos seuraava merkki on b, niin w_i päättyy siihen.
   - Jos seuraavat kaksi merkkiä ovat aa, niin w_i päättyy siihen.

- Jos w_i alkaa aab ja a- ja b-merkkien lukumäärien erotus on 3, sen viimeinen merkki on a. Siis $w_i = \mathrm{aab}w'\mathrm{a}$. Merkkijonon w' on oltava muotoa $\mathrm{a(ba)}^*$, koska muuten lukumäärien erotus menisi nollaan tai kolmoseen ennen w_i:n loppua.
- Jos w_i alkaa aab ja a- ja b-merkkien lukumäärien erotus on 0, sen viimeinen merkki on b. Siis $w_i = \mathrm{aab}w''\mathrm{b}$. Merkkijonon w'' on oltava tyhjä tai muotoa $\mathrm{a(ba)}^*b$, koska muuten lukumäärien erotus menisi nollaan tai kolmoseen ennen w_i:n loppua.

Yhdistämällä edelliset neljä vaihtoehtoa ja toistamalla saadaan säännöllinen lauseke

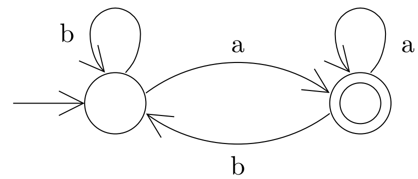
$$(\mathrm{ab} \cup \mathrm{ba} \cup \mathrm{aa(ba)}^*(\mathrm{a} \cup \mathrm{bb}) \cup \mathrm{bb(ab)}^*(\mathrm{b} \cup \mathrm{aa}))^*$$

(Tämäkin lauseke on ehkä helpompi ymmärtää tarkastelemalla automaatin polkuja.)

Arvostelu: Kustakin virheettömästä automaatista ja lausekkeesta saa yhden pisteen. Pikkuvirheestä vähennetään neljännespiste, isommasta puoli. Tyypillisiä pikkuvirheitä olivat (a)-kohdan lausekkeesta unohtuneet sulut ja (b)-kohdan DFA:n (ei siis malliratkaisun NFA:n) virheelliset paluusiirtymät. (c)-kohdassa voi saada neljännespisteen ratkaisusta, joka on selvästi väärin, mutta sisältää oikeita elementtejä.

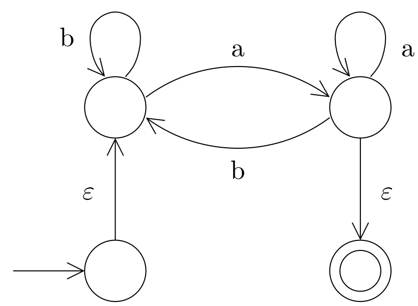
2. [*1+5 pistettä*]

(a) Annettu automaatti:

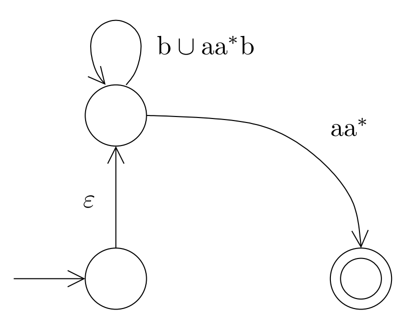


Automaatin tunnistama kieli koostuu kaikista aakkoston { a, b } merkkijonoista, jotka päättyvät merkkiin a. Säännöllinen lauseke: $(\mathrm{a} \cup \mathrm{b})^*\mathrm{a}$.

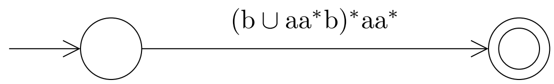
(b) Lisätään alku- ja lopputila:



Poistetaan yksi tila:



Poistetaan toinen tila:



Säännöllinen lauseke on $(\mathrm{b} \cup \mathrm{aa}^*\mathrm{b})^*\mathrm{aa}^*$.

Kommentti: Alkuperäisen automaatin kaksi tilaa voi poistaa kahdessa järjestyksessä. Vaihtoehtoinen poistamisjärjestys tuottaa säännöllisen lausekkeen $\mathrm{b}^*\mathrm{a}(\mathrm{a} \cup \mathrm{bb}^*\mathrm{a})^*$.

Arvostelu: a)-kohdassa annettiin puoli pistettä sanallisesta kuvauksesta sekä puoli pistettä säännöllisestä lausekkeesta, jos ne vastasivat tunnistettavaa kieltä.

b)-kohdassa jaettiin yksi piste kustakin seuraavasta kohdasta:

- Muodostettu GNFA, jossa on uusi alkutila ja uusi lopputila.
- Tilat poistetaan yksi kerrallaan.
- Ensimmäinen poisto oikein.
- Toinen poisto oikein. (riippumaton siitä, oliko ensimmäinen poisto oikein vai ei)
- Lopullinen säännöllinen lauseke vastaa tunnistettavaa kieltä.

Pisteitä vähennettiin myös seuraavin perustein: epäselvä teksti, monitulkintaisuus ja yleinen sekavuus ajatuksenjuoksussa.

Joissakin tapauksissa säännölliseen lausekkeeseen oli päädytty käyttämällä jotakin muuta kuin kirjassa esitettyä menetelmää. Nämä ratkaisuyritykset tuottivat lähes poikkeuksetta virheellisen lopputuloksen, eikä niistä annettu yhtään pisteitä. Vaikka säännöllinen lauseke vastasi tunnistettavaa kieltä, arvioitiin myös, kävikö näin vain sattumalta.

Yleisiä virheitä: menetelman ainoa operaation ei ollut hallussa ja tuotetut lausekkeet eivät vastanneet tunnistettavaa kieltä. Monesti myös todettiin, että kieli päättyy merkkiin a, mikä on virhe, sillä vain kielen sisältämät merkkijonot päättyvät merkkiin a, ei kieli sinänsä. Virheelliset säännölliset lausekkeet kummeksuttavat arvostelijaa, sillä niiden tarkistaminen olisi ollut helppoa.

3. [*6 pistettä*] **Väite:** Aakkoston $\{\,\mathrm{a}, \mathrm{b}\,\}$ kieli

$$A = \{\, \mathrm{a}^m \mathrm{b}^p \mathrm{a}^n \mathrm{b}^q \mid m + n = p + q \,\}$$

ei ole säännöllinen.

Todistus: Tehdään vastaoletus, että kieli A on säännöllinen. Säännöllisyydestä seuraa, että kielellä A on jokin pumppauspituus $k \in \mathbb{N}$. Muodostetaan merkkijono $s = \mathrm{a}^k\mathrm{b}^k\mathrm{a}^1\mathrm{b}^1$, jossa k on mielivaltainen. Koska $s \in A$ ja $|s| \geq k$, niin pumppauspituuden määritelmän mukaan merkkijono s voidaan jakaa sellaisiin osiin $s = xyz$, että

(a) $xy^iz \in A$ kaikilla $i \in \mathbb{N}$

(b) $|y| > 0$

(c) $|xy| \leq k$.

Olkoot x, y ja z tämän mukaiset. Koska $|xy| \leq k$ ja $|y| > 0$, niin kaikki mahdolliset jaot ovat muotoa $x = \mathrm{a}^r$ ja $y = \mathrm{a}^t$, missä $r + t \leq k$ ja $t > 0$. Tällöin $z = \mathrm{a}^{k-r-t}\mathrm{b}^k\mathrm{ab}$. Valitsemalla $i = 2$ saadaan $xy^iz = xy^2z = \mathrm{a}^r\mathrm{a}^{2t}\mathrm{a}^{k-r-t}\mathrm{b}^k\mathrm{ab} = \mathrm{a}^{k+t}\mathrm{b}^k\mathrm{ab} \notin A$, mikä on ristiriita pumppausominaisuuden kohdan (a) kanssa. Siis vastaoletus säännöllisyydestä on virheellinen eli kieli A ei ole säännöllinen. $\square$

Kommentti: Oletetaan, että kielen määrittelyn muuttujille pätee $m, p, n, q \in \mathbb{N}$. Koska muuttuja p on jo käytössä kielen A määrittelyssä, malliratkaisu käyttää pumppauspituudelle symbolia k sekaannusten välttämiseksi.

Valittu merkkijono s ei ole ainoa toimiva valinta, esimerkiksi $\mathrm{a}^k\mathrm{b}^k\mathrm{a}^k\mathrm{b}^k$ ja $\mathrm{a}^k\mathrm{b}^k\mathrm{a}^0\mathrm{b}^0$ toimivat myös.

Arvostelu:

Seuraavista täytetyistä kohdista sai yhden pisteen kustakin.

- Maininta vastaoletuksesta: kieli A on säännöllinen. Tästä seuraa pumppauspituus k.
- Merkkijono s valitaan niin, että ehdot $s \in A$ ja $|s| \geq k$ pätevät.
- Valittu merkkijono s kelpaa onnistuneeseen todistukseen.
- Merkkijonon s jako osiin $s = xyz$ on tehty ehtojen mukaisesti $|y| > 0$ ja $|xy| \leq k$.
- Jakoon kuuluvan merkkijonon y muotoilu mielivaltaisille ehdot toteuttaville jaoille $s = xyz$.
- Vastaesimerkin tuottaminen ja loppupäätelmä kielen A säännöllisyydestä.

Täysiin pisteisiin riitti myös vastaus, jossa osoitettiin, että mikään äärellinen automaatti ei kykene tunnistamaan kieltä A. Tädellinen todistus perustui kielen A tunnistavan äärellisen automaatin rakenteeseen, josta johdettiin ristiriita sille, kuuluvatko tietyt merkkijonot kieleen A. Tämä tuli osoittaa mielivaltaiselle tilojen määrälle ℓ, jolloin kielen A tunnistavaa äärellista automaattia ei ole olemassa. Vastaukset, joissa oli todettu vain tämän todistuksen intuitio, olivat yhden pisteen arvoisia. Myös muut pelkkään intuitioon nojaavat todistusyritykset saivat yhden pisteen.

Lisäksi pisteitä vähennettiin seuraavin perustein: epäselvä matemaattinen notaatio, sanallisen argumentoinnin puute ja ajatuksenjuoksun epäjohdonmukaisuus.

Pumppauslemman tulkitseminen kahden pelaajan K ja E pelinä hyväksyttiin tämän tentin yhteydessä, joskin tehtävän arvostelija suosittelee pumppauslemman käyttämistä jatkossa sellaisenaan. Symbolin p käyttämisestä pumppauspituudelle ei tällä kertaa sakotettu, vaikka symbolia p oli jo käytetty toisessa merkityksessä kielen A määrittelyssä.

Yleisiä virheitä: merkkijonoa s ei osattu muodostaa, kielen määrittelyn muuttujat m, p, n, q oli sotkettu mukaan merkkijonon s muodostukseen, tarkasteltiin vain yhtä jakoa $s = xyz$ kaikkien mahdollisten jakojen sijaan, sekavuus todistusten esitysmuodossa, ei-formaalit käsienheiluttelargumentit ja ennen kaikkea epäjohdonmukaiset ajatusketjut ilman perusteluita.

4. [*2+4 pistettä*] Määritellään, että kieli A on *symmetrinen*, jos $A^{\mathcal{R}} = A$.

   (a) Kaikki symmetriset kielet eivät ole säännöllisiä. Esim. kieli $\{\, 0^n 1 0^n \mid n \in \mathbb{N} \,\}$ tiedetään ei-säännölliseksi, mutta se on selvästi symmetrinen.

   Kaikki säännölliset kielet eivät ole symmetrisiä. Esim. kieli $\{\, 01 \,\}$ on säännöllinen, mutta ei symmetrinen.

   (b) **Väite:** Symmetristen kielten luokka on suljettu yhdisteen suhteen.

   Todistus: Olkoot A ja B symmetrisiä. Nyt

$$
\begin{aligned}
w \in (A \cup B)^{\mathcal{R}} \quad &\Leftrightarrow \quad w^{\mathcal{R}} \in A \cup B \\
&\Leftrightarrow \quad w^{\mathcal{R}} \in A \text{ tai } w^{\mathcal{R}} \in B \\
&\Leftrightarrow \quad w \in A^{\mathcal{R}} \text{ tai } w \in B^{\mathcal{R}} \\
&\Leftrightarrow \quad w \in A \text{ tai } w \in B \\
&\Leftrightarrow \quad w \in A \cup B,
\end{aligned}
$$

   missä askel 4 käytti hyväksi kielten A ja B symmetrisyyttä. Siis $(A \cup B)^{\mathcal{R}} = A \cup B$ eli $A \cup B$ on symmetrinen. $\square$

   Sen sijaan symmetristen kielten luokka ei ole suljettu konkatenaation suhteen. Esim. kielet $A = \{\, 0 \,\}$ ja $B = \{\, 1 \,\}$ ovat symmetrisiä, mutta niiden konkatenaatio $A \circ B = \{\, 01 \,\}$ ei ole.

Arvostelu: Kustakin oikeasta vastauksesta ilman perusteluja saa puoli pistettä. (a)-kohdassa kustakin pätevästä vastaesimerkistä saa puoli pistettä lisää. (b)-kohdassa konkatenaation vastaesimerkistä saa pisteen ja yhdisteen virheettömästä todistuksesta kaksi pistettä. Pienestä virheestä tai epätarkkuudesta vähennetään neljännespiste. Tyypillinen pieni epätarkkuus on esimerkiksi ilmaista (a)-kohdan ensimmäinen vastaesimerkkikieli muodossa 0^n10^n, joka voitaisiin myös tulkita kieleksi, joka sisältää yhden merkkijonon.

Tyypillinen isompi virhe on ymmärtää symmetrinen kieli kieleksi, jonka kaikki merkkijonot ovat palindromeja. Näinhän ei ole. Esimerkiksi kieli $A = \{\mathrm{ab}, \mathrm{ba}\}$ ja tehtävän 1(c) kieli ovat symmetrisiä. Tästä ei ole sakotettu, jos annettu vastaesimerkki on pätevä myös oikean tulkinnan mukaan, mutta (b)-kohdan yhdistetodistuksessa tästä on vähennetty piste.