

Esame 5/07/2018

Andrea Passerini
andrea.passerini@unitn.it

Informatica

Programma python

Scrivere un programma che prenda in ingresso:

- un file con il mapping tra codici CATH e nomi delle strutture 3D corrispondenti
- un file con i dati dei codici CATH di ciascun dominio
- un livello tra 1 e 4
- un valore di soglia

e stampi l'elenco delle strutture (sia codice sia nome) che compaiono in un numero di domini superiore alla soglia, ordinate per frequenza. Il livello rappresenta la sottoparte di codice da considerare nel calcolare le statistiche (il codice CATH è gerarchico).

Esempio input

```
> head cath-b-newest-names
```

```
1 Mainly Alpha
2 Mainly Beta
3 Alpha Beta
4 Few Secondary Structures
1.10 Orthogonal Bundle
1.20 Up-down Bundle
1.25 Alpha Horseshoe
1.40 Alpha solenoid
1.50 Alpha/alpha barrel
2.10 Ribbon
```

```
> head cath-b-selection
```

```
101mA00 v4_2_0 1.10.490.10 0-153:A
1021A00 v4_2_0 1.10.530.40 1-162:A
102mA00 v4_2_0 1.10.490.10 0-153:A
1031A00 v4_2_0 1.10.530.40 1-162:A
103mA00 v4_2_0 1.10.490.10 0-153:A
1041A00 v4_2_0 1.10.530.40 1-162:A
1041B00 v4_2_0 1.10.530.40 1-162:B
104mA00 v4_2_0 1.10.490.10 1-153:A
105mA00 v4_2_0 1.10.490.10 1-153:A
106mA00 v4_2_0 1.10.490.10 0-153:A
```

Esempio esecuzione

```
> python cath_stats.py
CATH names file: cath-b-newest-names
CATH domains file: cath-b-selection
CATH level (1-4): 2
Frequency threshold: 500
3.40 (3-Layer(aba) Sandwich)      2003
1.10 (Orthogonal Bundle)          1796
2.60 (Sandwich) 1337
3.30 (2-Layer Sandwich) 1289
2.40 (Beta Barrel)      956
3.10 (Roll)             504
```

Programma python: suggerimento

Si possono implementare 5 funzioni separate:

- 1 una che legga il file di mapping e restituisca un dizionario da codice a nome
- 2 una che legga il file di dati e restituisca un elenco di dominio + codice
- 3 una che a partire da un elenco dominio+codice ed un livello, restituisca un dizionario che mappi dal codice (ristretto al livello in ingresso) ai domini con tale codice
- 4 una che data la mappa da codice a domini, la mappa da codici a nomi ed una soglia, stampi i codici (ed i rispettivi nomi) con frequenza superiore alla soglia, ordinati per frequenza
- 5 una (o un main) che realizzi il programma richiesto usando le funzioni di cui sopra

Esercizi da linea di comando

Contare il numero di Glutammine (Q) presenti nelle sequenze contenute all'interno della directory `fastas/`

Risultato atteso

721

Esercizi da linea di comando

Elencare le sequenze all'interno della directory `fastas/` che cominciano con Alanina (A) o con Acido Glutammico (E).

Risultato atteso

```
AAILGDEYLWSGGVIPYTFAGVSG...  
AATTLPDGAAAESLVESSEVAVIG...  
APLAVNDGDMRLADGGATNQGRVE...  
...
```

Modalita' di esecuzione e consegna

- 1 Avviare la macchina in modalita' `ESAME`
- 2 Autenticarsi con nome utente `sci-esame` e password fornita dal docente
- 3 Il testo del compito ed i file necessari si trovano in una cartella `Testo` sul Desktop
- 4 Realizzare il programma python come file `programma.py` e scrivere gli esercizi da linea di comando in un file di testo `linea_di_comando.txt`
- 5 Creare sul Desktop una cartella con *nome_cognome* e metterci i due file realizzati.
- 6 Eseguire il logout ma NON spegnere la macchina