

Lezione7

December 11, 2019

1 Lezione 7

1.1 Esercizio 1

Dato il file words.txt, scrivere delle procedure che forniscano il seguente output: - per ogni parola, restituire il carattere dell'alfabeto con maggiore frequenza (qualora ci fossero frequenze uguali di valore massimo, prendere l'ultimo carattere di valore massimo) - restituire la parola di lunghezza massima (parola più lunga che inizia con a, parola più lunga che inizia con b, etc)

Attenzione: la soluzione proposta deve applicare in modo corretto il concetto di modularizzazione

1.2 Esercizio 2

Si consideri il file source.pdb. Scrivere un programma python che:

- usa un dizionario per leggere il nome degli aminoacidi (nella forma di codice a tre lettere) dalle linee SEQRES del file pdb;
- li converte nel codice ad una lettera;
- li concatena in una stringa al fine di ottenere una sequenza proteica;
- Infine, la sequenza e' stampata in un file output.fasta nel formato FASTA.

Ogni linea SEQRES puo' essere divisa in 17 colonne. La prima contiene la parola chiave "SEQRES", la seconda contiene il numero della sequenza (che parte da 1), la terza contiene l'ID della catena, e la quarta contiene il numero di aminoacidi che compongono la sequenza. Gli aminoacidi (rappresentati per mezzo del codice a tre lettere) sono riportati a partire dalla quinta colonna all'ultima. Le colonne sono separate mediante uno spazio bianco.

Un file FASTA ha la seguente struttura:

```
[1]: >NOMEFILE
MTEITAAMVKELRESTGAGMMDCKNALSETNGDFDKAVQLLREKGLGKAAKKADRLAAEG
LVSVKVSDDFTTIAAMRPSYLSYEDLDMTFVENEYKALVAELEKENEERRRLKDPNKPEHK
IPQFASRKQLSDAILKEAEEKIKEELKAQGKPEKIWDNIIPGKMNSFIADNSQLDSKLT
MGQFYVMDDKKTVEQVIAEKEKEFEFGGKIKIVEFICFEVGEGLKKTEDFAAEVAAQL
```

Ogni riga deve contenere al massimo 60 caratteri.

```
[3]: aa_codes = {
'ALA': 'A', 'CYS': 'C', 'ASP': 'D', 'GLU': 'E',
'PHE': 'F', 'GLY': 'G', 'HIS': 'H', 'LYS': 'K',
'ILE': 'I', 'LEU': 'L', 'MET': 'M', 'ASN': 'N',
```

```
'PRO': 'P', 'GLN': 'Q', 'ARG': 'R', 'SER': 'S',  
'THR': 'T', 'VAL': 'V', 'TYR': 'Y', 'TRP': 'W'} # Dizionario elenco aminoacidi
```

[: