582206 Laskennan mallit (syksy 2010)

2. kurssikoe, ratkaisuja

Tehtävän 1 tarkasti Juha Kärkkäinen, tehtävän 2 Jyrki Kivinen ja tehtävän 3 Esa Junttila.

1. (a)

$$\begin{aligned} S &\rightarrow 0S1 \mid UV \\ U &\rightarrow 1U \mid \varepsilon \\ V &\rightarrow 0V \mid \varepsilon \end{aligned}$$

Arvostelu:

Tehtävässä on sallittu sekä tulkinta $0 \in \mathbb{N}$ että $0 \notin \mathbb{N}$. Välimuodosta, jossa esimerkiksi m voi olla nolla mutta k ja n eivät, on vähennetty puoli pistettä. Yleisesti pikkuvirheistä on vähennetty neljännespiste tai puoli pistettä ja isommasta virheestä piste.

(b)

$$\begin{aligned} S &\rightarrow S \cup T \mid T \\ T &\rightarrow T \circ F \mid F \\ F &\rightarrow U^* \mid U \\ U &\rightarrow 0 \mid 1 \mid (S) \end{aligned}$$

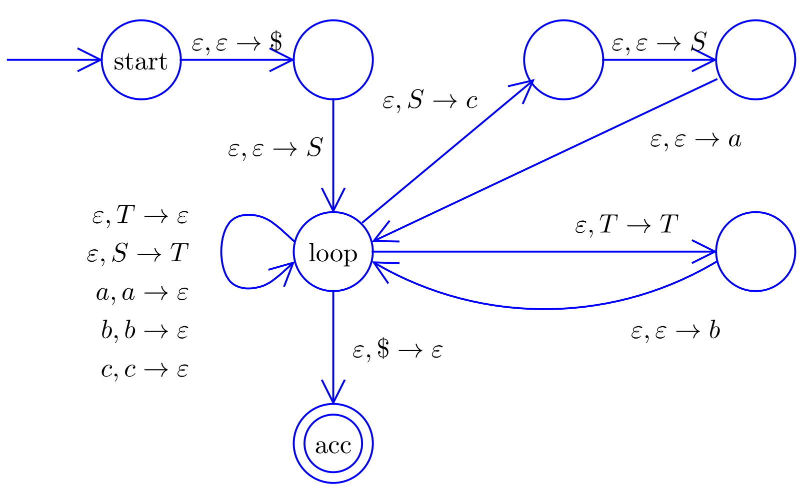
Arvostelu:

Tehtävänannossa mainittu pääteaakkosto ei sisällä merkkejä $\emptyset$ ja ε ja näitä merkkejä sisältäviä säännöllisiä lausekkeita ei siis pystytä tuottamaan. Niiden mukaan ottamisesta ei kuitenkaan rangaista. ε on tällöin ongelmallinen, koska se voidaan tulkita joko päätesymboliksi tai tyhjäksi merkkijonoksi. Arvostelussa on käytetty aina suotuisinta tulkintaa.

Säännöllisiä lausekkeita voi kieltä muuttamatta muokata mm. lisäämällä sulkuja, monistamalla *-operaattoria tai jättämällä $\circ$-operaattori merkitsemättä. Siis esimerkiksi $0 \circ 1^*$ ja $((0)1^{**})$ kuvaavat saman kielen. Tehtävässä ei ole edellytetty, että kielioppi tuottaa kaikki näin saatavat variaatiot. Epäyhtenäisyydestä tässä suhteessa vähennetään kuitenkin puoli pistettä.

Monessa vastauksessa oli ongelmana, että ne eivät pysty tuottamaan mielivaltaisen syviä sisäkkäisiä lausekkeita. Tällaisesta vastauksesta on tyypillisesti saanut yhden pisteen. Vähemmän merkittävällä tavalla virheellisestä vastauksesta on tyypillisesti saanut kaksi pistettä.

(c) Pinoautomaatti:



Merkkijonon aabcc vasen johto (muita johtoja ei olekaan):

$$S \Rightarrow \mathrm{a}S\mathrm{c} \Rightarrow \mathrm{aa}S\mathrm{cc} \Rightarrow \mathrm{aa}T\mathrm{cc} \Rightarrow \mathrm{aab}T\mathrm{cc} \Rightarrow \mathrm{aabcc} \qquad .$$

Pinon sisällöt ja lukematta oleva syöte hyväksyvässä laskennassa:

(tyhjä)	aabcc
\$	aabcc
S\$	aabcc
c\$	aabcc
Sc\$	aabcc
aSc\$	aabcc
Sc\$	abcc
cc\$	abcc
Scc\$	abcc
aScc\$	abcc
Scc\$	bcc
Tcc\$	bcc
Tcc\$	bcc
bTcc\$	bcc
Tcc\$	cc
cc\$	cc
c\$	c
\$	(tyhjä)
(tyhjä)	(tyhjä)

Arvostelu:

Pisteistä noin puolet tulee automaatista ja toinen puoli johdosta ja simulaatiosta. Pikkuvirheestä vähennetään neljännespiste. Tyypillisiä pikkuvirheitä ovat väärin merkityt siirtymät automaatissa ja puuttuvat välivaiheet johdossa tai simulaatiossa. Jos johdon sijasta on annettu jäsennyspuu, tästä on vähennetty puoli pistettä. Muulla tavalla kuin mainitulla menetelmällä tuotetusta automaatista on vähennetty piste.

2. **Huom.** tehtävässä oli virhe: kuvattu ongelma on nimeltään solmupeite (vertex cover), ei dominoiva joukko (dominating set). Tällä ei sinänsä ole vaikutusta ratkaisuun. Ratkaisu on tässä esitetty tehtävässä kuvatulle ongelmalle, jolle yleensä käytetään merkintää VC.

   (a) Kieli DS voidaan tunnistaa epädeterministisessä polynomisessa ajassa seuraavasti:

      i. Jos syöte ei ole muotoa $\langle G, k \rangle$, missä $G = (V, E)$ on verkko ja k on luonnollinen luku, niin hylkää.
      ii. Alusta $V \to \emptyset$.
      iii. Jokaisella $v \in V$ valitse epädeterministisesti: joko lisää v joukkoon V tai pidä V ennallaan.
      iv. Jos $|V| \neq k$, niin hylkää.
      v. Toista kaikilla $(u, v) \in E$: jos $u \notin V$ ja $v \notin V$, niin hylkää.
      vi. Hyväksy.

   Siis $DS \in \mathrm{NP}$.

(b) Koska DS on NP-täydellinen, sille on olemassa polynomisessa ajassa toimiva (deterministinen) ratkaisualgoritmi, jos ja vain jos P = NP. Tämä koskee myös vastaavaa etsintäongelmaa, jossa on annettu verkko G ja halutaan löytää pienin kelvollinen solmujoukko V. Käytännössä yleensä halutaankin ratkaista nimenomaan tämä etsintäongelma. Mahdollisia lähestymistapoja ovat mm. ei-polynomiset algoritmit (riittävän pienillä verkoilla), approksimointialgoritmit (jotka löytävät solmujoukon, joka todistettavasti ei ole kovin paljon suurempi kuin optimaalinen) ja erilaiset heuristiikat, jotka toimivat nopeasti ja voivat tuottaa kelvollisia ratkaisuja mutta joiden toimivuudesta ei ole takuita.

(c) Tällainen on esimerkiksi luennoilla käsitelty Turingin koneen hyväksymisongelma

$$A_{\mathrm{TM}} = \{\, \langle M, w \rangle \mid M \text{ on Turingin kone, joka hyväksyy syötteen } w \,\}.$$

Koska kieli on tunnistettava, on olemassa Turingin kone, joka tunnistaa sen, ts. hyväksyy kieleen A_{TM} kuuluvat merkkijonot, ja kieleen kuulumattomilla merkkijonoilla joko hylkää tai menee ikuiseen silmukkaan. Ei-ratkeavuudesta seuraa, että ikuinen silmukka -vaihtoehdosta ei päästä eroon: mikä tahansa kielen A_{TM} tunnistava Turingin kone jää ikuiseen silmukkaan joillain syötteillä. On siis mahdotonta laatia sellaista algoritmia, joka pysähtyisi kaikilla syötteillä ja ratkaisisi, pysähtyykö annettu Turingin kone annetulla syötteellä.

Arvostelu: Kohdassa (a) oleellista on epädeterministisen laskennan perusajatus. Sen takia ratkaisut, joissa on yritetty kehittää ongelmaan determinististä algoritmia, ovat antaneet nolla pistettä. (Ehdotetut deterministiset algoritmit ovat luonnollisesti kaikki olleet ei-polynomisia tai virheellisen tuloksen tuottavia.)

Kohtien (b) ja (c) muotoilu oli sellainen, että niihin ei ole mitään yksikäsitteistä oikeaa vastausta. Pisteitä on annettu melko avomielisesti vastauksista, jotka osoittavat opiskelijan ymmärtäneen peruskäsitteet. Esim. (b)-kohdasta on voinut saada täydet pisteet, jos on selittänyt (hyvin), miten ongelman ratkaiseminen liittyy "P = NP?"-ongelmaan.

3. (a) **Väite:** Yhteydettömien kielten luokka ei ole suljettu leikkauksen suhteen.
Todistus: Yhteydettömien kielten luokka on suljettu leikkauksen suhteen, jos ja vain jos kaikkien yhteydettömien kielten leikkaukset ovat yhteydettömiä. Merkitään annettua ei-yhteydellistä kieltä kieltä $L = \{\, 0^n1^n0^n1^n \mid n \in \mathbb{N} \,\}$ ja muodostetaan vastaesimerkki, jossa yhteydettömien kielten leikkaus on ei-yhteydetön kieli L.
Määritellään seuraavat kielet

$$\begin{aligned} A &= \{0^i1^k0^k1^i \mid i,k \in \mathbb{N}\} \\ B &= \{0^i1^j0^k1^k \mid i,j,k \in \mathbb{N}\}. \end{aligned}$$

Kielet A ja B ovat yhteydettömiä, sillä ne voidaan muodostaa seuraavilla yhteydettömillä kieliopeilla (kieliä vastaavat lähtösymbolit ovat S_A ja S_B)

$$\begin{aligned} S_A &\to 0S1 \mid T \\ T &\to 1T0 \mid \varepsilon \end{aligned} \qquad\qquad \begin{aligned} S_B &\to 0S_B \mid U \\ U &\to 1U \mid V \\ V &\to 0V1 \mid \varepsilon. \end{aligned}$$

Nyt $A \cap B = L$. Yhteydellisten kielten leikkaus ei siis ole aina yhteydellinen, mikä todistaa väitteen. □

Arvostelijan EJ huomioita a)-kohdasta (3 pistettä):
Pisteytyksen suuntaviivat:

- +1 piste: vastauksen rakenne on johdonmukainen, vastaukseen on kirjoitettu tarpeellinen johdatteluteksti ja sanalliset perustelut.
- +1 piste: vastaesimerkkiin soveltuvat yhteydettömät kielet keksitty.

- $+0{,}5$ pistettä: valittujen kielten yhteydettömyys osoitettu (esim. antamalla yhteydettömät kieliopit)
- $+0{,}5$ pistettä: valittujen kielten leikkaus muodostettu oikein (annettuna valitut kielet)
- $-?$ pistettä: sekavuudesta, monitulkintaisuudesta, epätäsmällisyydestä, väärinkäsityksistä, huonosti ymmärrettävästä kielestä, epäselvästä käsialasta ja vastaavista heikentävistä seikoista.

Vaihtoehtoisesti vastaesimerkin voi muodostaa myös kolmen yhteydellisen kielen leikkauksesta.

Harmillisen yleinen virhe (joukko-opin alkeet) oli väittää, että seuraavien kielien A ja B leikkaus olisi L.

$$A = \{\, 0^m 1^m 0^n 1^n \mid m, n \in \mathbb{N} \,\} \text{ ja } B = \{\, 0^n 1^n 0^m 1^m \mid m, n \in \mathbb{N} \,\}$$
$$A = \{\, 0^n 1^n \mid n \in \mathbb{N} \,\} \text{ ja } B = \{\, 0^n 1^n \mid n \in \mathbb{N} \,\}$$
$$A = \left\{\, 0^n 1^n 0^i 1^k \mid n, i, k \in \mathbb{N} \,\right\} \text{ ja } B = \left\{\, 0^i 1^k 0^n 1^n \mid n, i, k \in \mathbb{N} \,\right\}$$

Kahdessa ensimmäisessä tapauksessahan $A = B = A \cap B$, joten vastaesimerkki ei toimi. Virheet: konkatenaatio ja leikkaus sekoitettu keskenään; joukoissa A ja B käytetyt *muuttujat* m ja n oletettu vakioarvoiksi tai muuten riippuvaisiksi toisistaan. Kolmannessa tapauksessa leikkauskieli on $\left\{\, 0^i 1^i 0^k 1^k \mid i, k \in \mathbb{N} \,\right\}$, koska joukkojen määrittelyissä käytetyt muuttujat n,i ja k ovat toisistaan riippumattomia.

(b) **Väite:** Yhteydettömien kielten luokka ei ole suljettu komplementoinnin suhteen.

Todistus: Tehdään vastaoletus, että yhteydettömien kielten luokka *on* suljettu komplementoinnin suhteen. Olkoot A ja B mielivaltaisia yhteydettömiä kieliä. De Morganin lain mukaan $A \cap B = \overline{\overline{A} \cup \overline{B}}$. Tunnetusti yhteydettömien kielten yhdiste on yhteydetön, ja vastaoletuksen mukaan myös yhteydettömän kielen komplementti on yhteydetön, joten $A \cap B$ on yhteydetön. Yhteydettömien kielten luokka on siis suljettu leikkauksen suhteen, mikä on ristiriita a)-kohdan tuloksen kanssa. Siis vastaoletus on väärin ja väite tosi.$\square$

Arvostelijan EJ huomioita b)-kohdasta (3 pistettä):

Pisteytyksen suuntaviivat:

$+1$ piste: vastauksen rakenne on johdonmukainen, vastaukseen on kirjoitettu tarpeellinen johdatteluteksti ja sanalliset perustelut.

- $+1$ piste: todettu ja käytetty yhteydettömien kielten sulkeumaominaisuutta yhdisteen suhteen
- $+1$ piste: havaittu $A \cap B = \overline{\overline{A} \cup \overline{B}}$ ja todistuksen yksityiskohdat kunnossa
- $-?$ pistettä: sekavuudesta, monitulkintaisuudesta, epätäsmällisyydestä, väärinkäsityksistä, huonosti ymmärrettävästä kielestä, epäselvästä käsialasta ja vastaavista heikentävistä seikoista.

Vaihtoehtoinen vastaus: määritellään jokin kieli A, muodostetaan sen komplementti $\overline{A}$ eksplisiittisesti joukkonotaatiolla ja osoitetaan, että kielistä A ja $\overline{A}$ yksi on yhteydetön ja toinen ei-yhteydetön. Harmillisen yleinen virhe oli virheellinen komplementtikielen määrittely — yleensä komplementtiin sisällytettiin vain osa siihen kuuluvista merkkijonoista.

(c) **Lause:** Kieli $D = \{\, \langle M \rangle \mid \langle M \rangle \notin L(M) \,\}$ ei ole Turing-tunnistettava.

Todistus: Tehdään vastaoletus, että kieli D *on* Turing-tunnistettava. On siis olemassa jokin Turingin kone, jonka tunnistama kieli on D. Olkoon M_D tämä kone ja $L(M_D) = D$ sen tunnistama kieli. Tarkastellaan, hyväksyykö kone M_D syötteen, joka on koneen oma merkkijonoesitys $\langle M_D \rangle$.

$$\langle M_D \rangle \in L(M_D) \Leftrightarrow \langle M_D \rangle \in D \Leftrightarrow \langle M_D \rangle \notin L(M_D)$$

Ensimmäinen ekvivalenssi seuraa Turingin koneen M_D valintakriteeristä ja toinen ekvivalenssi kielen D määritelmästä. Seurauksena on ristiriita sille, kuuluuko $\langle M_D \rangle$ koneen tunnistamaan kieleen $L(M_D)$ vai ei. Kone M_D ei siis tunnistakaan kieltä D, joten vastaoletus on väärin ja väite oikein.$\square$

Arvostelijan EJ huomioita c)-kohdasta (3 pistettä):
Kyseinen kieli D on luennoillakin esitelty laskennallisesti hankala diagonaalikieli. Yleistä diagonaalimenetelmää käytetään tuottamaan vastaesimerkkejä tilanteisiin, joissa erilaisia *ongelmia* on "enemmän" kuin *ratkaisuja*. Esimerkiksi reaalilukujen ylinumeroituvuus voidaan osoittaa olettamalla numeroituvuus ja muodostamalla sitten vastaesimerkki (reaaliluku) diagonaalimenetelmällä. Tässä diagonaalikieli osoittaa, ettei kaikkia kieliä voida tunnistaa Turingin koneilla, millä on merkittäviä seurauksia laskennallisten ongelmien ratkaisulle tietokoneilla.

Pisteytyksen suuntaviivat:

- $+1$ piste: vastauksen rakenne on johdonmukainen, vastaukseen on kirjoitettu tarpeellinen johdatteluteksti ja sanalliset perustelut.
- $+1$ piste: Turingin koneen M_D merkkijonoesitystä $\langle M_D \rangle$ käytetty syötteenä
- $+1$ piste: todistuksen viimeistelyn yksityiskohdat ja ristiriidan osoittaminen
- $-?$ pistettä: sekavuudesta, monitulkintaisuudesta, epätäsmällisyydestä, väärinkäsityksistä, huonosti ymmärrettävästä kielestä, epäselvästä käsialasta ja vastaavista heikentävistä seikoista.

Arvostelijan EJ yleisiä huomioita tehtävästä 3

- Tällä kertaa lyhimmät todistukset perustuivat vastaoletusten kumoamiseen ja vastaesimerkkien keksimiseen. Joskus tämä on ainoa vaihtoehto, mutta tavallisesti suositaan ns. konstruktiivista todistusta, jossa oletukset ja tunnetut tulokset todistavat väitteen suoraan ilman ristiriitoja. Konstruktiiviset todistukset ovat tavallisesti helpompia seurata ja vähemmän virhealttiita.
- Todistusyritysten heikko kirjallinen taso yllätti. Todistus on *vakuuttava perustelu väitteen paikkansapitävyydestä* ja tarkoitettu luettavaksi kelle vain, jolla on riittävät ennakkotiedot. Jos todistusta näytetään kanssaopiskelijalle, hänen tulisi uskoa väite ehdoitta, eikä mitään epäselvyyksiä tulisi ilmetä. Jotta tähän päästäisiin, todistus vaatii seurakseen johdattelevaa tekstiä, joka kertoo lukijalle todistuksen kulun. Muuten todistus jää irrallisten matemaattisten merkintöjen sekamelskaksi, jonka sisällön vain kirjoittaja itse voi tulkita (eikä siis kelpaa todistukseksi). Periaate on sama kuin suomen kielen puhumisessa tai kirjoittamisessa: viestin välittäminen lukijalle niin selkeästi kuin mahdollista.
- Sanojen *selvästi* ja *triviaali* käyttöä tulisi välttää, sillä kirjoittaja voi muuten vaikuttaa ylimieliseltä.
- Romanian kielen kurssista ei pääse läpi ilman kielen huolellista opettelua. Vastaavasti puutteet matematiikan kielitaidossa vaikeuttavat läpipääsyä matemaattisilla kursseilla, joihin Laskennan mallitkin osittain kuuluu.