

Harjoitus 2

1. UNIX:in egrep-komennolla (extended grep) voi etsiä tekstistä hahmoja, jotka on määritelty säännöllisinä lausekkeina. egrepin 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 kotisivun tiedostolla esim.c (<http://www.cs.joensuu.fi/pages/whamalai/tepe04/esim.c>)!

Millaisia hahmoja löydät seuraavalla komennolla?
`egrep '(uu)|(sata)|(issa)|(oita)|(alla)'` esim.c

Keksitkö yksinkertaisemman hahmon, joka tulostaisi vain runon rivit?

2. Millaisella egrep-komennolla löydät tiedostosta esim.c 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. Mitä merkkijonoja kuuluu kieleen $L(\emptyset^*)$? Entä $L(\epsilon^*)$?

6. 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) $(a \cup ab)(a^* \cup ab)^*b$

7. 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) $\{w \in \{a, b\}^* | w \text{ sisältää merkkijonon } aba \text{ mutta ei merkkijonoa } bab\}$

8. 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) $\{w \in \{a, b\}^* | w:n \text{ sisältämien } b\text{-merkkien lukumäärä on kolmella jaollinen}\}$

9. Lue satu päätösongelmista

<http://www.cs.joensuu.fi/pages/whamalai/tepe04/satu.html> ja täydennä se loppuun!