

Esame 07/02/2014

Andrea Passerini
passerini@disi.unitn.it

Informatica

Programma python

Scrivere una funzione `findCommonStarts(fastafile, seed)` che:

- prenda in ingresso:
 - un nome di file `fastafile` con un elenco di sequenze in formato fasta
 - una sequenza `seed`
- e stampi i nomi delle sequenze che condividono con `seed` una sottosequenza iniziale (anche solo il primo residuo), e le sottosequenze in comune

Esempio esecuzione

```
>>> import utility
>>> utility.findCommonStarts('proteins.fasta','ADSFSGSH')
1ok0:a AD
2ask:a A
1pyb:d A
1jbe:a AD
1w6n:a A
2ite:a A
1lo7:a A
2jba:b A
3bv8:a A
1f66:c A
1xg0:a A
1jfi:a A
1ay7:b A
1cks:a A
3cq1:a A
2g7o:a A
```

Programma python: suggerimento

Si possono implementare 4 funzioni separate:

- 1 una che legga il file `fastafile` e restituisca il dizionario di sequenze contenute nel file
- 2 una che prenda il dizionario di sequenze ed il seed, e per ogni sequenza trovi la piu' lunga sottosequenza iniziale che condivide con seed, e metta nome e sottosequenza iniziale in un dizionario (se tale parte iniziale e' non vuota). La funzione deve restituire il dizionario costruito.
- 3 una che stampi il dizionario di nomi e sottosequenze iniziali
- 4 una che realizzi il programma richiesto usando le funzioni di cui sopra

Shell: Esercizio 1

Il file 1A26 contiene informazioni su sequenza e struttura tridimensionale di una proteina. Le righe che cominciano per SEQRES contengono la sequenza di residui, con codice a tre lettere.

Stampare su una singola riga la sequenza di residui di cui e' composta la proteina

Risultato

```
ALA LEU THR VAL SER ALA GLY THR LYS SER LYS LEU ALA LYS PRO ILE GLN ASP LEU ILE LYS MET ILE P  
GLU PHE GLU ILE ASP LEU GLN LYS MET PRO LEU GLY LYS LEU SER LYS ARG GLN ILE GLN SER ALA TYR S  
ASP GLY GLY SER GLU SER GLN ILE LEU ASP LEU SER ASN ARG PHE TYR THR LEU ILE PRO HIS ASP PHE G  
GLU TYR ILE GLN ALA LYS VAL GLN MET LEU ASP ASN LEU LEU ASP ILE GLU VAL ALA TYR SER LEU LEU A  
ILE ASP ILE ASN TYR GLU LYS LEU ARG THR ASP ILE LYS VAL VAL ASP LYS ASP SER GLU GLU ALA LYS I  
ALA THR HIS ASN ALA TYR ASP LEU LYS VAL VAL GLU ILE PHE ARG ILE GLU ARG GLU GLY GLU SER GLN A  
GLN LEU LEU TRP HIS GLY SER ARG THR THR ASN PHE ALA GLY ILE LEU SER GLN GLY LEU ARG ILE ALA P  
GLY LYS GLY ILE TYR PHE ALA ASP MET VAL SER LYS SER ALA ASN TYR CYS HIS THR SER GLN ALA ASP P  
LEU GLY ASN MET TYR GLU LEU LYS ASN ALA SER HIS ILE THR LYS LEU PRO LYS GLY LYS HIS SER VAL L  
ALA THR THR THR LEU ASP GLY VAL GLU VAL PRO LEU GLY ASN GLY ILE SER THR GLY ILE ASN ASP THR C  
VAL ALA GLN VAL ASN LEU LYS TYR LEU LEU LYS LEU LYS PHE ASN TYR LYS THR SER LEU TRP
```

Shell: Esercizio 2

Contare quante delle sequenze nel file `proteins.fasta` contengono:

- Il motivo di due cisteine separate da due residui qualunque
- Ripetuto due volte, con nel mezzo una sequenza qualunque che contenga al suo interno una coppia di acidi (D o E) contigui

Esempio:

AFASCFECDFSHGSDDASFACSSCDFAF

Risultato

2

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` nella home (NON nel 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 nella home una cartella con *nome_cognome* e metterci i due file realizzati.
- 6 Eseguire il logout ma NON spegnere la macchina