T-79.148 **Kevät 2004**
Tietojenkäsittelyteorian perusteet
Harjoitus 2
Demonstraatiotehtävien ratkaisut

4. **Tehtävä:** Osoita, että mikä tahansa vähintään kaksimerkkinen aakkosto Σ on samanveroinen binääriaakkoston $\Gamma = \{0, 1\}$ kanssa siinä mielessä, että Σ:n merkkijonot voidaan helposti koodata Γ:n merkkijonoiksi ja kääntäen. Miten paljon merkkijonon pituus voi muuttua suunnittelemassasi koodauksessa? (Siis jos merkkijonon $w \in \Sigma^*$ pituus on $|w| = n$ merkkiä, mikä on sen vastinjonon $w' \in \Gamma^*$ pituus?) Onnistuisiko vastaava koodaus, jos kohdeaakkostossa olisikin vain *yksi* merkki, esim. $\Gamma = \{1\}$?

   Vastaus: Olkoon Σ jokin k-merkkinen aakkosto, $k > 1$. Σ:n merkkijonot voidaan koodata $\Gamma = \{0, 1\}$:n merkkijonoiksi seuraavasti:

   - Samaistetaan Σ:n aakkosto kokonaislukujen $\{1, \dots, k\}$ kanssa.
   - Nämä luvut (eli Σ:n aakkosto) voidaan esittää $\lceil \log_2 k \rceil$:n mittaisina binäärilukuina.
   - Jokainen Σ:n merkkijono voidaan esittää Γ:n merkkijonona korvaamalla Σ:n aakkoset merkkijonossa niiden binäärikoodauksella.

   Koodauksen purku $\Gamma^* \to \Sigma^*$ onnistuu vastaavasti ottamalla merkkijonosta $\lceil \log_2 k \rceil$:n mittaiset merkkijonot ja tulkitsemalla ne Σ:n aakkosiksi.

   Jos merkkijonon $w \in \Sigma^*$ pituus on $|w| = n$ merkkiä on tällöin sen vastinjonon $w' \in \Gamma^*$ pituus $|w'| = n \cdot \lceil \log_2 k \rceil$, sillä jokaisen Σ:n merkin koodaamiseen tarvitaan $\lceil \log_2 k \rceil$ merkkiä.

   Tarkastellaan esimerkiksi aakkostoa $\Sigma = \{a, b, c, d, e, f\}$ ja merkkijonoa *aacfd*. Koska $|\Sigma| = 6$, tarvitaan symbolien esittämiseen $\lceil \log_2 6 \rceil = \lceil 2.58 \rceil = 3$ bittiä. Yksi mahdollinen koodaus on:

   $$\begin{array}{ll} a \mapsto 001 & d \mapsto 100 \\ b \mapsto 010 & e \mapsto 101 \\ c \mapsto 011 & f \mapsto 110 \end{array}$$

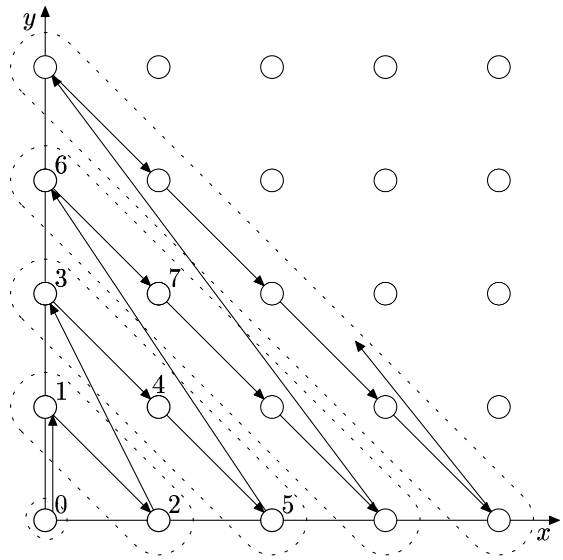
   Tällöin jonon *aacfd* esitys on 001001011110100.

   Vastaavaa koodausta ei voida rakentaa, mikäli $\Gamma = \{1\}$. Luvuille voidaan kyllä määrittää unaariesitys seuraavasti: $1 \mapsto 1$, $2 \mapsto 11$, $3 \mapsto 111$, $\dots$, mutta näin syntynyttä koodia ei voida enää purkaa yksikäsitteisesti. Esim. lukujot 1 1 1, 1 2, 2 1 ja 3 koodautuvat kaikki merkkijonoksi 111.

5. **Tehtävä:** Osoita, että karteesinen tulo $\mathbb{N} \times \mathbb{N}$ on numeroituvasti ääretön. (*Vihje:* Ajattele parit $(m, n) \in \mathbb{N} \times \mathbb{N}$ sijoitetuiksi euklidiseen (x, y)-tasoon $\mathbb{R}^2$. Numeroi parit suoran $y = -x$ suuntaisin vinorivein.) Päättele tämän tuloksen perusteella, että myös rationaalilukujen joukko $\mathbb{Q}$ on numeroituvasti ääretön.

   Vastaus: Joukko S on numeroituvasti ääretön, mikäli voidaan muodostaa bijektio $f : \mathbb{N} \to S$. Intuitiivisesti tämä tarkoittaa sitä, että kaikille joukon S alkioille voidaan asettaa yksikäsitteinen järjestysnumero.

   Joukon $\mathbb{N} \times \mathbb{N}$ alkiot $(x, y) \in \mathbb{N} \times \mathbb{N}$ voidaan numeroida seuraavan kuvan mukaisesti:



Ajatuksena on siis järjestää kaikki lukuparit suoran $y = -x$ suuntaisiin jonoihin, ja numeroida nämä jonot alkioittain yksi kerrallaan lyhyimmästä alkaen. Tässä numerointia ei voida suorittaa x-akselin suuntaisesti, sillä silloin kaikki indeksit kuluisivat jo y-akselin läpikäyntiin eikä yhtään lukuparia $(x, y), y > 0$ saavutettaisi koskaan.

Ylläoleva numerointi voidaan määritellä seuraavasti:

$$f(x, y) = x + \sum_{k=1}^{x+y} k = x + \frac{(x+y)(x+y+1)}{2}$$

Esimerkiksi $f(3, 1) = 13$, eli lukuparin $(3, 1)$ järjestysnumero on 13. Todettakoon vielä, että funktio $f(x, y)$ on todellakin bijektio, eli kutakin järjestysnumeroa vastaa yksikäsitteinen lukupari. Koordinaattiparin laskeminen annetusta indeksistä on kuitenkin hankalampaa, ja se jätetäänkin vastausten lopussa olevaan liitteeseen tiedoksi asiasta kiinnostuneille.

Positiiviset rationaaliluvut $\mathbb{Q}^+$ voidaan esittää $\mathbb{N} \times \mathbb{N}$ -lukuparina s.e.

$$(x, y) \equiv \frac{x}{y}, \qquad y \neq 0$$

Tämä on numeroituvasti äärettömän joukon $\mathbb{N} \times \mathbb{N}$ aito osajoukko, joten $\mathbb{Q}^+$ on joko äärellinen tai numeroituvasti ääretön. Jos $\mathbb{Q}^+$ olisi äärellinen, olisi olemassa joku rationaaliluku $\frac{x}{y}, x \in \mathbb{N}, y \in \mathbb{N}, y \neq 0$, jolla olisi suurin järjestysluku $n < \infty$ ($\mathbb{Q}^+$:n numeroinnissa). Kuitenkin voidaan aina löytää yo. kuvan perusteella rationaaliluku, jolla olisi järjestysluku $n' > n$, joten tämä on ristiriita oletuksen $\mathbb{Q}^+$ on äärellinen kanssa. Näin ollen $\mathbb{Q}^+$ on numeroituvasti ääretön.

Joukko $\mathbb{Q}^+$ voidaan laajentaa kaikkien rationaalilukujen joukoksi määrittelemällä joukko:

$$\mathbb{Q}^- = \{(-x, y) \mid (x, y) \in \mathbb{Q}^+\} \ .$$

Koska joukko $\mathbb{Q} = \mathbb{Q}^+ \cup \mathbb{Q}^-$ on kahden numeroituvasti äärettömän joukon yhdiste, on myös se numeroituvasti ääretön.

6. **Tehtävä:** Olkoon S mielivaltainen epätyhjä joukko.

   (a) Muodosta jokin injektiivinen funktio $f : S \to \mathcal{P}(S)$.

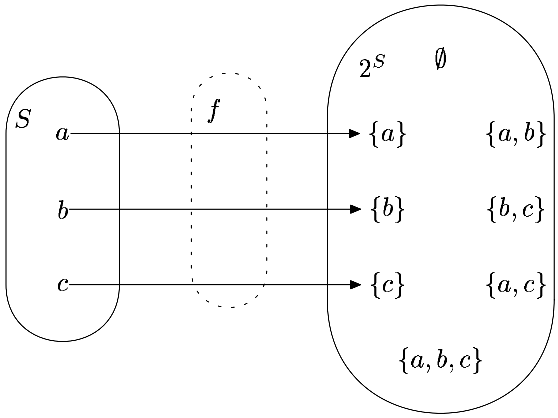
(b) Osoita, että ei ole mahdollista muodostaa injektiota $g : \mathcal{P}(S) \to S$. (*Vihje:* Oletetaan, että tällainen injektio g olisi olemassa. Tarkastellaan joukkoa $R = \{s \in S \mid s \notin g^{-1}(s)\}$ ja merkitään $r = g(R)$. Onko tällöin $r \in R$?)

Totea (b)-kohdan seurauksena, että minkä tahansa numeroituvasti äärettömän joukon S potenssijoukko on ylinumeroituva.

Vastaus: Oletetaan $S \neq \emptyset$ mielivaltainen.

(a) Määritetään $f : S \to \mathcal{P}(S)$ s.e f on injektio. Kaikilla $a \in S$ on olemassa $\{a\} \in \mathcal{P}(S)$ ja jos $a \neq b$, niin $\{a\} \neq \{b\}$, joten kuvaus $f : S \to \mathcal{P}(S)$, $f(a) = \{a\}$ on halutunlainen injektio.

Alla on esitetty kuvaus f joukolle $S = \{a, b, c\}$:



(b) Oletetaan, että on olemassa injektio $g : \mathcal{P}(S) \to S$. Määritellään osajoukko $S' \subseteq S$ seuraavasti:

$$S' = \{a \in S \mid \text{ on olemassa joukko } A \subseteq S \text{ siten, että } g(A) = a\} \ .$$

Huomataan, että S' ei voi olla tyhjä joukko, sillä $|\mathcal{P}(S)| > 0$ kaikilla joukoilla S. Tarkastellaan joukkoa $R = \{s \in S' \mid s \notin g^{-1}(s)\}$ ja merkitään $r = g(R)$. Jos $r \in R$, pätee $r \notin g^{-1}(r)$. Kuitenkin

$$g^{-1}(r) = g^{-1}(g(R)) = R,$$

joten kyseessä on ristiriita. Toisaalta, jos $r \notin R$, pätee $r \in g^{-1}(r) = g^{-1}(g(R)) = R$, joka on myös ristiriita. Koska sekä $r \in R$ että $r \notin R$ johtavat ristiriitaan, täytyy jokin aikaisemmin tehdyistä oletuksista olla väärä. Ainoa tällainen oletus on se, että injektio g on olemassa, joten voimme todeta, ettei ole mahdollista muodostaa injektiota $\mathcal{P}(S) \to S$.

Yllä määriteltiin joukko S' sen takia, että muutoin merkintä $g^{-1}(s)$ ei välttämättä olisi määritelty määriteltäessä joukkoa R.

Täten ei ole mahdollista muodostaa injektiota $g : \mathcal{P}(S) \to S$. Jos S on numeroituvasti ääretön, on olemassa bijektio $f : \mathbb{N} \to S$. Jotta $\mathcal{P}(S)$ olisi numeroituva, pitäisi olla bijektio $f' : \mathbb{N} \to \mathcal{P}(S)$. Oletetaan, että tällainen bijektio f' olisi olemassa. Kuvaus $g \circ f'^{-1} : \mathcal{P}(S) \to S$ on tällöin bijektio (eli injektio ja surjektio), mikä on ristiriidassa sen kanssa, että ei ole olemassa injektiota $\mathcal{P}(S) \to S$. Näin ollen $\mathcal{P}(S)$ on ylinumeroituva.

Liite: koordinaattiparin laskeminen järjestysnumerosta tehtävässä 4.

Olkoon annettuna järjestysnumero m ja halutaan laskea koordinaatit x ja y siten, että

$$x + \frac{(x+y)(x+y+1)}{2} = m \ . \tag{1}$$

Merkitään $z = x + y$, jolloin (1) muuttuu muotoon:

$$z - y + \frac{z(z+1)}{2} = m \quad . \tag{2}$$

Koska $z - y \geq 0$,

$$\begin{aligned} \frac{z(z+1)}{2} &\leq m \qquad\qquad , \text{joten} && (3)\\ z &\leq \frac{-1 \pm \sqrt{1+8m}}{2} && (4) \end{aligned}$$

Koska $z \in \mathbb{N}$ ja $m, x, y \geq 0$, huomataan, että:

$$z = \left\lfloor \frac{-1 + \sqrt{1+8m}}{2} \right\rfloor \tag{5}$$

Nyt voidaankin laskea sekä x että y käyttäen apuna indeksointifunktiota f:

$$\begin{aligned} x &= m - f(0, z) \\ y &= z - x \end{aligned} \tag{6}$$

Tässä $f(0, z)$ antaa (x, y):n vinorivin ensimmäisen alkion järjestysnumeron. Lasketaan esimerkiksi indeksiä $m = 13$ vastaava pari:

$$\begin{aligned} z &= \left\lfloor \frac{-1 + \sqrt{105}}{2} \right\rfloor = \lfloor 4.62 \rfloor = 4 \\ x &= 13 - f(0, 4) = 13 - 10 = 3 \\ y &= 4 - 3 = 1 \end{aligned}$$

Tulokseksi saatiin pari $(3, 1)$ niin kuin pitikin.