

## Tehtävä 1: Massadata (big data) ja sen käytön haasteet

Lue ensin seuraava taustamateriaali huolellisesti ja vastaa sen jälkeen siihen liittyviin kysymyksiin.

Maailmassa mitataan jatkuvasti monenlaisia asioita, kuten vaikkapa radiosignaalin laatua, ja tästä syntyviä mittausrvoja nimitetään dataksi. Myös sosiaalisen median päivitykset tai kauppaketjujen kassatiedot ovat dataa. Eri-laista ja eri lähteistä peräisin olevaa dataa syntyy maailmassa koko ajan enemmän ja sen jakaminen hetkessä ympäri maailmaa on helppoa. Puhutaan massadatasta ("big data"), kun tarkoitetaan niin laajoja ja monimutkaisia datamääriä, että niiden analysointi ja hyödyntäminen on erittäin haastavaa ja vaatii uusien menetelmien kehittämistä. Tästä huolimatta massadatan käytölle on asetettu suuria odotuksia. Ajatellaan, että se tuottaa uusia liiketoimintamahdollisuuksia ja toisaalta antaa mahdollisuuden ratkoa sellaisia ongelmia, joita ennen ei ole pystytty ratkaisemaan.

Kun massadataa luonnehditaan, siihen liitetään usein seuraavia sanoja:

- *Volyyymi*: Dataa syntyy jatkuvasti niin valtavia määriä, että sen käsitteleminen ja tallentaminen on hyvin haastavaa. Se myös saattaa sijaita fyysisesti useassa eri paikassa, mikä edelleen lisää haastetta datan käsittelyyn.
- *Vauhti*: Dataa syntyy myös jatkuvasti kiihtyvällä tahdilla. Jo tällä hetkellä on mahdotonta tallettaa kaikkea dataa ja tulevaisuudessa tilanne on aina vain pahempi. Toisaalta kuitenkin olisi tärkeää päästä käsiksi kaikkeen syntyvään dataan, jotta sitä voidaan hyödyntää älykkäästi.
- *Variaatio*: Dataa on monen erityyppistä ja erilaisten datalähteiden määrä kasvaa sekin koko ajan. Syntyvän datan standardoiminen käyttämään tiettyjä tallennusmuotoja ei ole toistaiseksi vielä toiminut tyydyttävällä tavalla. Tämä hankaloittaa datan analyysiä ja hyötykäyttöä. Eroja dataan syntyy myös muun muassa erilaisista käytetyistä mittaustavoista, mittayksiköistä, tavoista kirjata numeerisia tietoja ja jopa mittareiden kalibroinnista. Myös kieli- ja kulttuurierot varsinkin tekstimuotoisen datan tallentamisessa voivat aiheuttaa päänvaivaa. Visuaalisen datan erisitysmuodot voivat nekin aiheuttaa ongelmia, samoin paikkatieto erilaisten koordinaattijärjestelmien takia.
- *Oikeellisuus*: Syntyvä data ei välttämättä ole kovin täsmällistä. Se voi myös olla epäselvää tai jopa virheellistä. Joskus datan keruu- ja mittausten menetelmät ovat erilaisia tai niitä ei tunneta, ja siksi eri lähteistä

tuotetun datan vertailukelpoisuus ei ole itsestään selvää. Pahimmassa tapauksessa esimerkiksi lämpötiladatasta ei tiedetä onko sen mittayksikkö Celsius vai Fahrenheit.

- *Oleellisuus ja arvo:* Data on usein reaaliaikaista, jolloin sen hyödyllisyyskin on sidottuna ajanhetkeen. Data ei välttämättä ole kovin kiinnostavaa tai oleellista tietyllä ajanhetkellä tai tietyssä asiayhteydessä, mutta esimerkiksi myöhempanä ajanhetkenä tai toisessa asiayhteydessä sama data saattaa olla hyvin kiinnostavaa ja arvokasta. Toisaalta yhdelle toimijalle tai organisaatiolle data saattaa olla arvotonta, kun taas toiselle se on kullannarvoista. Joskus datan arvo, merkitys, tulee nimenomaan siitä että se tarjoaa tietoa datan muutoksista ajan suhteen. Esimerkkinä tästä on ilmastotutkimukseen liittyvä data, jossa esimerkiksi meren pinnan korkeuden muutokset nimenomaan ajan suhteen ovat olennainen ja kiinnostava tieto.

Nämä luonnehdinnat kertovat, että massadata-ilmiossa on kyse muustakin kuin vain valtavan suurista datamääristä. Jotta suuresta määrästä dataa on hyötyä, se pitää jalostaa tiedoksi. Data muuttuu tiedoksi silloin, kun data jäsennetään ja esitetään niin, että sillä on jokin merkitys. Vaikkapa matkapuhelinoperaattorilla voi olla tiedossa kaikkien asiakkaidensa kännyköiden reaaliaikainen sijainti, mutta tämä data muuttuu tiedoksi vasta, kun sitä hyödynnetään sopivasti, esimerkiksi tarjoamalla tietyssä sijainnissa olevien kännyköiden kautta palvelua siten, että kaupungin keskustassa tarjotaan eri palveluja kuin esikaupunkialueella. Vastaavasti sääasemien havaintodatalla ei tee mitään, ennen kuin se on analysoitu sääennusteeksi tai ilmaston muuttumista kuvaavaksi historiatiedoksi. Sähän liittyvän datan määrä kasvaa jatkuvasti ja tänään tietoa voidaan kerätä eri muodossa, eri mittarein ja mittayksiköin kuin vuosia tai vuosikymmeniä sitten. Ilmastodata (lämpötilat, kaasupitoisuudet, meren pinnan korkeudet, jne.), joka ulottuu tästä päivästä taaksepäin aina satojen tuhansien vuosien päähän menneisyyteen, on kerätty eri tekniikoin ja menetelmin, osin laskennallisesti, ja se on erittäin haasteellista analyysin kannalta. Massadatan haasteena esimerkiksi tällaisessa aineistossa on se, miten tuloksista saadaan merkityksellistä ilman että alkuperäistä dataa tulkitaan väärin tai virheellisin oletuksin.

Jotta data voi muuttua tiedoksi, tarvitaan paljon erilaisia apuvälineitä: monimutkaisia ohjelmistoja, laskentamalleja ja ratkaisuja, ja niitä tekemään tarvitaan monenlaista eri osaamista: muun muassa ohjelmointitaitoa, tilastojen ymmärtämistä ja liiketoiminnan tajua. Lisäksi itse datasta pitää olla riittävästi tietoa (miten, missä ja kuka mittasi ja millä tekniikoilla).

Massadatan käyttöön liittyy usein myös se, että monet toimijat avaavat syntyvän datan avoimesti kenen tahansa käytettäväksi. Esimerkiksi Suomes-

sa Maanmittauslaitos on avannut keräämänsä karttadatan avoimesti kaikkien saataville, ja tätä dataa hyödyntävät useat yritykset erilaisissa karttapalveluissa. Samoin on avattu Suomen Ilmatieteen laitoksen säätutkien tuottama data, jota hyödynnetään esimerkiksi säätietoja kertovissa sovelluksissa. Esimerkeistä näkyy kuinka uutta liiketoimintaa voi syntyä, kun data on avointa.

Tärkeänä osa-alueena massadatan käytössä on luottamus: data ei synny tyhjästä ja väärinkäytönkin vaara on olemassa. Paikkakuntien lämpötilatietoja ei ehkä kovin helposti käytetä väärin, mutta jos kännyköiden sijaintitietoihin kytketään käyttäjän henkilökohtaista tietoa, onkin jo väärinkäytön mahdollisuus suurempi. Toisaalta suurista tietomassoista on kuitenkin mahdollisuus jalostaa hyvinvointia, palveluja ja uutta liiketoimintaa. Sen vuoksi jokaisen yrityksen tai organisaation, joka harkitsee datan avaamista julkiseksi, on tehtävä se harkiten ja etukäteen mietittävä, mitä dataa voidaan julkaista ja millaisilla pelisäännöillä.

## Kysymykset

**Vastaa seuraaviin kysymyksiin edellä olevan taustamateriaalin ja yleistietosi perusteella. Kysymyksien yhteenlaskettu maksimipistemäärä on 25 pistettä.**

**Kysymys 1.** Kerro konkreettinen esimerkki viidestä eri massadataa luonnehtivasta piirteestä perusteluineen. Valitse esimerkit niin että ne poikkeavat toisistaan, eivätkä ole peräisin taustamateriaalista. Anna siis yhteensä 5 esimerkkiä. Anna 1-2 lausetta perusteluja yhtä esimerkkiä kohden. (maksimipistemäärä 5)

**Kysymys 2.** Viime aikoina on keskusteltu ajoneuvojen reaaliaikaisesta seurannasta autoverotuksen yhteydessä. Jotta ajoneuvon liikkumista voidaan seurata, siihen on käytännössä asennettava GPS-paikannin, joka lähettää jatkuvasti dataa auton sijainnista. Jos oletetaan, että Suomessa otetaan käyttöön tällainen järjestelmä, niin

- a) mihin muuhun hyödylliseen voitaisiin sillä kerättyä tietoa käyttää,
- b) mitä eettisiä ongelmia tiedon keräämisessä ja hyödyntämisessä voidaan nähdä,
- c) mitä teknisiä haasteita järjestelmän luomisessa ja käytössä voidaan kohdata,
- d) mitä haasteita datan muuttamisessa tiedoksi voidaan tätä järjestelmää käyttäessä kohdata?

## Tietojenkäsittelytieteen yhteisvalinta 26.5.2014

---

Mainitse kaksi esimerkkiä kustakin kohdasta (a, b, c, d). Perustele. (maksimipistemäärä 8)

**Kysymys 3.** Usein massadataa hyödynnettäessä olemassa olevaa dataa kerätään eri lähteistä analyysiä ja tiedon tuottamista varten. Eri lähteistä kerättävän datan mittaustapa ja tallennusmuoto voi kuitenkin aiheuttaa ongelmia. Esimerkkinä tästä on vaikkapa päivämäärän tallentaminen eri muotoon: amerikkalainen tapa 04/11/2014 ilmaisee vuoden 2014 huhtikuun 11. päivää, kun taas suomalaiselle kyseessä on marraskuun 4. päivä.

Anna neljä taustamateriaalista ja toisistaan poikkeavaa esimerkkiä massadatan analyysistä, jossa voi syntyä ongelmia siihen liittyen, että tiedot ovat peräisin eri lähteistä ja datan mittaus-, keräys- ja tallennusmuodot vaihtelevat. Pohdi myös kunkin ongelman kohdalla, miten näitä ongelmia voitaisiin ratkoa. (maksimipistemäärä 12)

## Tehtävä 2: Loppuosataulukko

Tutustu tarkoin seuraavaan tekstiin ja vastaa sitä hyväksi käyttäen tehtävän loppuosassa esitettyihin viiteen kysymykseen.

Annetun merkkijonon (ns. hahmo) esiintymän haku pidemmästä merkkijonosta (tekstistä) on yksi perustavanlaatuisen tietojenkäsittelytieteen sovellus. Haun voi toteuttaa selaamalla koko tekstin alusta loppuun samalla tutkien, esiintyykö hahmo jossain kohdassa. Tämä lähestymistapa on kuitenkin hidas, jos teksti on hyvin suuri. Hakua voidaan nopeuttaa luomalla tekstistä etukäteen sopiva indeksirakenne, jonka avulla tekstin läpisiselaaminen voidaan välttää. Eräs yksinkertainen ja paljon käytetty indeksirakenne on ns. loppuosataulukko. Menetelmää on käytetty yleisesti esimerkiksi laajoihin DNA-aineistoihin, kuten n. 3 miljardia merkkiä käsittävään ihmisen genomiin, kohdistuvien hakujen toteuttamiseen tehokkaasti.

Merkkijono koostuu peräkkäisistä merkeistä. Merkkijonoon voidaan viitata merkkijonomuuttujaa käyttäen. Merkkijonon merkkeihin viitataan antamalla merkin indeksi hakasuluissa välittömästi merkkijonomuuttujan perässä. Esimerkiksi  $x[1]$  tarkoittaa merkkijonon  $x$  ensimmäistä merkkiä. Merkintä  $x[i..j]$  tarkoittaa merkkijonon  $x$  alimerkkijonoa, joka alkaa indeksistä  $i$  ja päättyy indeksiin  $j$ . Alimerkkijono, joka jatkuu merkkijonon loppuun asti, on loppuosaa. Loppuosan alkukohtaa sanotaan loppuosan indeksiksi. Merkkijonojen  $x$  ja  $y$  yhteydessä käytämme vertailuoperaattoreita  $<$ ,  $\leq$ ,  $=$ ,  $\geq$  ja  $>$  kuvaamaan merkkijonojen aakkosjärjestyksestä. Esimerkiksi merkintä  $x \leq y$  tarkoittaa, että  $x$  on aakkosjärjestyksessä pienempi tai yhtäsuuri kuin  $y$ . Yhtäsuuruus tarkoittaa, että merkkijonot ovat keskenään samanlaiset.

**Esimerkki 1.** Jos merkkijono  $x$  on “kevät”, pätee mm. että:

- $x[1] = \text{“k”}$ ,  $x[2] = \text{“e”}$  ja  $x[4] = \text{“ä”}$ .
- $x[1..3] = \text{“kev”}$ ,  $x[3..3] = \text{“v”}$  ja  $x[2..5] = \text{“evät”}$ .
- Merkkijonon  $x$  loppuosat, niiden indeksien eli alkukohtien 1, 2, 3, 4 ja 5 mukaisessa järjestyksessä, ovat  $x[1..5] = \text{“kevät”}$ ,  $x[2..5] = \text{“evät”}$ ,  $x[3..5] = \text{“vät”}$ ,  $x[4..5] = \text{“ät”}$  ja  $x[5..5] = \text{“t”}$ .

Viittaamme tekstiin merkkijonomuuttujalla  $t$  ja tekstistä etsittävään hahmoon merkkijonomuuttujalla  $p$ . Lisäksi oletamme, että tekstin  $t$  pituus on  $n$ .

Tekstin  $t$  loppuosataulukko  $S$  on  $n$ -alkiainen kokonaislukutaulukko, joka luettelee tekstin loppuosien indeksit loppuosien aakkosjärjestyksen mukaisessa järjestyksessä. Taulukon  $S$  indeksin  $i$  arvo, josta käytämme merkintää  $S[i]$ , kertoo, mistä tekstin kohdasta tekstin  $i$ :nneksi pienin loppuosaa alkaa:  $S[1]$  kertoo tekstin aakkosjärjestyksessä pienimmän loppuosan indeksin,  $S[2]$  kertoo

aakkosjärjestyksessä sitä seuraavan loppuosan indeksin, jne. Asian voi ilmaista myös niin, että indekseillä  $i = 2, \dots, n$  pätee ehto  $t[S[i-1]..n] \leq t[S[i]..n]$ .

**Esimerkki 2.** Tekstin  $t = \text{“abababba”}$  loppuosat ovat  $\text{“abababba”}$ ,  $\text{“bababba”}$ ,  $\text{“ababba”}$ ,  $\text{“babba”}$ ,  $\text{“abba”}$ ,  $\text{“bba”}$ ,  $\text{“ba”}$  ja  $\text{“a”}$ . Alla on vasemmalla tekstin  $t$  loppuosat aakkosjärjestyksessä ja oikealla tekstin  $t$  loppuosataulukko  $S$ . Huomaa, että loppuosataulukon arvot ovat samat kuin vasemmalla esitetyt loppuosien indeksit. Esimerkiksi arvo  $S[5] = 7$  ilmaisee, että aakkosjärjestyksessä viidenneksi pienin loppuosa alkaa indeksistä 7 eli on  $t[7..8] = \text{“ba”}$ .

Loppuosa (aakkosjärj.)	Loppuosan indeksi	Indeksi $i$	$S[i]$
a	8	1	8
abababba	1	2	1
ababba	3	3	3
abba	5	4	5
ba	7	5	7
bababba	2	6	2
babba	4	7	4
bba	6	8	6

Loppuosien perusominaisuus on, että jos etsitty hahmomerkkijono  $p$  esiintyy jossain tekstin kohdassa, esiintyy  $p$  samalla tekstin kyseisestä kohdasta alkavan loppuosan alkuosana. Sanomme, että tällöin hahmo  $p$  täsmää kyseisen loppuosan kanssa. Hahmo  $p$  voidaan etsiä tekstistä etsimällä sellainen loppuosa, jonka kanssa  $p$  täsmää. Loppuosataulukon ansiosta tämä onnistuu tehokkaasti käyttäen ns. puolitushakua.

Puolitushaku ylläpitää tietoa siitä loppuosataulukon välistä, jonka sisällä voisi senhetkisten tietojen valossa olla hahmon  $p$  kanssa täsmävä loppuosa. Käytämme hakuvälin alkuindeksistä merkintää *alku* ja loppuindeksistä merkintää *loppu*. Lisäksi määritämme välin keskimmäisen indeksin *keski* kaavalla  $keski = (alku + loppu)/2$ , jonka arvo pyöristetään lähimpään kokonaislukuun (eli tässä aina ylöspäin), jos summan  $alku + loppu$  arvo on pariton. Alkutilanteessa jokainen loppuosa voisi periaatteessa täsmätä eli  $alku = 1$  ja  $loppu = n$ .

Puolitushaku vertaa hahmoa ja hakuvälin keskimmäistä loppuosaa  $t[S[keski]..n]$  keskenään. Jos hahmo täsmää, voidaan haku lopettaa<sup>1</sup>. Muuten pätee joko  $p < t[S[keski]..n]$  tai  $p > t[S[keski]..n]$ .

Jos  $p < t[S[keski]..n]$  eli hahmo on aakkosjärjestyksessä pienempi kuin indeksin  $S[keski]$  omaava loppuosa, ei mikään loppuosataulukon välin  $keski, \dots, loppu$  loppuosa täsmää hahmon kanssa. Tämä on suora seuraus siitä, että loppuosataulukko ilmaisee loppuosat niiden aakkosjärjestyksessä. Tällöin puoli-

---

<sup>1</sup>Tässä tehtävässä keskitytään yhden esiintymän hakuun; hahmo toki voi esiintyä tekstissä monessa eri kohdassa.

tushaku voi päivittää hakuvälin uudeksi ylärajaksi  $loppu = keski - 1$  ja verrata hahmoa seuraavaksi tämän uuden hakuvälin keskimmäiseen loppuosaan.

Jos  $p > t[S[keski]..n]$ , voidaan edellä esitettyyn tapaan todeta, ettei mikään välin  $alku, \dots, keski$  loppuosa voi täsmätä hahmon kanssa. Tällöin hakuvälin alaraja voidaan päivittää asettamalla  $alku = keski + 1$ .

Jos hakuväli tulee haun aikana tyhjäksi eli ehto  $alku > loppu$  astuu voimaan, lopetetaan haku tuloksettomana: teksti ei tässä tapauksessa sisällä yhtään hahmon  $p$  esiintymää. Alla on esitetty hahmon  $p$  haku tekstistä  $t$  loppuosataulukon  $S$  avulla vielä hieman täsmällisemmin askeleittain:

1. Aseta hakuvälin alkuindeksi  $alku = 1$  ja loppuindeksi  $loppu = n$ .
2. Jos  $alku > loppu$  eli hakuväli on tyhjä, lopeta haku: hahmo  $p$  ei esiinny tekstissä.
3. Laske alku- ja loppuindeksien puolivälissä oleva indeksi  $keski$ :  
 $keski = (alku + loppu)/2$ , pyöristäen arvo tarvittaessa ylöspäin.
4. Vertaile hahmoa  $p$  ja hakuvälin keskimmäistä loppuosaa  $t[S[keski]..n]$ .
  - Jos  $p$  täsmää loppuosan  $t[S[keski]..n]$  kanssa, lopeta haku palauttaen tieto, että hahmon  $p$  yksi esiintymä löytyi tekstin indeksistä  $S[keski]$  alkaen.
  - Jos  $p$  ei täsmää loppuosan  $t[S[keski]..n]$  kanssa, niin:
    - Jos  $p < t[S[keski]..n]$ , karsi hakuvälistä indeksit, jotka ovat  $\geq keski$ , asettamalla  $loppu = keski - 1$  ja jatka hakua palaamalla askeleeseen 2.
    - Jos  $p > t[S[keski]..n]$ , karsi hakuvälistä indeksit, jotka ovat  $\leq keski$ , asettamalla  $alku = keski + 1$  ja jatka hakua palaamalla askeleeseen 2.

**Esimerkki 3.** Hahmon  $p = \text{“ababbaba”}$  haku tekstistä  $t = \text{“abbaababbababbab”}$ .

Loppuosa	$i$	$S[i]$
aababbababbab	1	4
ab	2	15
ababbab	3	10
ababbababbab	4	5
abbaababbababbab	5	1
abbab	6	12
abbababbab	7	7
b	8	16
baababbababbab	9	3
bab	10	14
bababbab	11	9
babbab	12	11
babbababbab	13	6
bbaababbababbab	14	2
bbab	15	13
bbababbab	16	8

1. Alkuarvot  $alku = 1$  ja  $loppu = 16$ .
2.  $keski = (1 + 16)/2 = 9$ ,  
 $p < t[S[9]..16]$ ,  $loppu = keski - 1 = 8$ .
3.  $keski = (1 + 8)/2 = 5$ ,  
 $p < t[S[5]..16]$ ,  $loppu = keski - 1 = 4$ .
4.  $keski = (1 + 4)/2 = 3$ ,  
 $p > t[S[3]..16]$ ,  $alku = keski + 1 = 4$ .
5.  $keski = (4 + 4)/2 = 4$ ,  
 $p$  täsmää loppuosan  $t[S[4]..16]$  kanssa ja haku päättyy.

Esimerkin puolitusluku joutui tutkimaan 4 tekstinkohtaa. Puolituslukuun vahvuus on, että hakuaskeleiden määrä kasvaa varsin hitaasti tekstin koon kasvaessa. Esimerkiksi n. 3 miljardista merkistä koostuvaan ihmisen genomiin kohdistuva puolitusluku joutuu tutkimaan korkeintaan 32 tekstinkohtaa.

## Kysymykset

**Kysymys 1.** Esitä ne askeleet, jotka puolitusluku tekee etsiessään hahmoa  $p = \text{“babaa”}$  esimerkin 3 tekstistä  $t = \text{“abbaababbababbab”}$ . Esitä vastauksesi samassa muodossa kuin esimerkissä 3 eli anna kunkin askeleen kohdalla hakuväliä koskevat arvot  $alku$ ,  $loppu$  ja  $keski$ . (maksimipistemäärä 3)

**Kysymys 2.** Muodosta merkkijonon  $t = \text{“yhteisvalinta”}$  loppuosataulukko. (maksimipistemäärä 4)

**Kysymys 3.**

- a) Muodosta merkkijonon  $t = \text{“aacatcgatagctagaacat”}$  loppuosataulukko. (maksimipistemäärä 4)
- b) Esitä ne askeleet, jotka puolitusluku tekee etsiessään hahmoa  $p = \text{“cga”}$  kohdan a) tekstistä. Esitä vastauksesi samassa muodossa kuin esimerkissä 3 eli anna kunkin askeleen kohdalla hakuväliä koskevat arvot  $alku$ ,  $loppu$  ja  $keski$ . (maksimipistemäärä 3)



**Kysymys 4.** Muodosta jokin sellainen suomalaisen aakkoston pienistä kirjaimista koostuva merkkijono, jonka loppuosataulukko vastaa alla annettua taulukkoa. Tässä kysymyksessä sallittuun aakkostoon kuuluu siis merkit “a”, “b”, “c”, ..., “ö”. (maksimipistemäärä 4)

$i$	$S[i]$
1	8
2	7
3	6
4	5
5	4
6	3
7	2
8	1

**Kysymys 5.** Muodosta sellainen pelkästään merkkejä “a” ja “b” sisältävä merkkijono, jonka loppuosataulukko vastaa alla annettua taulukkoa. (maksimipistemäärä 7)

$i$	$S[i]$
1	16
2	13
3	3
4	14
5	11
6	9
7	4
8	6
9	15
10	12
11	2
12	10
13	8
14	5
15	1
16	7

## Tehtävä 3: Ongelmanratkaisutehtävä

Tässä tehtävässä esitetään kolme toisistaan riippumatonta yleistä ongelmanratkaisukykyä mittaavaa osatehtävää.

**Kysymys 1.** Pöydälle asetetaan neljä pelikorttia. Kaksi näistä korteista on etupuolelta punaisia ja kaksi mustia. Kortit ovat pöydällä selkäpuoli ylöspäin, joten niiden väri ei näy. Valitaan kaksi korttia satunnaisesti, millä todennäköisyydellä ne ovat samaa väriä? (maksimipistemäärä 4)

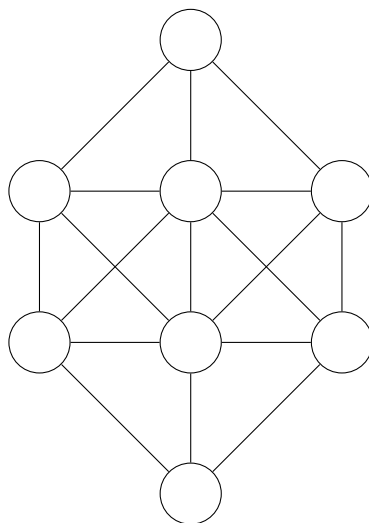
**Kysymys 2.**

(a) Piirrä alla oleva kuvio vastauspaperiin ja sijoita luvut yhdestä kahdeksaan siihen seuraavien ehtojen mukaisesti:

- kukin luku sijoitetaan omaan ympyräänsä
- kahta suuruusjärjestyksessä peräkkäistä lukua ei saa sijoittaa siten, että ne sisältävien ympyröiden välillä on viiva

Jos esimerkiksi luku 4 sijoitetaan ylimpään ympyrään, silloin lukuja 3 tai 5 ei saa sijoittaa yhteenkään kolmesta sen alla olevasta ympyrästä. (maksimipistemäärä 6)

(b) Kaikkiaan on olemassa neljä sellaista lukujen sijoittelua, jotka tuottavat a-kohdan ehtojen mukaisen ratkaisun. Oletetaan, että ratkaistaan a-kohdan tehtävä käymällä läpi kaikki mahdolliset tavat sijoitella luvut kuvioon. Jos yhden sijoittelun kokeiluun menee 10 sekuntia, kauanko tehtävän ratkaisuun pahimmillaan menee? (maksimipistemäärä 4)



**Kysymys 3.** Noppapelissä kaksi pelaajaa heittää kukin omaa 6-tahoista noppaansa ja suuremman lukeman saanut voittaa (tasatilanteessa heitetään uudestaan, kunnes peli ratkeaa). Käytettävät nopat ovat kuution muotoisia ja kukin kuudesta tahosta voi tulla heiton päällimmäiseksi yhtä todennäköisesti. Nopat ovat kuitenkin siitä epätavallisia, että niiden tahoihin merkityt arvot on voitu valita mielivaltaisesti lukujoukosta  $\{1, 2, 3, 4, 5, 6\}$ . Sama luku voi siis esiintyä useassa nopan tahossa, ja vastaavasti joitain lukuja voi puuttua kokonaan.

(a) Pöydällä on kolme noppaa (A, B ja C), joiden tahoissa on seuraavat lukemat:

- Noppa A: 2, 2, 2, 5, 5, 5
- Noppa B: 3, 3, 3, 3, 3, 6
- Noppa C: 1, 1, 1, 4, 4, 4

Pelaaja 1 saa valita ensin pöydältä noppansa, jonka jälkeen pelaaja 2 saa valita omansa jäljelle jääneistä. Perustele, miten pelaajan 1 pitäisi valita, että hänellä olisi aina etu pelissä (toisin sanoen yli 50% voittotodennäköisyys). (maksimipistemäärä 5)

(b) Tasoituksen vuoksi pelaaja 2 saa suunnitella uuden 6-tahoisen nopan kysymyksen 3 alussa kuvattujen ehtojen mukaisesti, ja asettaa sen nopan C tilalle. Jälleen, pelaaja 1 valitsee noppansa ensin, sitten pelaaja 2 valitsee jäljelle jääneistä. Suunnittele sellainen noppa C ja pelistrategia pelaajalle 2, joka varmistaa hänelle aina edun pelissä, tai osoita että tämä ei ole mahdollista. (maksimipistemäärä 6)

## Tehtävä 1: Massadata ("Big data") ja sen käytön haasteet

Oikeita vastauksia tehtävän kysymyksiin on useita ja tässä annetaan vain esimerkkivastauksia, jonka tyyppiset voidaan katsoa ratkaisuiksi. Arvostelussa huomioidaan erityisesti esimerkkien monipuolisuus ja niiden perustelut.

### Kysymys 1 mallivastaus:

Massadataa luonnehtivia piirteitä ovat volyyymi, vauhti, variaatio, oikeellisuus, oleellisuus ja arvo.

Volyymista esimerkkinä voidaan antaa kauppaketjujen kassapäätteisiin kertyvä data siitä, mitä tuotteita asiakkaat ostavat. Tätä dataa syntyy suurissa kauppaketjuissa päivittäin suuria määriä. Yksinkertaisimmillaan sitä voidaan käyttää esimerkiksi kaupan varastotilannetiedon ylläpitämiseen, mutta sitä analysoimalla voidaan myös yrittää ymmärtää asiakkaiden ostokäyttäytymistä: mitä he ostavat, mihin aikaan vuorokaudesta/vuodesta, kuinka paljon, minkä muiden ostosten yhteydessä jne. ja tätä tietoa voidaan hyödyntää esimerkiksi mainostamisessa.

Vauhdista esimerkkinä voidaan antaa Twitter-viestit, joita syntyy ympäri maailmaa satoja miljoonia päivittäin. Vaikka kukin viesti on vain 140 merkkiä pitkä, niistä päivittäin syntyvän datan määrä on huikea ja sen tallettaminen vaatii suuria määriä levytilaa. Palvelun käyttäjien ja siinä lähetettyjen viestien määrä kasvaa jatkuvasti. Twitter-viestejä analysoimalla voidaan saada selville hyvin monenlaisia asioita alkaen maailmanlaajusten yhteiskunnallisten mielipiteiden muutoksista ja päätyen esimerkiksi viihteessä sillä hetkellä pinnalla oleviin trendeihin.

Variaatiosta esimerkkinä voidaan antaa yhteisöpalvelu Facebookissa syntyvä data. Data voi olla luonteeltaan esimerkiksi käyttäjien tekstimuotoisia päivityksiä eri kielillä, kuvia, linkkejä tai paikkadataa. Facebookin omistava yhtiö voi hyödyntää tätä dataa esimerkiksi myymällä mainostilaa sopivien keskustelunaiheiden yhteydessä, mutta tekstimuotoisen, monilla eri kielillä tuotetun datan analysointi kulttuurierot huomioon ottaen on haastavaa.

Oikeellisuudesta esimerkkinä voidaan antaa neuvolassa pienten vauvojen punnituksessa vuosikymmenien kuluessa syntyvä painotieto. Yksittäisen vauvan kohdalla data on suhteellisen helppo tulkita tiedoksi: vauva esimerkiksi on kasvanut normaalia vauhtia tai kasvu on jostain syystä hidastunut, johon on aiheellista etsiä syytä. Pienen vauvan painossa kuitenkin pienetkin painon nousut ja laskut ovat merkityksellisiä ja vaa'assa oleva heitto suuntaan tai toiseen tai vauvan juuri syövä ateria voivat vaikuttaa punnitustulokseen. Yksittäisen vauvan kohdalla voidaan tarkkailla muitakin merkkejä kuin painoa (esimerkiksi kuinka virkeä vauva on), mutta mikäli halutaan ymmärtää vuosien saatossa kaikkien suomalaisten vauvojen tai kaikkien eurooppalaisten vauvojen painon kehitystä ja syitä siihen, täsmälleen oikea paino olisi hyödyllistä tietää. Tähän ei kuitenkaan koskaan voida päästä, joten datan tulkinnassa on ymmärrettävä virhemarginaalin tarpeellisuus.

Oleellisuudesta esimerkkinä voidaan antaa linja-autojen sijaintidata. Joillakin paikkakunnilla linja-autoissa on GPS-paikannin, joka kertoo reaaliaikaisesti missä linja-auto kulkee. Kun tämä data sijoitetaan kartalle ja yhdistetään linja-autoaikatauluihin, matkustajat voivat siitä nähdä kuinka kauan menee, että linja-auto saapuu pysäkille, tai onko linja-auto jo ohittanut pysäkin. Tämän datan hyödyllisyys ainakin matkustajan kannalta on siis sidottu ajanhetkeen.

### Kysymys 2 mallivastaus:

- a) Autojen reaaliaikaisella GPS-seurannalla voitaisiin tunnistaa liikenteen pullonkaulakohdat, esimerkiksi tieosuudet joilla on liian paljon liikennettä tien kapasiteettiin nähden jolloin liikenne ruuhkautuu ja hidastuu. Seurannalla voitaisiin myös tukea joukkoliikenteen suunnittelua tunnistamalla useimmin käytettyjä autoreittejä työmatkaliikenteessä ja siten voitaisiin uudet joukkoliikennepalvelut kohdistaa näille reiteille
- b) Autojen reaaliaikaisen GPS-seurannan avulla kaikkien autoilijoiden liikkeitä voitaisiin seurata tarkasti, mikä johtaisi yksityisyyden katoamiseen. Myös datan väärinkäyttö voisi olla ongelma, sillä esimerkiksi keltainen lehdistö haluaisi varmasti mielellään seurata julkisuuden henkilöiden autojen liikkeitä.
- c) Autojen reaaliaikaisen GPS-seurannan teknisiä haasteita ovat esimerkiksi paikannuksen epävarmuus sekä siirrettävän datan suuri määrä. Paikannuksessa voi syntyä epävarmuuksia satelliittisignaalien heikkenemisen vuoksi maanalaisissa tunneleissa tai pysäköintihalleissa, eikä paikannuksessa saa olla suuria heittoa heikosta signaalista huolimatta. Autojen suuresta määrästä johtuen seurannasta kerätään niin suuri datamäärä, että sen turvallinen säilyttäminen verotuksen vaatiman ajan voi tulla ongelmaksi.
- d) Suuresta datamäärästä pitää pystyä jalostamaan tieto yksittäisen ajoneuvon liikkeistä ja ajomäärästä verotusta ja mahdollisia valituksia varten. Epätarkan signaalin ja signaalin katoamisen aiheuttamat katkokset tallennetussa ajoreitissä pitää pystyä käsittelemään verotettavan kannalta edullisimmalla tavalla.

### Kysymys 3 mallivastaus:

1. Datan yksikköä ei ole tallennettu, vain arvo. Esimerkiksi dataan ei ole tallennettu mitä lämpötilan arvo 35 tarkoittaa, ja tämä pitää selvittää muualta kuin itse datasta, esimerkiksi kysymällä dataa hallinnoivilta ihmisiltä tai muualta dokumentaatiosta. Lämpötila voi olla joskus Celsius-, joskus Fahrenheit-asteikolla. Luonnollisesti yksikkö pitäisi lisätä datan yhteyteen tallenteisiin.
2. Valuuttatiedot esimerkkinä ajan suhteen muuttuvista arvoista: on ensinnäkin tiedettävä rahayksikkö, jossa numeerinen valuuttadata on tallennettu. Lisäksi, jos tarvitaan konversioita toiseen valuuttaan, on tiedettävä aika, jolloin valuuttatieto on tallennettu, koska vaihtokertoimet riippuvat ajasta. Tällöin esim. puhelimen hinnasta on tiedettävä hinnan 147 lisäksi valuutta euro, sekä ajankohta 15.6.2014 ja lisäksi paikka, Oulu, Suomi. Vasta näillä tiedoilla varustettuna hintatieto on vertailtavissa ajan ja paikan suhteen muihin tietoihin.
3. Jatkuvasti kerättävä data vanhenee tai on hyödytöntä, jos keräämistä ei jatketa. Esimerkkinä trendien muutokset tweettien kautta seurattuna: on jatkuvasti kerättävä Twitterin kautta dataa, jos halutaan seurata mitkä ilmiöt, brändit, ihmiset tai tapahtumat kiinnostavat ihmisiä, ja miten kiinnostuksen kohteet muuttuvat. On siis huolehdittava datan päivityksestä ja uuden datan tuomisesta mukaan analyysiin sitä mukaa, kun sitä syntyy. Tallennuskapasiteetti saattaa muodostua ongelmaksi, samoin tiedon arkistointi. Analyysistä tulee raskasta. Tarvitaan tehokkaita tietokantoja ja analyysisovelluksia.
4. Aineiston taustat voivat olla erilaiset, esim. jossain aineistossa data voi käsittää vain tietyn populaation tai osan siitä, kun taas toisissa aineistoissa data voi koskettaa ihan erilaista populaatiota. Näin datan analyysissä voidaan tehdä virheitä, jos ei ymmärretä, ettei data ole yhteismitallista populaatioiden erojen takia. Esimerkiksi jos tutkitaan datan avulla puhelimen käyttäjien tottumuksia maailmalla, eikä ymmärretä, että kehittyvissä maissa puhelin ei ole aina henkilökohtainen viestintäväline, vaan puhelimen omistaja saattaa vuokrata puhelintaan muiden käyttöön jatkuvasti. Näin puhelimen käytöstä kertyvä data ei kerro yksilön käyttäytymisestä, toisin kuin rikkaissa maissa. Datan yhteyteen pitää siis jollain tavoin metatietona kirjata millaisesta populaatiosta se on kerätty, jotta aineistot olisivat vertailukelpoisia, tai vaihtoehtoisesti se pitää jotenkin muuten selvittää, kun dataa analysoidaan. Dataa analysoitaessa nämä metatiedot tietysti pitää ottaa huomioon.

## Tehtävä 2: Loppuosataulukko - Vastaukset

### Kysymys 1.

1. Alkuarvot  $alku = 1$  ja  $loppu = 16$ .
2.  $keski = (1 + 16)/2 = 9$ ,  $p > t[S[9]..16]$ ,  $alku = keski + 1 = 10$ .
3.  $keski = (10 + 16)/2 = 13$ ,  $p < t[S[13]..16]$ ,  $loppu = keski - 1 = 12$ .
4.  $keski = (10 + 12)/2 = 11$ ,  $p < t[S[11]..16]$ ,  $loppu = keski - 1 = 10$ .
5.  $keski = (10 + 10)/2 = 10$ ,  $p > t[S[10]..16]$ ,  $alku = keski + 1 = 11$ .
6. Nyt  $alku > loppu$ , joten haku päättyi tuloksettomana. Hahmo  $p =$  "babaa" ei esiinny tekstissä.

### Kysymys 2.

Loppuosa	$i$	$S[i]$
a	1	13
alinta	2	8
eisvalinta	3	4
hteisvalinta	4	2
inta	5	10
isvalinta	6	5
linta	7	9
nta	8	11
svalinta	9	6
ta	10	12
teisvalinta	11	3
valinta	12	7
yhteisvalinta	13	1

Vastauksessa riittää, että arvoja  $S[i]$  kuvaava sarake on annettu.

Kysymys 3.

a) Loppuosa	$i$	$S[i]$
aacat	1	16
aacatcgatagctagaacat	2	1
acat	3	17
acatcgatagctagaacat	4	2
agaacat	5	14
agctagaacat	6	10
at	7	19
atagctagaacat	8	8
atcgatagctagaacat	9	4
cat	10	18
catcgatagctagaacat	11	3
cgatagctagaacat	12	6
ctagaacat	13	12
gaacat	14	15
gatagctagaacat	15	7
gctagaacat	16	11
t	17	20
tagaacat	18	13
tagctagaacat	19	9
tcgatagctagaacat	20	5

Vastauksessa riittää, että arvoja  $S[i]$  kuvaava sarake on annettu.

- b)
1. Alkuarvot  $alku = 1$  ja  $loppu = 20$ .
  2.  $keski = (1 + 20)/2 = 11$ ,  $p > t[S[11]..20]$ ,  $alku = keski + 1 = 12$ .
  3.  $keski = (12 + 20)/2 = 16$ ,  $p < t[S[16]..20]$ ,  $loppu = keski - 1 = 15$ .
  4.  $keski = (12 + 15)/2 = 14$ ,  $p < t[S[14]..20]$ ,  $loppu = keski - 1 = 13$ .
  5.  $keski = (12 + 13)/2 = 13$ ,  $p < t[S[13]..20]$ ,  $loppu = keski - 1 = 12$ .
  6.  $keski = (12 + 12)/2 = 12$ ,  $p$  täsmää loppuosan  $t[S[12]..20]$  kanssa.  
Hahmo  $p = \text{“cga”}$  löydettiin alkaen tekstin kohdasta  $S[12] = 6$ .

**Kysymys 4.**

Tehtävän avainkohta lienee havaita, että annetun loppuosataulukon arvot ovat laskevassa järjestyksessä eli mitä aikaisemmaksi loppuosa sijoittuu aakkosjärjestyksessä, sitä suurempi sen indeksi on. Tai toisin päin: mitä pienempi loppuosan indeksi on, sitä myöhempänä se on aakkosjärjestyksessä. Tämän ominaisuuden sekä tehtävän muut vaatimukset täyttäväksi merkkijonoksi käy mikä tahansa suomalaisen aakkoston merkeistä koostuva 8 merkin pituinen merkkijono  $t$ , jossa mikään merkki ei ole aakkosjärjestyksessä suurempi kuin sitä edeltävä merkki. Tämän voi ilmaista täsmällisemmin esimerkiksi niin, että kaikilla  $i = 2, \dots, 8$  pitää päteä  $t[i - 1] \geq t[i]$ . Tämä on sekä riittävä että välttämätön ehto minkä tahansa tehtävässä sallituista merkeistä koostuvan vastauksen laillisuudelle.

Vastaukseksi kelpaa siten esimerkiksi mikä tahansa merkkijonoista “hgfedcba”, “öäwwnkkk” ja “mmmmmmmm”.



**Kysymys 5.**

Tehtävän ratkaisu on yksikäsitteinen:  $t = \text{“bbaababbababaaba”}$ . Tämä merkkijono voidaan muodostaa esimerkiksi seuraavanlaisella päättelyketjulla:

1. Jos merkkijono olisi  $t = \text{“bbbbbbbbbbbbbbbb”}$ , olisivat loppuosataulukon arvot, kysymyksen 4 tapaan, laskevassa järjestyksessä. Koska tämä ei päde, on merkkijonossa oltava ainakin yksi merkki “a”.
2. Päätelmän 1 perusteella aakkosjärjestyksessä ensimmäisen loppuosan ensimmäisen merkin on oltava  $t[S[1]] = t[16] = \text{“a”}$ .
3. Jos pätesi  $t[15] = \text{“a”}$ , olisi loppuosa  $t[15..16] = \text{“aa”}$  aakkosjärjestyksessä toisena. Näin ei kuitenkaan ole, sillä  $S[2] \neq 15$ , joten  $t[15] = \text{“b”}$ .
4. Kun ainoastaan merkit “a” ja “b” ovat sallittuja, olisi yhden merkin pituinen merkkijono “b” ainoa merkillä “b” alkava loppuosa, joka on aakkosjärjestyksessä aikaisemmin kuin loppuosa  $t[15..16] = \text{“ba”}$ . Koska jo tiedetään, että  $t[16] = \text{“a”}$  eli merkkijono  $t$  ei sisällä loppuosaa “b”, on jokaisen loppuosaa  $t[15..16] = \text{“ba”}$  aakkosjärjestyksessä aikaisemman loppuosan ensimmäisen merkin oltava “a”. Koska  $S[9] = 15$  eli “ba” on aakkosjärjestyksessä 9:s loppuosa, sitä pienemmillä indekseillä  $i = 1, \dots, 8$  pätee  $t[S[i]] = \text{“a”}$ . Nyt saatiin täydennettyä merkkijonoon merkit  $t[13] = \text{“a”}$ ,  $t[3] = \text{“a”}$ ,  $t[14] = \text{“a”}$ ,  $t[11] = \text{“a”}$ ,  $t[9] = \text{“a”}$ ,  $t[4] = \text{“a”}$  ja  $t[6] = \text{“a”}$ .
5. Toisaalta jokaisen loppuosan, joka on aakkosjärjestyksessä loppuosan  $t[15..16] = \text{“ba”}$  jälkeen, on alettava merkillä “b”. Siten indekseillä  $i = 10, \dots, 16$  pätee  $t[S[i]] = \text{“b”}$ . Nyt saatiin asetettua loput puuttuvat merkit  $t[12] = \text{“b”}$ ,  $t[2] = \text{“b”}$ ,  $t[10] = \text{“b”}$ ,  $t[8] = \text{“b”}$ ,  $t[5] = \text{“b”}$ ,  $t[1] = \text{“b”}$  ja  $t[7] = \text{“b”}$ . Lopputuloksena syntyi merkkijono  $t = \text{“bbaababbababaaba”}$ .

## Tehtävä 3: Esimerkkiratkaisut

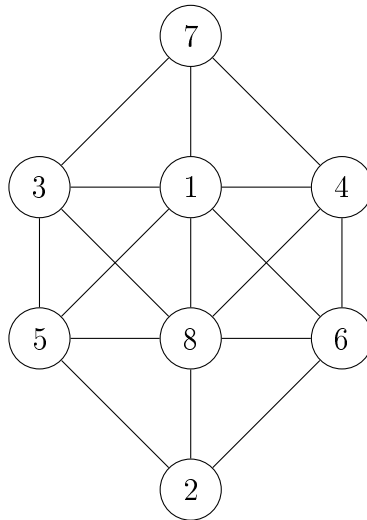
### Kysymys 1.

Oikea vastaus on  $\frac{1}{3}$ . Tämä voidaan päätellä esimerkiksi seuraavalla tavalla:

Kaksi neljästä kortista voidaan valita kuudella eri tavalla: (1,2), (1,3), (1,4), (2,3), (2,4) ja (3,4). Kahdella yhdistelmällä väri on sama, eli (musta, musta) tai (punainen, punainen), loppuilla neljällä värin täytyy olla eri. Kaikkiaan on siis kuusi vaihtoehtoa, joissa kahdessa korteilla on sama väri. Voidaan olettaa kaikki tapaukset yhtä todennäköisiksi, todennäköisyys saada samanväriset kortit on siis  $\frac{2}{6} = \frac{1}{3}$ .

### Kysymys 2.

- (a) Ratkaisusta on neljä variaatiota, loput saadaan kääntämällä ja/tai peilaamalla esitetty ratkaisu.
- 1 ja 8: Kaikilla luvuilla paitsi 1 ja 8 on kaksi vierekkäistä lukua. Kuvaajassa keskimmaisilla ympyröillä on yhteys kaikkiin paitsi yhteen muuhun lukuun, ainoa vaihtoehto on siis sijoittaa 1 ja 8 keskelle.
  - 2:n ja 7:n sijoittamiseen on nyt molemmille vain yksi vaihtoehto
  - loput on jo helppo sijoitella
- (b) Ensimmäisen numeron sijoittamiseen on 8 mahdollisuutta, tämän jälkeen toisen 7 jne. Kaikkiaan vaihtoehtoja on siis  $8 * 7 * \dots * 2 * 1 = 8! = 40320$  kappaletta, joista vain 4 oikeita. Pahimmassa tapauksessa oikeat ratkaisut löytyvät vasta lopuksi, kestää siis  $40320 * 10 = 403200$  sekuntia käydä läpi kaikki vaihtoehdot (tarkkaan ottaen kestää  $40320 * 10 - 30 = 403170$  sekuntia löytää ensimmäinen ratkaisu, tämän huomioiminen ei välttämättömtä). Toisin sanoen tehtävän ratkaisuun kuluu 112 tuntia. Neljä kertaa nopeampi haku olisi myös mahdollinen, jos samaistetaan toisiinsa saman sijoittelun käännettyt ja peilatut versiot, tätä ei tarvitse huomioida saadakseen täydet pisteet.



**Kysymys 3.**

(a) Kahden nopan heitojen tuloksista syntyy  $6 \cdot 6 = 36$  mahdollista yhdistelmää, jotka kaikki yhtä todennäköisiä. Lasketaan kullekin noppaparille, monessako tapauksessa ensimmäinen noppa voittaa, tämä jaettuna 36:lla on voittotodennäköisyys pelaajalle, joka pelaa ensimmäisellä nopalla toista vastaan (huomaa, että annetuilla nopilla tasapelit eivät ole mahdollisia). Noppien vertailu:

- $P(\text{"A voittaa B:n"}) = \frac{15}{36} = \frac{5}{12}$
- $P(\text{"A voittaa C:n"}) = \frac{27}{36} = \frac{3}{4}$
- $P(\text{"B voittaa C:n"}) = \frac{21}{36} = \frac{7}{12}$

Vastaavasti esim.  $P(\text{"B voittaa A:n"}) = 1 - P(\text{"A voittaa B:n"}) = \frac{7}{12}$ . Noppa B on paras valinta, koska sillä voittaa A:n  $\frac{7}{12}$  todennäköisyydellä, ja C:n samoin  $\frac{7}{12}$  todennäköisyydellä.

(b) Ratkaisu ei ole itsestäänselvä:

- Jos lisäämme nopan jolla todennäköisemmin häviää B:lle, pelaaja 1:llä on yhä etu, mikäli hän valitsee B:n.
- Jos lisäämme kopion B:stä, voittotodennäköisyydet menevät tasan jos valinnat B ja C, mutta pelaaja 2 ei vielä saa etua.
- Olisiko olemassa noppa, jolla todennäköisemmin voittaa B:n, mutta pelaaja 1 ei varmistaisi etuaan vaikka valitsisikin sen?

Kuvatkoon mille tahansa kahdelle nopalle merkintä  $X \succ Y$  ominaisuutta "X antaa edun Y:tä vastaan". Edellisen perusteella pöydällä olevilla nopilla toteutuu seuraava:  $B \succ A$ ,  $B \succ C$  ja  $A \succ C$ . Koska B:llä saa edun sekä A:ta että C:tä vastaan, ensimmäinen pelaaja voi varmistaa aina etunsa valitsemalla sen. Luodaan nyt noppa C, jolle voimassa sekä  $C \succ B$ , että  $A \succ C$ , eli luodaan peliin kivi-paperi-sakset tyyppinen suhde noppien välille. Tällainen noppa voidaan luoda vaihtamalla 1 tai 2 arvoa 1 nykyisessä noppa C:ssä arvoksi 4.

Nopalla  $C = 1, 4, 4, 4, 4, 4$  on voimassa

- $P(\text{"A voittaa C:n"}) = \frac{21}{36} = \frac{7}{12}$
- $P(\text{"C voittaa B:n"}) = \frac{25}{36}$

Vastaavasti nopalla  $C = 1, 1, 4, 4, 4, 4$

- $P(\text{"A voittaa C:n"}) = \frac{24}{36} = \frac{2}{3}$
- $P(\text{"C voittaa B:n"}) = \frac{20}{36} = \frac{5}{9}$

Voittostrategia pelaajalle 2:

- Pelaaja 1 valitsee nopan A, Pelaaja 2 nopan B
- Pelaaja 1 valitsee nopan B, Pelaaja 2 nopan C
- Pelaaja 1 valitsee nopan C, Pelaaja 2 nopan A

Tällä strategialla pelaajalla 2 on aina yli 50% voittotodennäköisyys.