

Esame 10/06/2018

Andrea Passerini
andrea.passerini@unitn.it

Informatica

Programma python

Scrivere un programma che prenda in ingresso due file di annotazioni GO, e stampi, per ciascun termine GO che si trova in entrambi i file:

- 1 il termine GO
- 2 il numero di geni annotati con quel termine nel primo file, e l'elenco delle descrizioni di tali geni
- 3 il numero di geni annotati con quel termine nel secondo file, e l'elenco delle descrizioni di tali geni

I termini devono essere ordinati in maniera decrescente in base alla somma dei geni nel primo e nel secondo file annotati con un termine. Vanno ignorate le descrizioni vuote e le proteine ipotetiche.

Esempio File di annotazioni GO

```
> head gene_association.PAMGO_Ddadantii
```

Esempio esecuzione

```
> python common_go.py
Primo file annotazioni GO: gene_association.PAMGO_Ddadantii
Secondo file annotazioni GO: gene_association.PAMGO_Mgrisea
GO: GO:0044409
1:      pectin acetylerase
11:     cAMP-dependent protein kinase catalytic subunit (cpkA) gene, c
GO:0005615
9:      Secreted protease C precursor, type III secreted protein, xylan
3:      Osteopontin precursor (Bone sialoprotein-1) (Secreted phospho
GO:0030163
3:      Secreted protease C precursor, Secreted protease A precursor, S
7:      "RPT1, ATP-dependent 26S proteasome regulatory subunit", Proba
...
```

Programma python: suggerimento

Si possono implementare 4 funzioni separate:

- 1 una che legga il file di annotazioni e restituisca un dizionario termine GO → lista di descrizioni dei geni annotati con quel termine, ignorando descrizioni vuote e proteine ipotetiche
- 2 una che prenda in ingresso due mappe e restituisca una lista di termini GO in comune tra le due e le liste di descrizioni dei rispettivi geni
- 3 una che stampi termini e liste secondo il formato richiesto (vedasi esempio), dove i termini sono in ordine decrescente in base alla somma dei geni annotati con il termine nei due file
- 4 una (o un main) che realizzi il programma richiesto usando le funzioni di cui sopra

Esercizi da linea di comando

Elencare i termini GO del file

`gene_association.PAMGO_Mgrisea` che hanno associate proteine ipotetiche, ordinati per il numero di proteine ipotetiche associate a ciascuno

Risultato atteso

```
3 GO:0046872
3 GO:0008270
3 GO:0008150
3 GO:0005634
3 GO:0005575
1 GO:0052108
1 GO:0050661
1 GO:0044408
1 GO:0030163
. . . .
```

Esercizi da linea di comando

Stampare a schermo i nomi dei file e le sequenze che:

- Non hanno acidi aspartici (D) o glutammici (E) nelle prime 10 posizioni, e ne hanno uno (aspartico o glutammico) nelle successive cinque

Risultato atteso

```
fastas/1a4rA.f:MQTIKCVVVGDVAVGKTCLL...
fastas/1cnsA.f:SVSSIIVSRAQFDRMLLHRND...
fastas/1d8wA.f:GHHHHHHXTTQLEQAWELAK...
fastas/1dqcA.f:YLAFRCGRYSPCLDDGPNVN...
fastas/1dvoA.f:PPKWVKVKKQKLAEKAAREAE...
fastas/1dw9A.f:XIQSQINRNIRLDLADAILL...
fastas/1gjsA.f:MKAIFVLNAQHDEAVDANSL...
fastas/1h3qA.f:MSGSFYFVIVGHHDNPVFEM...
fastas/1tsk_.f:VVIGQRCYRSPDCYSACKKL...
fastas/2pth_.f:TIKLIVGLANPGA EYAATRH...
```

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