

Εισαγωγή στον προγραμματισμό - 1η Άσκηση

Στο αρχείο `ecoli1.dat` θα βρείτε τις $\sim 4.5 \times 10^6$ βάσεις του γονιδιώματος του βακτηρίου *e-coli*. Οι βάσεις στο αρχείο συμβολίζονται με τα γράμματα a, c, g, t.

Να γραφεί πρόγραμμα σε C με όνομα `e-coli.c` που θα διαβάζει το αρχείο `ecoli1.dat` και θα εμφανίζει:

- το ποσοστό εμφάνισης της κάθε βάσης στο γονιδίωμα (αριθμός εμφανίσεων / σύνολο βάσεων)
- πόσες φορές εμφανίζεται στο γονιδίωμα η τριπλέτα “tga” ανάμεσα από τις θέσεις 800.000 και 810.000 καθώς και ανάμεσα από τις θέσεις 2.510.000 και 2.520.000.
- τον αριθμό των βάσεων που υπάρχουν ανάμεσα στις δύο εμφανίσεις του συμπλέγματος 'ttggat' στο γονιδίωμα ανάμεσα από τις θέσεις 800.000 και 810.000 .

και θα γράφει σε αρχείο με όνομα `results.txt` τις θέσεις εμφάνισης των δύο τελευταίων ζητούμενων με τον παρακάτω τρόπο:

tga -> 800000 - 810000

800070

800089

.....

tga -> 2510000 – 2520000

2510088

2510091

.....

ttggat -> 800000 – 810000

.....

Παραδοτέα

Ένα αρχείο zip που θα περιέχει:

- Ένα αρχείο pdf με όνομα e-coli.pdf στο οποίο:
 - Θα υπάρχει εξώφυλλο με το όνομα και τον αριθμό μητρώου του φοιτητή.
 - Μια περιγραφή της κάθε συνάρτησης του προγράμματος και της λειτουργίας της
 - Τα αποτελέσματα που εμφανίζονται στο τερματικό.
 - Τις παρατηρήσεις και τα συμπεράσματα σας.
- Το αρχείο e-coli.c με τον κώδικα
- Το αρχείο results.txt

Προθεσμία Υποβολής

Κυριακή 9/1/2022

Καλή Επιτυχία !