

## Tietorakenteet, laskuharjoitus 8, 22-26.3

Seuraavat **TRAKLA2**-deadlinet 21.3. ja 28.3.

Voit valita käyttämäsi ohjelmointikielen vapaasti, paitsi jos teet tehtävän 5, joka on tehtävä Javalla.

Ja jälleen kerran: **älä suorita ohjelmia koneilla melkki, melkinkari, melkinpaasi**

1. Muuta merkkijonot JONO ja PINO kokonaisluvuiksi monisteen sivun 278 menetelmällä. Mitä merkkijonoa vastaa kokonaisluku 17625834700?

Seuraava Java-koodi kertoo merkin ASCII-koodin:

```
char merkki = 'A';
int koodi = (int)merkki;
System.out.println(koodi);
```

ASCII-koodit on listattu myös esimerkiksi seuraavalla sivustolla:

<http://www.asciitable.com/>

2. Toteuta haluamallasi ohjelmointikielillä hajautusfunktio, joka kuvaa merkkijonon
  - (a) kokonaisluvuksi  $i \in \{0, 1, 2, 3, 4, 5, \dots, 2^j - 1\}$  missä  $j$  on kokonaisluku
  - (b) kokonaisluvuksi  $i \in \{0, 1, 2, 3, 4, 5, \dots, m - 1\}$  missä  $m$  on alkuluku
  - (c) siten, että se sopii käytettäväksi kaksoishajautuksessa toisena hajautusfunktiona yhdessä edellisen kohdan funktion kanssa

3. Toteuta haluamallasi ohjelmointikielillä hajautustaulu ylivuotoketjuja tai avointa hajautusta käyttäen. Voit olettaa, että hajautustaulun avaimet ovat merkkijonoja. Tässä tehtävässä riittää toteuttaa operaatiot **hash-search** ja **hash-insert**. Riittää, että taulukkosu pysyy koko ajan samankokoisena, eli uudelleenhajautusta ei tarvita.

Yleensä sama avain talletetaan hajautustauluun vain kertaalleen. Jos näin toimitaan on hash-insertin yhteydessä pakko käydä koko ylivuotoketju läpi. Tällöin hash-insert ei ole enää parhaimman tapauksen vaatimuudeltaan vakioaikainen.

4. Testaa edellisen tehtävän hajautustauluasi lisäämällä sinne Aleksis Kiven romaanin Seitsemän veljestä kaikki sanat. Romaani on saatavilla tekstimuodossa seuraavassa osoitteessa:

<http://www.gutenberg.org/files/11940/11940-8.txt>

Kokeile erilaisia hajautusfunktioita. Jos toteutuksesi perustuu ketjutukseen, selvitä mikä on eri tapauksissa pisin ylivuotoketjun pituus ja keskimääräinen ylivuotoketjun pituus. Miten taulukon koko vaikuttaa ketjujen pituuksiin?

Jos toteutuksesi perustuu avoimeen hajautukseen, tarkastele suorituskykyä eri mittaisilla taulukoilla ja eri törmäyksenhallintastrategioilla.

Tiedoston lukeminen ja sen käsittely sana sanalta onnistuu Javalla seuraavasti:

```
import java.io.File;
import java.util.Scanner;

public class Main {
    public static void main(String[] args) throws Exception {
        // avataan tiedosto seitseman.txt
```

```

Scanner teksti = new Scanner(new File("seitseman.txt"));
// tai          = Scanner(new File("seitseman.txt"), "ISO-8859-1");
// jos tiedosto on ISO-8859-1-formaatissa

// jatketaan niin kauan kuin sanoja jäljellä
while ( teksti.hasNext() ) {
    String mj = teksti.next(); // haetaan seuraava sana
    System.out.println(mj);
}
}
}

```

5. Vertaa oman hajautustaulutaulusi, Javan hajautustaulun HashSet:in ja Javan tasapainoitettun hakupuun TreeSet:in suorituskykyä. Jos et tehnyt edellistä tehtävää, saat tästä rastin vertaamalla HashSet:iä ja TreeSet:iä ja tekemällä bonuksista ainakin ensimmäisen kohdan. Ohje suorituskykymittauksille löytyy laskareiden 4 tehtäväpaperissa.

Tee seuraavat mittaukset:

- Lisää rakenteiseen Seitsemän veljeksien sanat ja mittaa tähän kuluva aika
- Edellisen jälkeen testaa jokaisesta Dostojevskin romaanin Rikos ja rangaistus englanninkielisen käännöksen sanasta löytyykö se Seitsemän veljeksien sanojen joukosta ja mittaa tähän kuluva aika. Käy läpi kaikki Rikoksen ja rangaistuksen sanat alusta loppuun, sama sana saa siis toistua kyselyssä niin monta kertaa kuin se tulee tekstissä vastaan.

Rikos ja rangaistus löytyy osoitteesta:

<http://www.gutenberg.org/files/2554/2554-8.txt>

Bonuskysymykset:

- TreeSet ylläpitää avaimia järjestyksessä. TreeSet:issä olevat avaimet on mahdollista käydä järjestyksessä läpi *iteraattorin* (engl. iterator) avulla. Selvitä miten iteraattori toimii ja tulosta Seitsemän veljeksien sanat aakkosjärjestyksessä.  
Myös HashSetin sisältämät avaimet on mahdollista käydä iteraattorilla läpi. Mikä on tällöin avainten järjestys?
  - Tulosta Seitsemän veljeksien ja Rikoksen ja rangaistuksen yhteiset sanat. Tulosta jokainen yksittäinen sama sana vain kertaalleen.
6. Hajautustaulun avaimina halutaan käyttää binääripuita. Voit olettaa, että binääripuun solmuihin on talletettu kokonaislukuja. Esitä menetelmä, jolla binääripuun voi muuttaa kokonaisluvuksi. Perustele sitten seuraavat asiat menetelmästäsi: Voivatko kaksi eri binääripuuta saada arvokseen saman kokonaisluvun? Vastaako jokaista kokonaislukua jokin binääripuu?