Puude teisenduskauguse arvutamine

Raivo Laanemets

26. mai 2009. a.

Probleemi kirjeldus

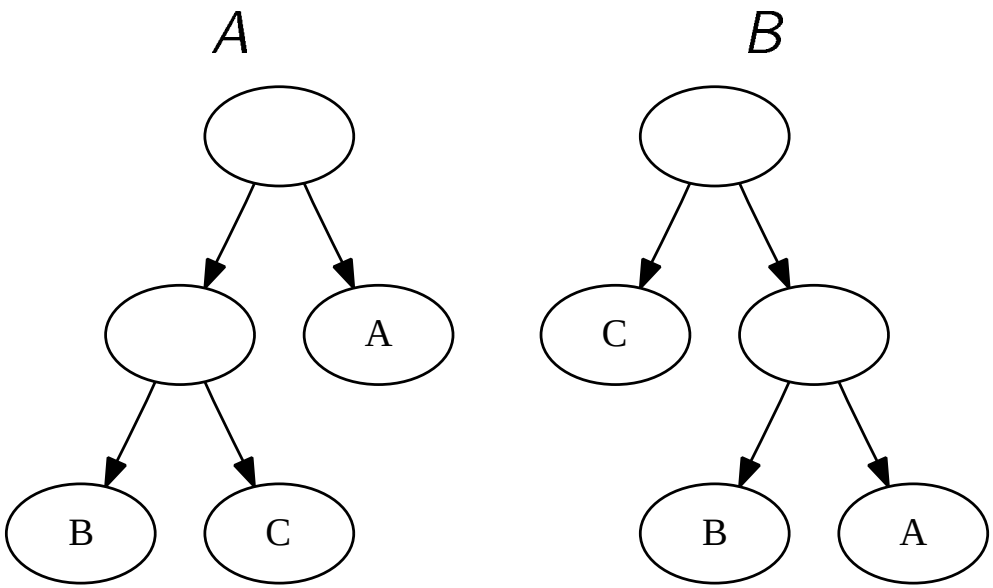
- Antud on kaks puud A ja B.
- Puude peal saab teostada operatsioone:
  - Tipu kustutamine
  - Lisamine
  - Ümbernimetamine
  - Alampuude ümbertõstmine
- Leida lühim/vähima kaaluga selline op. jada, mis teisendab puu A puuks B.

Võimalikud lahendused

Puid on kahte tüüpi:

- *ordered tree* - tipu alamate järjestus on oluline
  - Teada on $O(n^4)$ kiiruse ja mäluga algoritmid
  - Dünaamiline programmeerimine metsade salvestamisega
- *unordered tree* - tipu alamate järjestus on mitteoluline
  - Üldisel juhul NP-täielik probleem
  - Toore jõuga lahendamine
  - Tehisintellekti algoritmid (A*)

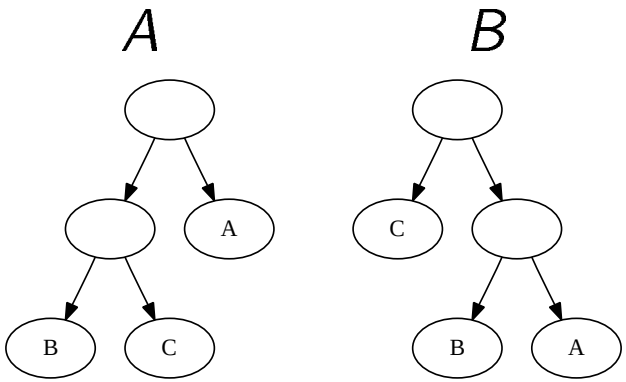
Näide 1



Projektist

- Algoritm ja realisatsioon *unordered*-tüüpi fülogeneetiliste puude jaoks
- Fülogeneetiline puu
  - Täielik kahendpuu
  - Andmed lehtedes
  - Alampuude järjestus pole oluline
  - Salvestatud NHX formaadis: `(A,(B,C))`
- Rakendus: võrrelda kahe erineva meetodiga saadud puude "sarnasust"
- Algoritm kasutab alt-üles lähenemist
- Üritab mõlemad puud ühtlustada kuni alles jääb üks tipp mõlemas puus
- Kaks operatsiooni: lehtede ühendamine ja lehtede ümbervahetamine

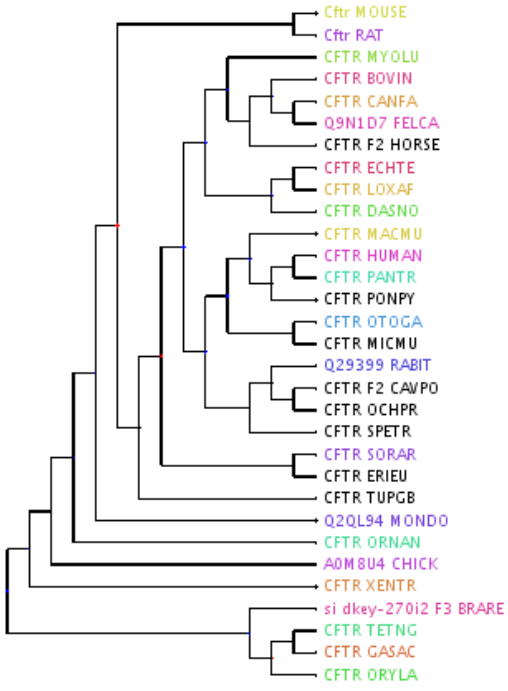
Näide 2



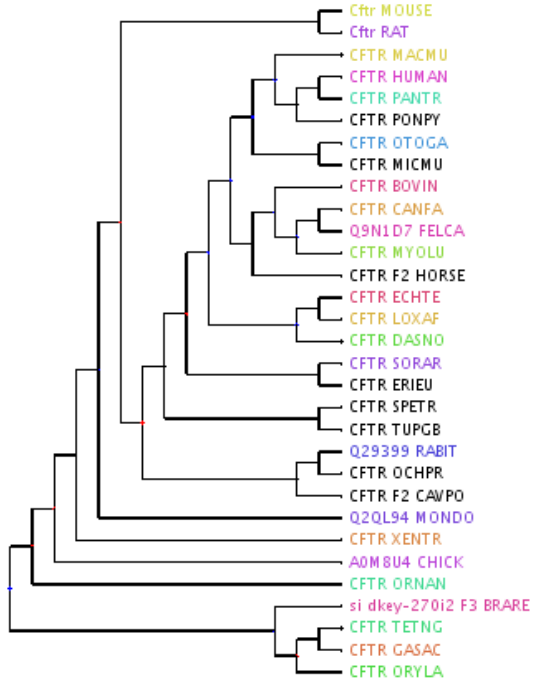

```
td.sh -in1 data/test/A.txt -in2 data/test/B.txt
Swap A and B in first tree
Merge A and C in both trees into L0
Merge B and L0 in both trees into L1
Swaps: 1
```

Näide 3

A


Cftr MOUSE
Cftr RAT
CFTR MYOLU
CFTR BOVIN
CFTR CANFA
Q9N1D7 FELCA
CFTR F2 HORSE
CFTR ECHTE
CFTR LOXAF
CFTR DASNO
CFTR MACMU
CFTR HUMAN
CFTR PANTR
CFTR PONPY
CFTR OTOGA
CFTR MICMU
Q29399 RABIT
CFTR F2 CAVPO
CFTR OCHPR
CFTR SPETR
CFTR SORAR
CFTR ERIEU
CFTR TUPGB
Q2QL94 MONDO
CFTR ORNAN
A0M8U4 CHICK
CFTR XENTR
si dkey-270i2 F3 BRARE
CFTR TETNG
CFTR GASAC
CFTR ORYLA

B


Cftr MOUSE
Cftr RAT
CFTR MACMU
CFTR HUMAN
CFTR PANTR
CFTR PONPY
CFTR OTOGA
CFTR MICMU
CFTR BOVIN
CFTR CANFA
Q9N1D7 FELCA
CFTR MYOLU
CFTR F2 HORSE
CFTR ECHTE
CFTR LOXAF
CFTR DASNO
CFTR SORAR
CFTR ERIEU
CFTR SPETR
CFTR TUPGB
Q29399 RABIT
CFTR OCHPR
CFTR F2 CAVPO
Q2QL94 MONDO
CFTR XENTR
A0M8U4 CHICK
CFTR ORNAN
si dkey-270i2 F3 BRARE
CFTR TETNG
CFTR GASAC
CFTR ORYLA

Näide 3 ja tulemused

- Suurte puude korral algoritm optimaalset teisenduskaugust ei leia
- Randomiseeritud variant genereerib palju puid ja leiab ligilähedase kauguse
- Ühendamise operatsioonile saab ahne kaalu anda - algoritm eelistab väiksemaid puid
- Reaalse kasutamise jaoks vajab paremaid heuristikaid
- Kitsendatud variant ka NP-täielik?