

# Esame 08/06/2015

Andrea Passerini  
passerini@disi.unitn.it

Informatica

## Programma python

Scrivere un programma python che:

- prenda in ingresso un nome di file `filename` che contiene proteine in formato FASTA, in cui l'intestazione delle sequenze contiene localizzazione e organismo della proteina.
- stampi per ciascuna localizzazione il numero di proteine divise per organismo, dall'organismo più frequente al più raro.

# Esempio ingresso

```
>7B2_HUMAN:Secretory
MVS RMVSTMLSGLLFWLASGWTPAFAYSPRTPDRVSEADIQRLLHGVME ...
>A1AG1_MUSCR:Secretory
MALH MILVMLSLLPLLEAQNP EHVNITIGEPITNETLSWLSDKWFFIGA ...
>A1BG_HUMAN:Secretory
MSMLVVFLLLWGV TWGPVTEAAIFYETQPSLWAESESLKPLANVT LTC ...
...
>ZNF22_RAT:Nucleus
MRLGKPQKGGISR SATQ GKTYESRRKTARQRQKWGV AIRFDSGLSRRRR ...
>ZNF42_HUMAN:Nucleus
MRPAVLGSPDRAPPEDEGPVMVKLEDSEEEGEAALWDPGPEAARLRFR C ...
```

# Esempio esecuzione

```
> python program.py
Name of file: sequences.fasta
Mitochondrion
107:HUMAN 27:BOVIN 18:DROME 16:MOUSE 7:RAT 3:ASCSU ...
Secretory
191:HUMAN 79:MOUSE 59:DROME 35:BOVIN 30:RAT 19:CHICK ...
Nucleus
636:HUMAN 188:DROME 145:MOUSE 65:CAEEL 27:RAT ...
Cytoplasm
203:HUMAN 72:MOUSE 47:DROME 30:RAT 20:CAEEL 13:BOVIN ...
```

## Programma python: suggerimento

Si possono implementare 4 funzioni separate:

- 1 una che legga il file e restituisca un dizionario con localizzazione come chiave e come valore la lista di nomi di proteine con quella localizzazione.
- 2 una che prenda in ingresso il dizionario letto e restituisca un nuovo dizionario, con localizzazione come chiave e come valore un dizionario di conteggi organismo → numero di occorrenze.
- 3 una che prenda in ingresso il dizionario calcolato con la funzione precedente, e per ogni localizzazione stampi la lista di conteggi, ordinata per occorrenza decrescente
- 4 una che realizzi il programma richiesto usando le funzioni di cui sopra

## Shell: esercizio #1

Dati il file `sequences.fasta` ed i due seguenti motivi:

- Un amino acido qualunque; una prolina (P); una treonina (T) oppure una serina (S); una alanina (A); una prolina; un amino acido qualunque.
- Un triptofano (W); tre aminoacidi qualunque; una fenilalanina (F) oppure una tirosina (Y).

calcolare (1) quante sequenze includono il primo motivo *seguito a distanza arbitraria* dal secondo, e (2) quante includono *almeno uno* dei due motivi.

## Risultato atteso

1 20

2 624

## Shell: esercizio #2

Dato il file `sequences.fasta`, stampare a schermo i cinque organismi che appaiono più volte, in ordine decrescente di numero di occorrenze.

## Risultato atteso

```
1137 HUMAN
312  MOUSE
312  DROME
99   CAEEL
94   RAT
```

## 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 `utility.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