

Praktikos darbo programų reikalavimai

Saulius Gražulis

Bioinformatika III

Daugeliui bioinformatikos užduočių atlikti reikės didesnių duomenų kiekių (seku, struktūrų). Diskuose vietos gali būti per maža visoms struktūroms nukelti. Kad apeiti šį apribojimą, galima panaudoti keletą gudrybių:

1. Nukėlinėkite struktūras iš Interneto betarpiškai prieš skaičiavimus; iškart po skaičiavimų struktūras ištrinkite. Nukėlimui galite panaudoti `wget` komandą arba `wgetentry` skriptą (struktūrų nukėlimui iš PDB). Šią komandą galima panaudoti skaičiavimus valdančiame Makefile'e.

Kad galima būtų nustatyti, su kokia failo versija buvo suskaičiuoti rezultatai, galima prie rezultatų (komentaruose) užrašyti duomenų bazės įrašo kodą (pvz. PDB ID 1KNV) ir nukelto failo MD5 ir/arba SHA* kontrolines sumas.

2. Kooperuokitės. Jei 5 žmonių grupei reikia imties iš 100 PDB failų, kiekvienas gali nukelti po 20 struktūrų į visiems skaitymui prieinamą direktoriją, o bendrą PDB failų imtį kiekvienas suformuos, nustatę simbolines nuorodas (angl. symbolic links, arba symlinks) į savo ir draugų išsaugotus PDB failus. Žinoma, reikia susitarti ir PDB failų neištrinti, kol visi nebaigs darbų.