

## Tehtävä: Lentävät lautaset

National UFO Reporting Center (NUFORC) ylläpitää tunnistamattomien lentävien esineiden havaintotietokantaa. Havaintotietokanta sisältää tiedon havaintoajasta, paikasta, osavaltiosta, maasta, tunnistamattoman lentävän esineen muodosta, havainnon kestosta, koordinaateista sekä hieman tarkemmasta kuvauksesta.

Tehtävänäsi on tehdä sovellus, joka saa syötteenä ufohavaintoja sisältävän tiedoston, ja tulostaa havaintoihin liittyviä tilastoja.

Tiedoston muoto on seuraava:

```
10/10/1974 23:00;hudson;ks;us;light;1200;38.1055556;-98.6597222;The light chased us.
10/10/1984 05:00;houston;tx;us;circle;60;29.7630556;-95.3630556;2 experience with unkown
10/10/2002 06:00;seattle;wa;us;sphere;90;47.6063889;-122.3308333;Light falls from sky
10/10/2007 01:00;stockbridge;ga;us;changing;3600;33.5441667;-84.2338889;i was abducted
10/10/2009 20:30;anaheim;ca;us;light;300;33.8352778;-117.9136111;UFO over Disneyland.
10/15/2009 22:50;lancaster;sc;us;flash;1.5;34.7202778;-80.7711111;Bright blue orb.
10/15/2009 23:00;dothan;al;us;flash;1200;31.2230556;-85.3905556;blinking lights
6/5/2004 03:20;polk (south of);mo;us;unknown;20;37.7291667;-93.2991667;Object/being
```

Jokainen tiedoston rivi sisältää yhden havainnon. Rivin sisältö on tarkemmin seuraava:

```
String r = "10/10/1974 23:00;hudson;ks;us;light;1200;38.1055556;-98.6597222;The light chased us."
String[] palat = r.split(";");

System.out.println(palat[0]); // sisältää kk/pp/vvvv hh:mm -muodossa olevan päivämäärän
System.out.println(palat[1]); // sisältää paikan
System.out.println(palat[2]); // sisältää osavaltion
System.out.println(palat[3]); // sisältää maan
System.out.println(palat[4]); // sisältää lentävän esineen muodon
System.out.println(palat[5]); // sisältää havainnon keston sekunteina (myös puolikkaita)
System.out.println(palat[6]); // sisältää pituusasteen
System.out.println(palat[7]); // sisältää leveysasteen
System.out.println(palat[8]); // sisältää vaillinaisen tekstuaalisen kuvauksen tapahtumasta
```

Datassa saattaa esiintyä myös epäselvyyksiä. Esimerkiksi yllä olevassa esimerkissä lentävän esineen muoto on tuntematon. Huomaa myös, että päivämäärät esitetään muodossa kuukausi/päivä/vuosi tunti:minuutti.

Ohjelman tulee aluksi kysyä käyttäjältä havainnot sisältävän tiedoston nimeä. Tämän jälkeen ohjelma tulostaa (alempana olevan esimerkin tapaan muotoiltuna) seuraavat tiedot:

- Havaintojen lukumäärä
- Lyhimmän havainnon kesto sekunteina
- Pisimmän havainnon kesto sekunteina
- Havainnon keskimääräinen kesto sekunteina
- Lentävien esineiden muodot
  - Muodot tulee tulostaa aakkosjärjestyksessä
  - Jokaisen esineen yhteydessä tulee kertoa myös kyseiseen esineeseen liittyvien havaintojen lukumäärä
- Muoto, johon liittyy eniten havaintokertoja
- Havaintojen lukumäärät vuosittain
- Havaintojen lukumäärät ja prosentuaaliset osuudet tunneittain
- Lentävien esineiden muodot havaintojen keston pituuden mukaan ryhmiteltynä
  - Havainnot tulee jakaa havainnon keston mukaan kolmeen yhtä suureen tai lähes yhtä suureen ryhmään
  - Tämän jälkeen tulostetaan esinekohtaiset havainnot kussakin ryhmässä
- Havaintojen lukumäärät keston pituuden ja tunnin mukaan ryhmiteltynä
  - Havainnot tulee jakaa havainnon keston mukaan kolmeen yhtä suureen tai lähes yhtä suureen ryhmään
  - Tämän jälkeen tulostetaan tuntikohtaiset havaintojen lukumäärät kussakin ryhmässä
- Yleisimmät tekstuaalisessa kuvauksessa esiintyvät sanat sanojen esiintymiskertojen mukaan järjestettynä
  - Käsittele kaikki sanat pienessä kirjoitusasussa (String-luokan metodi `toLowerCase` auttaa tässä)
  - Poista erikoismerkit kuten `./?;-;)(*-_` sekä numerot
  - Tulosta vain 15 yleisintä sanaa. Jos sanoja on vähemmän kuin 15, tulosta kaikki.
    - Jos sanan yleisyys ei ole yksikäsitteistä (useampia sanoja esiintyy esimerkiksi sama määrä), voit valita sanat saman määrän esiintyvistä sanoista haluamallasi tavalla.
- Yleisimmät tekstuaalisessa kuvauksessa esiintyvät sanaparit sanaparien esiintymiskertojen mukaan järjestettynä.
  - Seuraa samoja ohjeita kuin edellä.

- Tulosta vain 10 yleisintä sanaparia. Jos sanapareja on vähemmän kuin 10, tulosta kaikki.

Ohjelman tulee toimia sekä pienellä että isoilla tiedostoilla. Käytössäsi on kaksi tiedostoa (havainnot-mini.csv ja havainnot.csv), jotka löytyvät näyttökokeessa käytettävän koneen työpöydältä kansioista Data.

Tulostusesimerkki seuraavalla sivulla.

tiedosto: /home/cs-mooc/Desktop/Data/havainnot-mini.csv

Havainnot: 8

Lyhimmän havainnon kesto sekunteina: 3.0

Pisimmän havainnon kesto sekunteina: 3600.0

Keskimääräinen havainnon kesto sekunteina: 809.125

Esineiden muodot

changing, havainnot: 1

circle, havainnot: 1

flash, havainnot: 2

light, havainnot: 2

sphere, havainnot: 1

unknown, havainnot: 1

Eniten havainnot: flash, 2 kertaa

Havainnot vuosittain

1974, havainnot: 1

1984, havainnot: 1

2002, havainnot: 1

2004, havainnot: 1

2007, havainnot: 1

2009, havainnot: 3

Havainnot tunneittain

1-2, havainnot: 1, 12.5%

3-4, havainnot: 1, 12.5%

5-6, havainnot: 1, 12.5%

6-7, havainnot: 1, 12.5%

20-21, havainnot: 1, 12.5%

22-23, havainnot: 1, 12.5%

23-24, havainnot: 2, 25.0%

Muodot havainnon keston mukaan ryhmiteltynä

lyhyt: 0-60.0, havainnot: 3

keskipitkä: 60.0-1200.0, havainnot: 2

pitkä: 1200.0-3600.0, havainnot: 3

muoto	lyhyt	keskipitkä	pitkä
-------	-------	------------	-------

changing	0	0	1
----------	---	---	---

circle	1	0	0
--------	---	---	---

flash	1	0	1	
light	0	1	1	
sphere	0	0	1	0
unknown	1	0	0	0

Tunnit havainnon keston mukaan ryhmiteltynä

	tunti	lyhyt	keskipitkä	pitkä
1-2	0	0	1	
3-4	1	0	0	
5-6	1	0	0	
6-7	0	1	0	
20-21	0	1	0	
22-23	1	0	0	
23-24	0	0	2	

Yleisimmät sanat esiintymiskertojen mukaan järjestettynä

falls, 1  
 from, 1  
 i, 1  
 lights, 1  
 object, 1  
 orb, 1  
 over, 1  
 sky, 1  
 the, 1  
 ufo, 1  
 unkown, 1  
 us, 1  
 was, 1  
 with, 1  
 light, 2

Yleisimmät sanaparit esiintymiskertojen mukaan järjestettynä

disneyland bright, 1  
 experience with, 1  
 falls from, 1  
 from sky, 1  
 i was, 1  
 light chased, 1  
 light falls, 1  
 lights object, 1  
 object being, 1  
 orb blinking, 1

Jos käyttäjän syöttämää tiedostoa ei ole olemassa tai sitä ei voida lukea, ohjelman tulee ilmoittaa tästä ja lopettaa toimintansa ilman virhetilanteita.

tiedosto: **havainnot.txt**

tiedostoa ei voi avata