



ΠΟΛΥΤΕΧΝΕΙΟ ΚΡΗΤΗΣ ΤΜΗΜΑ ΗΜΜΥ

ΕΡΓΑΣΤΗΡΙΟ ΔΙΑΝΕΜΗΜΕΝΩΝ
ΠΛΗΡΟΦΟΡΙΑΚΩΝ ΣΥΣΤΗΜΑΤΩΝ ΚΑΙ ΕΦΑΡΜΟΓΩΝ
<http://courses.ece.tuc.gr>

ΗΜΜΥ ΠΛΗ 101

ΕΙΣΑΓΩΓΗ ΣΤΟΥΣ ΗΛΕΚΤΡΟΝΙΚΟΥΣ ΥΠΟΛΟΓΙΣΤΕΣ ΚΑΙ ΤΗΝ ΠΛΗΡΟΦΟΡΙΚΗ
ΕΡΓΑΣΤΗΡΙΑΚΗ ΑΣΚΗΣΗ

ΗΜΕΡΟΜΗΝΙΑ ΠΑΡΑΔΟΣΗΣ: 10.01.2012

ΗΜΕΡΟΜΗΝΙΑ ΕΞΕΤΑΣΗΣ: 19-20.01.2012

ΥΠΕΥΘΥΝΟΙ ΕΡΓΑΣΤΗΡΙΟΥ: Ανέστης Γιώργος, Καζάσης Φώτης, Μουμουτζής Νεκτάριος

ΕΠΙΚΟΙΝΩΝΙΑ: {ganest, fotis, nektar}@ced.tuc.gr

Πρώτος Κύκλος Ασκήσεων

Οι παρακάτω ασκήσεις είναι μέρος των εργαστηρίων του μαθήματος ΗΜΜΥ 101, και είναι υποχρεωτικές. Ο πρώτος κύκλος αντιστοιχεί στο 40% του συνολικού βαθμού του εργαστηρίου. Οι ασκήσεις γίνονται **ατομικά**.

Ο πρώτος κύκλος ασκήσεων πρέπει να σταλεί, σύμφωνα με τις οδηγίες που δίνονται παρακάτω, μέχρι τις **15:00 η ώρα της 10ης Ιανουαρίου 2012**.

Η υποβολή του προγράμματος αυτόματα δηλώνει ότι είσθε οι μοναδικοί συγγραφείς της λύσης της άσκησης. **Εάν το πρόγραμμα ή μέρος αυτού έχει αντιγραφεί, θα πρέπει να αναφέρεται η πηγή του και ο συγγραφέας του.** Τα μέρη της λύσης που έχουν αντιγραφεί δε θα βαθμολογούνται αναλογικά.

ΠΡΩΤΗ ΑΣΚΗΣΗ (50%)

Περιγραφή

Τα **νουκλεοτίδια** είναι οργανικά μόρια, που αποτελούν τις δομικές μονάδες των **νουκλεϊκών οξέων** (DNA) όταν τρία (3) ή περισσότερα ενώνονται μεταξύ τους για να σχηματίσουν ένα μακρομόριο νουκλεϊκού οξέος. Ένα νουκλεοτίδιο αποτελείται (μεταξύ άλλων) από μια αζωτούχα βάση. Οι αζωτούχες βάσεις στο DNA είναι η αδενίνη (adenine), η γουανίνη (guanine), η θυμίνη (thymine) και η κυτοσίνη (cytosine) και συνακόλουθα μπορούμε να συμβολίζουμε τα αντίστοιχα νουκλεοτίδια που το συνθέτουν με το αρχικό γράμμα της αζωτούχας βάσης καθενός. Έτσι η αλληλουχία νουκλεοτιδίων σε ένα μακρομόριο DNA μπορεί να αναπαρασταθεί ως μια ακολουθία των χαρακτήρων 'a', 'c', 'g' και 't', πχ.

c a a t g c t a g c a a t t g c t a g c

Κάθε τέτοια αλληλουχία κωδικοποιεί τη δομή κάποιας πρωτεΐνης ως εξής: Ανά τρία τα νουκλεοτίδια του μακρομορίου DNA προσδιορίζουν ένα μόριο αμινοξέως. Η αλληλουχία των αμινοξέων που προσδιορίζονται από τις διαδοχικές τριάδες των νουκλεοτιδίων του DNA αντιστοιχεί στο μακρομόριο μιας πρωτεΐνης. Προφανώς, ο προσδιορισμός της πρωτεΐνης εξαρτάται από το σημείο από το οποίο θα ξεκινήσει αυτή η διαδικασία αντιστοίχισης τριάδων νουκλεοτιδίων με μόρια αμινοξέων. Η ακριβή εύρεσης του σημείου αυτού (εφόσον υπάρχει) για την ορθή κωδικοποίηση μιας δοσμένης πρωτεΐνης αποτελεί το αντικείμενο της παρούσας άσκησης.

Για τους σκοπούς αυτής της άσκησης, ορίζουμε μια **γενετική ακολουθία** ως μια αλληλουχία χαρακτήρων καθένας από τους οποίους μπορεί να πάρει τις τιμές 'a', 'c', 'g' ή 't'. Η κωδικοποίηση ενός αμινοξέος είναι μια γενετική ακολουθία μήκους 3. Υπάρχουν 25 αμινοξέα, στα οποία θα αναφερόμαστε με τα κεφαλαία γράμματα (του αγγλικού αλφάβητου) από 'A' έως 'Y'. Μια **πρωτεΐνη** αντιστοιχεί σε μια αλληλουχία αμινοξέων, δηλαδή σε μια αλληλουχία χαρακτήρων καθένας από τους οποίους μπορεί να πάρει τις τιμές από 'A' έως και 'Y'.

Δεδομένου ότι ένα αμινοξύ, όπως ορίστηκε, αντιστοιχεί σε μια γενετική ακολουθία τριών χαρακτήρων (κωδικών νουκλεοτιδίων) και εφόσον οι δυνατοί χαρακτήρες σε κάθε μία από τις τρεις (3) θέσεις της αλληλουχίας είναι

τέσσερις (4), προκύπτει ότι συνολικά υπάρχουν $4^3 = 64$ διαφορετικές κωδικοποιήσεις των αμινοξέων, σε καθεμία από τις οποίες αντιστοιχεί ακριβώς ένα αμινοξύ (η κωδικοποίηση είναι “καλά ορισμένη”), αλλά ένα αμινοξύ μπορεί να έχει περισσότερες από μία κωδικοποιήσεις (η συνάρτηση αντιστοίχισης δεν είναι “1-1”).

Έστω τώρα ότι δίνεται μια πρωτεΐνη και μια μεγάλη γενετική ακολουθία η οποία ενδέχεται να περιέχει την συγκεκριμένη πρωτεΐνη. Το ζητούμενο είναι να καθοριστεί η γενετική κωδικοποίηση των αμινοξέων της πρωτεΐνης. Ένα παράδειγμα θα καταδείξει τη διαδικασία που χρησιμοποιούμε για να κάνουμε αυτό:

Ας υποθέσουμε ότι γνωρίζουμε ότι η πρωτεΐνη M S I Q H M R βρίσκεται κάπου στη γενετική αλληλουχία attgctagcaatgctagcaattgctagcaatt αλλά δεν ξέρουμε πού, και δεν γνωρίζουμε καμία κωδικοποίηση για τα αμινοξέα. Κατ'αρχάς, δοκιμάστε να ευθυγραμμίσετε τους πρώτους χαρακτήρες:

```
a t t g c t a g c a a t g c t a g c a a t t g c t a g c a a t t
M      S      I      Q      H      M      R
```

Αυτή η προσέγγιση θα μπορούσε να λειτουργήσει εάν ένας κώδικας για το M ήταν att, για το S ήταν gct, κλπ. Αλλά αυτό δεν ισχύει διότι το gct δεν μπορεί να είναι μια κωδικοποίηση και για το S και για το H, όπως φαίνεται ανωτέρω. Το να έχουμε δύο διαφορετικές γενετικές κωδικοποιήσεις για ένα αμινοξύ (για παράδειγμα, το M κωδικοποιείται ως att ή agc στην περίπτωση αυτή) επιτρέπεται, αλλά κάθε γενετικός κώδικας πρέπει να αντιπροσωπεύει μόνο ένα αμινοξύ.

Οι ακολουθίες δεν ευθυγραμμίζονται με την πρώτη προσπάθεια, γι' αυτό δοκιμάζουμε να ευθυγραμμίσουμε την πρωτεϊνική ακολουθία από το δεύτερο χαρακτήρα της γενετικής ακολουθίας:

```
a t t g c t a g c a a t g c t a g c a a t t g c t a g c a a t t
      M      S      I      Q      H      M      R
```

Και πάλι, το cta δεν μπορεί να κωδικοποιήσει δύο διαφορετικά αμινοξέα (S και H), έτσι ούτε αυτή η ευθυγράμμιση δεν μπορεί να γίνει αποδεκτή. Το ίδιο συμβαίνει και στην περίπτωση που ξεκινήσουμε από τον τρίτο χαρακτήρα της γενετικής ακολουθίας. Ας δούμε, τέλος, τι γίνεται αν ξεκινήσουμε από τον τέταρτο χαρακτήρα:

```
a t t g c t a g c a a t g c t a g c a a t t g c t a g c a a t t
          M      S      I      Q      H      M      R
```

Σε αυτή την περίπτωση υπάρχουν δύο προβληματικές τριάδες: Η agc αντιστοιχεί στα S και H και η gct αντιστοιχεί στα M και Q. Απορρίπτεται λοιπόν κι αυτή η ευθυγράμμιση. Συνεχίζοντας τη μετατόπιση της πρωτεϊνικής ακολουθίας κατά μήκος της γενετικής ακολουθίας με αυτόν τον τρόπο, θα έχουμε τελικά μια ευθυγράμμιση που θα μπορούσε να λειτουργήσει:

```
a t t g c t a g c a a t g c t a g c a a t t g c t a g c a a t t
                        M      S      I      Q      H      M      R
```

Η ευθυγράμμιση αυτή δείχνει ότι atg και agc είναι ο γενετικός κώδικας για το αμινοξύ M, cta είναι ένας κωδικός για το S, gca για το I, κοκ.

Αυτό είναι ένα τεχνητό παράδειγμα για να δείξουμε πως η μέθοδος λειτουργεί. Σε πραγματικά δεδομένα, οι ακολουθίες είναι πολύ μεγαλύτερες.

Ξεκινήστε από το σκελετό προγράμματος που δίνεται μαζί με την εκφώνηση της άσκησης (αρχείο *main.c*). Στο πρόγραμμα αυτό αρχικοποιείται ένας πίνακας με τους χαρακτήρες μιας συγκεκριμένης γενετικής ακολουθίας. Επεκτείνετε το πρόγραμμα ώστε να ζητάει από τον χρήστη να δώσει μια πρωτεϊνική ακολουθία από το πληκτρολόγιο. Θεωρήστε ότι η πρωτεϊνική ακολουθία έχει μήκος το πολύ 300 χαρακτήρες. Το πρόγραμμα θα διαβάσει την πρωτεΐνη και στη συνέχεια θα προσπαθεί να βρει μια έγκυρη ευθυγράμμιση με τη συγκεκριμένη γενετική ακολουθία. Αν υπάρχει τέτοια ευθυγράμμιση, θα τη βρίσκει και θα εκτυπώνει στην οθόνη την γενετική κωδικοποίηση των αμινοξέων καθώς και τη θέση στον πίνακα της γενετικής ακολουθίας από όπου ξεκινά η ευθυγράμμιση που βρέθηκε. Αν δεν υπάρχει έγκυρη ευθυγράμμιση, το πρόγραμμα θα τυπώνει μήνυμα ότι δεν μπόρεσε να βρει έγκυρη γενετική κωδικοποίηση των αμινοξέων.

ΠΡΟΣΟΧΗ: Το πρόγραμμά σας θα πρέπει να εκτελεί όλους τους απαραίτητους ελέγχους που θα εξασφαλίζουν τη σωστή λειτουργία του.

Τι θα πρέπει να παραδώσετε

Το αρχείο με τον πηγαίο κώδικά σας με όνομα της μορφής x_set1_exer1.c, , όπου x θα το αντικαταστήσετε με τον κωδικό της ομάδας σας (π.χ. LAB101030_set1_exer1.c). Προσοχή στη στοίχιση του κώδικα και τα σχόλια, τα οποία

θα πρέπει να είναι επεξηγηματικά του κώδικα. Στις πρώτες γραμμές του πηγαίου κώδικα να αναφέρετε το όνομά σας και τον αριθμό του φοιτητικού μητρώου σας.

ΔΕΥΤΕΡΗ ΑΣΚΗΣΗ (50%)

Περιγραφή

Η ακολουθία Fibonacci ορίζεται αναδρομικά για κάθε μη αρνητικό ακέραιο n ως εξής:

$$F(n) = F(n-1) + F(n-2), n > 1 \text{ και } F(0)=0, F(1)=1.$$

Με βάση αυτό τον ορισμό, ο υπολογισμός της τιμής της ακολουθίας για δοσμένο n μπορεί να υλοποιηθεί είτε με μια αναδρομική συνάρτηση, σύμφωνα με τον παραπάνω ορισμό, είτε με μη αναδρομική συνάρτηση που ξεκινά τον υπολογισμό από τις μικρότερες τιμές του n και προχωρά μέχρι την τιμή για το δοσμένο n κρατώντας σε κάθε βήμα τις δύο τελευταίες τιμές της ακολουθίας που έχουν υπολογιστεί μέχρι στιγμής.

ΑΝΑΠΤΥΞΗ ΑΡΧΙΚΟΥ ΠΡΟΓΡΑΜΜΑΤΟΣ

Αναπτύξτε ένα πρόγραμμα το οποίο θα υλοποιεί και τις δύο μεθόδους υπολογισμού της ακολουθίας Fibonacci μέσω δύο συναρτήσεων, μίας αναδρομικής και μίας μη αναδρομικής αντίστοιχα.

Το πρόγραμμα θα ζητά στην αρχή από τον χρήστη να εισάγει ένα θετικό ακέραιο και στη συνέχεια θα υπολογίζει την τιμή της ακολουθίας Fibonacci με είσοδο αυτό τον αριθμό καλώντας διαδοχικά τις δύο συναρτήσεις που υλοποιήσατε και τυπώνοντας το τελικό αποτέλεσμα που δίνει η καθεμία αναφέροντας ποια συνάρτηση έδωσε το κάθε αποτέλεσμα.

ΠΡΩΤΗ ΕΠΕΚΤΑΣΗ

Στη συνέχεια καλείστε να επεκτείνετε το πρόγραμμα που φτιάξατε ώστε κάθε φορά που ξεκινά η κλήση μίας συνάρτησης να τυπώνει ένα κατάλληλο μήνυμα του τύπου:

Computing fib(...)

όπου στη θέση των αποσιωπητικών θα εμφανίζεται η τιμή εισόδου.

Όταν τελειώνει η εκτέλεση της κλήσης θα τυπώνει μήνυμα του τύπου:

fib(...) finished with result ...

Στην περίπτωση αναδρομικής συνάρτησης θα πρέπει τα ανωτέρω μηνύματα να εμφανίζονται για κάθε αναδρομική κλήση δίνοντας τη δυνατότητα στο χρήστη να δει τον τρόπο με τον οποίο λειτουργεί η αναδρομή.

Αλλάξτε το όνομα της συνάρτησης (fib) στα ανωτέρω ενδεικτικά μηνύματα ώστε να ανταποκρίνεται στα ονόματα που έχετε χρησιμοποιήσει εσείς για τις δύο συναρτήσεις που υλοποιήσατε.

ΔΕΥΤΕΡΗ ΕΠΕΚΤΑΣΗ

Τέλος επεκτείνετε το πρόγραμμα που προέκυψε μετά την πρώτη επέκταση ώστε να υπολογίζει το πλήθος των προσθέσεων που εκτελούνται κατά τη διάρκεια κάθε κλήσης μίας συνάρτησης. Σε περίπτωση αναδρομικής συνάρτησης ο υπολογισμός του πλήθους για κάθε κλήση θα προσθέτει και το πλήθος των πράξεων που έγιναν από τις επιμέρους αναδρομικές κλήσεις. Τι παρατηρείτε; Ποια από τις δύο συναρτήσεις είναι πιο οικονομική από πλευράς πλήθους εκτελούμενων προσθέσεων;

ΠΑΡΑΔΟΤΕΟ

Το παραδοτέο σας θα είναι το πρόγραμμα που προέκυψε μετά την υλοποίηση και της δεύτερης επέκτασης. Το πρόγραμμα αυτό θα περιλαμβάνει, μεταξύ των άλλων, και λεπτομερή σχόλια για τις παρατηρήσεις που κάνατε συγκρίνοντας το πλήθος των εκτελούμενων πράξεων κατά την εκτέλεσή του (μετά την υλοποίηση και της δεύτερης επέκτασης).

ΠΡΟΣΟΧΗ: Το πρόγραμμά σας θα πρέπει να εκτελεί όλους τους απαραίτητους ελέγχους που θα εξασφαλίζουν τη σωστή λειτουργία του.

Τι θα πρέπει να παραδώσετε

Το αρχείο με τον πηγαίο κώδικά σας με όνομα της μορφής x_set1_exer2.c, , όπου x θα το αντικαταστήσετε με τον κωδικό της ομάδας σας (π.χ. LAB101030_set1_exer2.c). Προσοχή στη στοίχιση του κώδικα και τα σχόλια, τα οποία θα πρέπει να είναι επεξηγηματικά του κώδικα. Στις πρώτες γραμμές του πηγαίου κώδικα να αναφέρετε το όνομά σας και τον αριθμό του φοιτητικού μητρώου σας.

ΟΔΗΓΙΕΣ ΓΙΑ ΤΗΝ ΗΛΕΚΤΡΟΝΙΚΗ ΥΠΟΒΟΛΗ ΤΗΣ ΑΣΚΗΣΗΣ

Αποστολή του 1ου κύκλου ασκήσεων: Ο 1ος κύκλος ασκήσεων αποτελείται από δύο (2) ασκήσεις.

(Βήμα 1) Συμβουλευτείτε τα τμήματα «Τι θα πρέπει να παραδώσετε» των ασκήσεων

(Βήμα 2) Δημιουργήστε ένα συμπιεσμένο αρχείο (.zip) με όνομα της μορφής x_set1.zip, όπου x θα το αντικαταστήσετε με τον κωδικό της ομάδας σας (π.χ. LAB101030_set1.zip). Στο αρχείο αυτό συμπεριλάβετε τα αρχεία που δημιουργήσατε στο βήμα 1.

Για παράδειγμα, αν είστε η ομάδα LAB101030, θα δημιουργήσετε ένα zip αρχείο με όνομα LAB101030_set1.zip και περιεχόμενο τα παρακάτω αρχεία:

LAB101030_set1_exer1.c

LAB101030_set1_exer2.c

(Βήμα 3) Αφού *βεβαιωθείτε ότι έχετε εισάγει σωστά τον κωδικό της ομάδας σας*, υποβάλετε το συμπιεσμένο αρχείο που δημιουργήσατε μέσω της ιστοσελίδας του μαθήματος.