

Harjoituksia säännöllisistä lausekkeista

1. UNIX:in `egrep`-komennolla (extended grep) voi etsiä tekstistä hahmoja, jotka on määritelty säännöllisinä lausekkeina. `egrep`in perussyntaksi on seuraava: `egrep <lauseke> <tiedosto>`, missä lauseke voi olla
 - hakasuluissa lista merkkejä, esim. `[abcd]`: mikä tahansa merkeistä a, b, c, d
 - `(lauseke)(lauseke)`: kahden lausekkeen katenaatio
 - `(lauseke1)|(lauseke2)`: joko lauseke1 tai lauseke2
 - `(lauseke)*`: lauseke toistuu 0 kertaa tai useampia kertoja (sulkeuma)
 - `\b`: tyhjä merkki sanan reunassa `\B`: tyhjä merkki sanan keskellä

Huom! Lauseke kannattaa laittaa hipsuihin ('lauseke'). Lisää tietoa komennolla `man egrep`.

Testaa `egrep`-komentoa! Voit käyttää syötetiedostona mitä tahansa (C-)ohjelmakoodia, esim. kotisivun tiedostoa esim. `c`.

Millaisia hahmoja etsivät seuraavat komennot:

```
egrep '[1]*'  
egrep '[1][0]'  
egrep '[1]||[0]'
```

2. Millaisella `egrep`-komennolla löydät seuraavat rivit:
 - a) Rivit, joilla on numeroita
 - b) Rivit, joilla esiintyy sana *while* tai *for*
 - c) Rivit, joilla esiintyy numero 10
 - d) Rivit, joilla esiintyy kokonaislukuja. (Huom! Kommentosi ei siis saa hyväksyä desimaalilukuja.)
3. Tarkastellaan seuraavia aakkoston $\Sigma = \{a, b\}$ kieliä. Anna kustakin kielestä kaksi merkkijonoa, jotka kuuluvat kieleen, ja kaksi, jotka eivät kuulu kieleen!
 - a) a^*b^*
 - b) $a(ba)^*b$
 - c) $a^* \cup b^*$

- d) $(aaa)^*$
- e) $(\epsilon \cup a)b$
- f) $\Sigma^*a\Sigma^*a\Sigma^*a\Sigma^*$

4. Mitä merkkijonoja kuuluu seuraavaan lausekkeen kuvaamaan kieleen?
 $(c \cup h \cup m \cup r)at((c \cup t)a \cup (s \cup t)o)ught(m \cup l \cup tw \cup r)ice$
5. Kuinka määrittelet säännöllisen lausekkeen, jolla löydät tietoa lumimyrskystä? Huomaa, että ilmaisu voi esiintyä myös taipuneena (sanavartalo ei onneksi taivu) tai yhdistettynä, esim. ”lumi- ja ukkosmyrskyvaroitus”.
6. Mitä merkkijonoja kuuluu kieleen $L(\emptyset^*)$? Entä $L(\epsilon^*)$?
7. Etsi lyhin merkkijono, joka kuuluu seuraavan lausekkeen kuvaamaan kieleen!

- a) $a^*(b \cup abb)b^*b$
- b) $a^*b^*b(a \cup (ab)^*)^*b^*$
- c) $(a \cup ab)(a^* \cup ab)^*b$

8. Muodosta seuraavia kieliä vastaavat säännölliset lausekkeet:
 - a) $\{w \in \{a, b\}^* | w:n \text{ kolmanneksi viimeinen merkki on } a\}$
 - b) $\{w \in \{a, b\}^* | w \text{ sisältää joko merkkijonon } ab \text{ tai } ba\}$
 - c) $\{w \in \{a, b\}^* | w \text{ sisältää merkkijonon } aba \text{ mutta ei merkkijonoa } bab\}$
9. Muodosta seuraavia kieliä vastaavat säännölliset lausekkeet:
 - a) $\{w \in \{a, b\}^* | w \text{ sisältää parillisen määrän merkkiä } a\}$
 - b) $\{w \in \{a, b\}^* | w:n \text{ pituus on pariton}\}$
 - c) $\{w \in \{a, b\}^* | w:n \text{ sisältämien } b\text{-merkkien lukumäärä on kolmella jaollinen}\}$
10. Esitä yksikertaisemmassa muodossa seuraavat lausekkeet! (s.e. ne yhä generoivat saman kielen!)
 - a) $(0 \cup 1 \cup 01 \cup 11)^*$
 - b) $(0^* \cup 10^*)^*$
 - c) $1^*(011^*)^* \cup 1^*(011^*)^*0$