

Esame 13/09/2011

Andrea Passerini
passerini@disi.unitn.it

Informatica

Programma python

Scrivere una funzione

`frequentPatterns(filename, kmin, kmax, freq)` che:

- prenda in ingresso un file fasta, una lunghezza minima `kmin` e massima `kmax` di pattern ed una loro frequenza minima `freq`
- legga le sequenze nel file
- stampi tutti i pattern (sottosequenze) di lunghezze tra `kmin` e `kmax` che compaiano almeno `freq` volte nelle sequenze (considerando anche ripetizioni in una sequenza)

Esempio

```
>>> import utility
>>> utility.frequentPatterns('seq.fasta',3,5,5)
...
...
...
DGKLK 5
IKQLQ 5
LKKLE 5
QKVAR 5
EIAAA 5
VYGVG 6
```

Programma python: suggerimento

Si possono implementare 5 funzioni separate:

- 1 una che legga da un file FASTA in ingresso le sequenze proteiche e le restituisca (e.g. dizionario)
- 2 una che prenda in ingresso le sequenze lette ed una lunghezza k , e restituisca un dizionario con tutte le sottosequenze di lunghezza k presenti, con la loro frequenza
- 3 una che prenda in ingresso un dizionario di sottosequenze ed una frequenza $freq$, e restituisca un dizionario con solo le sottosequenze con frequenza almeno $freq$ (con le loro rispettive frequenze)
- 4 una che prenda in ingresso un dizionario di sottosequenze e stampi tutte le coppie sottosequenza-frequenza
- 5 una che realizzi il programma richiesto usando le funzioni di cui sopra

Esercizi da linea di comando

- Selezionare tra le sequenze in `seq.fasta` quelle che contengono una *histag*, ossia (approssimativamente):
 - almeno quattro Istidine (H)
 - precedute da al massimo quattro residui all'inizio della sequenza
 - oppure seguite da al massimo quattro residui alla fine della sequenza
- e.g. **M****R****G****S****H****H****H****H**HHGSNIGETLGEKWKSR
- e.g. ...AIRAKAAALSLEHH**H****H****H****H**

Nota

- Nel file `seq.fasta` una singola sequenza può stare su più righe se è lunga
- Tutti i nomi di sequenza sono fatti di cinque caratteri

Esercizi da linea di comando

- Elencare le informazioni (permessi, proprietario, etc) dei due file più recenti contenuti nella vostra home directory
- Esempio di output (mia home):

```
total 140
drwxr-xr-x 17 andrea andrea 4096 2011-09-12 12:53 Personal
drwxr-xr-x  2 andrea andrea 8192 2011-09-12 12:19 Spool
```