

LE BIOSEQUENZE

Esercizio

L'elemento fondamentale in Bioinformatica è la Sequenza Nucleotidica. Le sequenze Nucleotidiche (DNA ed RNA) vengono rappresentate mediante stringhe di 4 elementi (A, T (U per RNA), C, G), nelle quali ciascuna lettera rappresenta un singolo nucleotide.

Il formato più diffuso per rappresentare per le sequenze Nucleotidiche è il Pearson FASTA.

La sequenza viene scritta dall'inizio alla fine senza spazi, ma può contenere interlinee.

Il formato FASTA impone una riga di intestazione che comincia con il simbolo ">" e contiene generalmente alcune informazioni basilari (non obbligatorie), relative alla sequenza.

Il file `input.txt` contiene le sequenze in formato FASTA.

Per semplicità assumiamo che ciascuna linea del file ha una lunghezza inferiore o uguale a 101 (incluso l'interlinea), ma una sequenza può essere scritta su più linee.

Dare in output:

- Numero di sequenze
- Lunghezza della sequenza più lunga
- Numero totale di nucleotidi

Nota: utilizzare la funzione `fgets` e `strlen`.

Esempio

Esempio di file "input.txt"

```
>Titolo sequenza 1
GCACTGTACC
>Titolo sequenza 2
GAACAGAAAGCTC
>Titolo sequenza 3
ACACAGG
>Titolo sequenza 4
ACCGGTA
TATAGGATAAGAA
>Titolo sequenza 5
TGGTACAAACGGG
>Titolo sequenza 6
TAAAGGGA
AG
>Titolo sequenza 7
TACACGAA
```

Output

```
Numero sequenze: 7
Lunghezza della sequenza più lunga: 20
Numero totale di nucleotidi: 81
```