

# Esame 09/05/2016

Andrea Passerini  
passerini@disi.unitn.it

Informatica

## Programma python

Scrivere un programma python che:

- prenda in ingresso:
  - due nomi di file, con informazioni sull'arricchimento di termini di Gene Ontology per un certo numero di rna-binding proteins (RBP), in due condizioni diverse.
  - una soglia sul numero minimo di termini arricchiti in comune da riportare
- stampi per tutte e sole le RBP che hanno un numero di termini arricchiti in comune almeno pari alla soglia, il nome dell'RBP, l'id e la descrizione dei termini in comune.

```
>cat Testo/FEFE.enrichment.txt
```

# Esempio esecuzione

```
> python shared_enrichment.py
Inserire primo file di enrichment: FEFE.enrichment.txt
Inserire secondo file di enrichment: IDID.enrichment.txt
Inserire soglia: 5
BP:      QKI
GOTERMS:      GO:1902679 (negative regulation of RNA biosynt
RBP:      FUS
GOTERMS:      GO:0005829 (cytosol),GO:0035254 (glutamate rec
RBP:      LARP4B
GOTERMS:      GO:0016607 (nuclear speck),GO:0006357 (regulat
RBP:      PARK7
GOTERMS:      GO:0005654 (nucleoplasm),GO:2001244 (positive
...
```

## Programma python: suggerimento

Si possono implementare 5 funzioni separate:

- 1 Una che legga un file, e restituisca un dizionario che associ ad ogni RBP una mappa da `GO.ID` a `Term`
- 2 Una che dati due dizionari letti dai dati, restituisca un dizionario di intersezione con per ogni RBP, tutti e soli i termini in comune tra i due dizionari di ingresso.
- 3 Una che dato un dizionario ed una soglia, restituisca un dizionario filtrato con solo le RBP con un numero di termini almeno pari alla soglia
- 4 Una che stampi il dizionario
- 5 una che realizzi il programma richiesto usando le funzioni di cui sopra

## Shell: esercizio #1

Tra le colonne in `FEFE.enrichment.txt`, ci sono anche: ID del termine GO (`GO.ID`), nome esteso del termine GO (`Term`) e *p*-value (`pval`). Stampare a schermo i cinque termini GO nel namespace “biological process” con *p*-value piu’ alto.

**Nota:** la colonna `Term` include il testo “process” in questo caso. (Le altre due colonne non contengono mai “process”).

## Risultato atteso

GO:0006464	7.33
GO:0009893	7.58
GO:0006464	8.3
GO:0051252	9.72
GO:0032774	10.3

## Shell: esercizio #2

Dato il file `sequences.fasta`, contare quante sequenze:

- Contengono tre tirosine (Y), seguite da al più tre aminoacidi qualunque, seguite da una istidina (H).
- Contengono caratteri che non identificano nessuno dei 20 aminoacidi “standard”.

Ci sono sequenze che soddisfano entrambe le condizioni?

**Nota:** i 20 caratteri standard sono A R N D C E Q G H I L K M F P S T W Y V.

## Risultato atteso

3, 8, No.

## 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