
SANAINDEKSOIJAN KÄYTTÖOHJEET

1. Yleistä

Tämä sovellus on tarkoitettu merkkijonojen etsimiseen tekstitiedostojen sisällöstä. Sovelluksella luetaan sisään haluttu määrä tekstitiedostoja ja etsitään niiden sisällöstä haluttuja merkkijonoja. Hakutuloksia on mahdollista tarkentaa tekemällä uusia merkkijonohakuja, jotka kohdistuvat aiempaan hakutulokseen. Merkkijonot tallennetaan hakupuuhun, joka mahdollistaa nopeat haut.

Sovellus käsittää tallennettaviksi ja haettaviksi merkeiksi suomalaiset aakkoset (a-ö), numerot (0-9) sekä sanojen välissä olevat tyhjät merkit. Mikäli tekstistä tai haun merkkijonosta löydetty merkki ei kuulu näihin, sovellus ohittaa sen. Esimerkkinä merkkijono **(Marja-Liisako?)**, jota sovellus käsittelee merkkijonona **'marjaliisako'**.

Sovelluksen käyttöliittymä on merkkipohjainen.

2. Sovelluksen käynnistäminen

Käynnistäminen NetBeansista:

Sovellus on toteutettu NetBeans-toteutusvälineellä, josta sen main-metodin sisältävä Sanaindeksoija.java-luokka voidaan käynnistää hakemistosta **/Tiraharj/Sanaindeksoija/Source Packages/Kayttoliittyma**. Tällöin sisäänluettavien tiedostojen tulee sijaita hakemistossa **/Tiraharj/Sanaindeksoija** **tai** kunkin tiedoston sijainti pitää antaa polkuna ohjelman kysyessä tiedoston nimeä.

Käynnistäminen komentotulkista:

Komentotulkista sovellus käynnistyy komennolla

```
java -jar "</hakemisto/hakemisto/>Sanaindeksoija.jar"
```

, jossa **</hakemisto/hakemisto/>**-kohdan tilalle kirjoitetaan se hakemistopolku, johon käynnistystiedosto Sanaindeksoija.jar on tallennettu, esimerkiksi

/Tiraharj/Sanaindeksoija/dist/. Jos käynnistysketkellä ollaan Sanaindeksoija.jar-tiedoston sijaintihakemistossa, ei polkua tarvitse määrittää. Ennen sovelluksen käynnistämistä tulee sisäänluettavat tekstitiedostot tallettaa

</hakemisto/hakemisto/>-kohdassa määritetyn polun juurihakemistoon. Jos käynnistys tehdään sijaintihakemistosta, tulee sisäänlukutiedostot tallentaa samaan hakemistoon kuin Sanaindeksoija.jar. Jos tiedostot sijaitsevat muualla, on oikea hakemisto mahdollista antaa ohjelmalle tiedoston nimen antamisen yhteydessä.

3. Tiedostojen sisäänluku

Sovellus kysyy käyttäjältä sisäänluettavien tekstitiedostojen lukumäärän. Jos käyttäjän antama vastaus ei ole luku, antaa sovellus virheilmoituksen ja pyytää antamaan lukumäärän uudelleen. Jos käyttäjä antaa lukumääräksi nollan (0), päättyy ohjelma.

```
Tervetuloa Sanaindeksoijaan!  
Anna sisäänluettavien tiedostojen lukumäärä:  
2
```

Kuva 1. Tiedostojen lukumäärä

Tämän jälkeen sovellus pyytää käyttäjää antamaan tekstitiedoston nimet yhden kerrallaan. Jos annetun nimistä tekstitiedostoa ei löydy sovelluksen hakemistosta, antaa sovellus tästä virheilmoituksen. Kun sovellus on lukenut sisään käyttäjän ilmoittaman määrän tekstitiedostoja, siirtyy ohjelma automaattisesti eteenpäin. Jos käyttäjä haluaa keskeyttää sisäänluvut ennen kuin ilmoitettu määrä tiedostoja on luettu sisään sovellukseen, onnistuu se antamalla tiedoston nimeksi tyhjän.

```
Anna 1. sisäänluettavan tiedoston nimi:  
testi3.txt
```

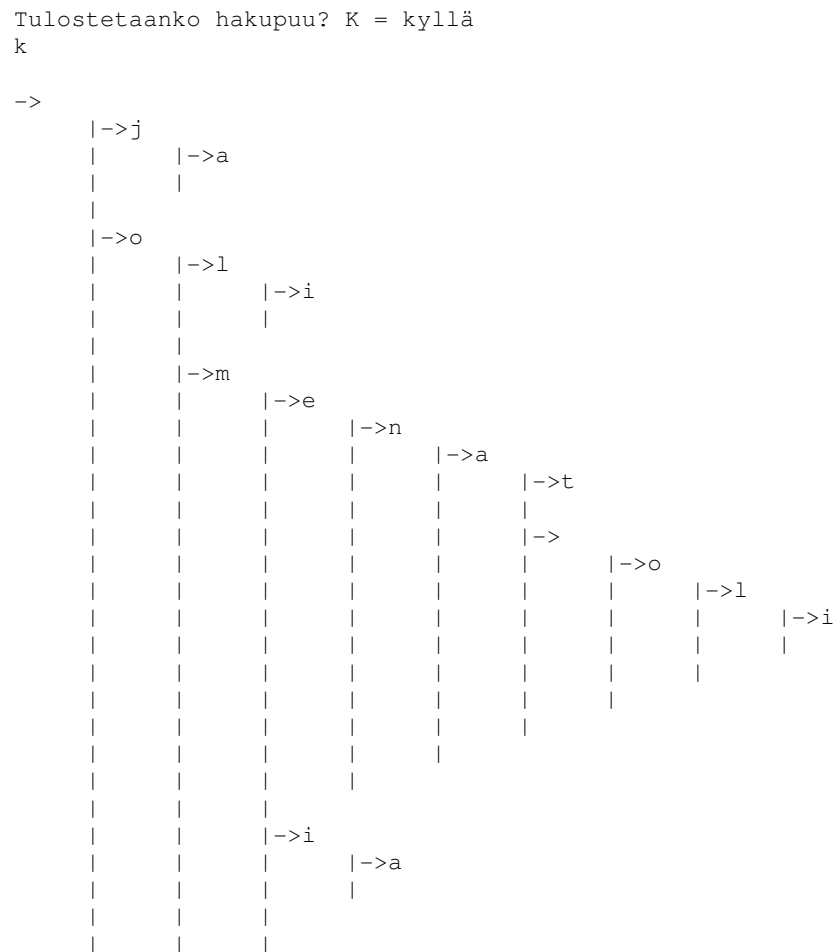
Kuva 2a. Tiedoston nimen antaminen ilman hakemistopolkua (sijaitsevat käynnistyksessä annetun polun juuressa)

```
Anna 1. sisäänluettavan tiedoston nimi:  
/cs/fs/home/ninbarlu/Tiraharj/testi3.txt
```

Kuva 2b. Tiedoston nimen antaminen hakemistopolun kanssa (voivat sijaita muualla kuin käynnistyspolulla)

Tekstitiedostojen sisältö tallennetaan hakupuuhun kunkin sanan alusta alkaen kyseisen tekstirivin loppuun asti. Uusi sana alkaa aina rivin alusta tai tyhjän merkin jälkeen. Tallennus jättää huomioimatta muut merkit paitsi suomalaisen aakkoston ja numerot sekä tallennettavien sanojen välissä olevat tyhjät merkit. Esimerkiksi tekstirivi **Omena? Appelsiini! Kirsik-** tallennetaan hakupuuhun kolmena merkkijonona, joita voi hakea kunkin merkkijonon alusta lähtien: **'omena appelsiini kirsik'**, **'appelsiini kirsik'** ja **'kirsik'**.

Jokaisen tekstitiedoston sisäänluvun jälkeen sovellus tiedustelee käyttäjältä, tulostetaanko siihen asti toteutettu hakupuu. Vastaamalla kysymykseen K (tai k) hakupuu tulostuu näytölle (suositeltavaa vain pienillä tiedostoilla).



Kuva 3. Hakupuun tulostaminen

4. Hakutoiminto

Kun tekstitiedostot on luettu sisään sovellukseen, siirtyy käyttäjä hakutoimintoon. Hakutoiminnossa käyttäjä voi hakea sisäänluettujen tekstitiedostojen riveiltä haluamiaan merkkijonoja. Sovellus pyytää käyttäjältä merkkijonon ja hakee sitä tekstitiedostojen tiedoista sekä tulostaa näytölle löytyneet tekstitiedostojen rivit sekä niiden yhteenlasketun lukumäärän.

Hakutoiminto päättyy, kun merkkijonoksi annetaan tyhjä.

```
Siirrytään hakutoimintoon.
```

```
Aloitetaan uusi haku. Tyhjä päättää hakutoiminnon.  
Anna haettava merkkijono:  
ei a
```

```
rautatie.txt, rivi 16: Ei ole harakka vielä aamiaista saanut, ei aukea  
rautatie.txt, rivi 596: --Vai ei annettu hevosta?  
rautatie.txt, rivi 597: --Ei annettu, enkä minä häntä hyvin kättänyt  
rautatie.txt, rivi 1124: --Kyllä se on niin, että siinä ei auta nyt  
rautatie.txt, rivi 1127: --Sitä, että sanoppa mitä tahansa, niin siinä  
ei auta mikään ... ei  
rautatie.txt, rivi 2551: Yhdentyylinen oli takki ja lakki kuin  
inspektorilla, vaikka ei aivan  
rautatie.txt, rivi 2581: Ja Matti ja Liisa ajattelivat, että nyt siinä  
ei auta mikään ja että nyt  
yhteensä 7 osumaa
```

Kuva 4. Haku

Käyttäjän antamasta merkkijonosta käsitellään vain suomalaisen aakkoston merkit, numerot ja sanojen välissä olevat tyhjät merkit, muut merkit ohitetaan kuten tallennusvaiheessa. Esimerkiksi haku **olik** tuo tulokseksi rivit, joissa esiintyy sana **oliko** tai **olikin**, mutta haku **'oli k'** antaa tulokseksi rivin, josta löytyy sanapari **'oli koleaa'**. Nimihaku **Marja-Leena** puolestaan löytää myös rivit, joilla nimi on kirjoitettu myös muotoon **Marjaleena**, mutta ei rivejä, joilla nimi on muodossa **Marja Leena**.

Käyttäjä voi halutessaan tarkentaa saamaansa hakutulosta tekemällä uuden haun, joka kohdistuu edelliseen hakutulokseen. Tällöin haku antaa tulokseksi vain ne rivit, joissa esiintyy kaikissa hauissa annetut merkkijonot.

```
Jos haluat tarkentaa hakua, anna uusi merkkijono. Tyhjä päättää  
tarkennuksen.  
s
```

```
rautatie.txt, rivi 16: Ei ole harakka vielä aamiaista saanut, ei aukea  
rautatie.txt, rivi 1124: --Kyllä se on niin, että siinä ei auta nyt  
rautatie.txt, rivi 1127: --Sitä, että sanoppa mitä tahansa, niin siinä  
ei auta mikään ... ei  
rautatie.txt, rivi 2581: Ja Matti ja Liisa ajattelivat, että nyt siinä  
ei auta mikään ja että nyt  
yhteensä 4 osumaa
```

Kuva 5. Tarkennettu haku

Tarkennettu haku päättyy, kun merkkijonoksi annetaan tyhjä. Tällöin hakutoiminto aloittaa uuden haun koko materiaalista. Tarkennetusta hausta siirrytään suoraan uuteen koko materiaalin hakuun myös silloin, kun hakutulos on tyhjä.

5. Tilaston tulostus

Kun hakutoiminto päättyy, siirrytään tilaston tulostukseen. Sovellus ilmoittaa käyttäjälle seuraavat tiedot:

- sisäänluettujen tiedostojen lukumäärän
- sisäänluettujen tekstirivien lukumäärän ja keskiarvon per tiedosto
- tehtyjen hakujen lukumäärän: tähän lasketaan ne haut, jotka on tehty kohdistuen koko materiaaliin
- tehtyjen alahakujen eli tarkennusten lukumäärän ja keskiarvon per varsinainen haku
- hauissa tulleiden osumien lukumäärän ja keskiarvon per varsinainen haku: määrä lasketaan hakutuloksista alahakujen jälkeen

Tilasto:

- sisäänluettuja tiedostoja 2 kpl
- sisäänluettuja rivejä 2997 kpl, k.a per tiedosto 1498.5
- tehtyjä hakuja 1 kpl
- tehtyjä alahakuja 1 kpl, k.a per haku 1.0
- saatuja osumia yhteensä 4 kpl, k.a per haku 4.0

Kuva 6. Tilaston tulostus

Tilaston tulostus päättää ohjelman.