ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ

ΣΧΟΛΗ ΤΕΧΝΟΛΟΓΙΩΝ ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ ΕΠΙΚΟΙΝΩΝΙΩΝ

ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ

Μάθημα Προπτυχιακών Σπουδών:

Βιοπληροφορική

Ακαδημαϊκό Έτος: 2022 - 2023

Εξάμηνο: 6ο

Απαλλακτική Εργασία

<u>Ομάδα Εργασίας:</u>

Θεόδωρος Κοξάνογλου Π20094,

Απόστολος Σιαμπάνης Π20173,

Δημήτρης Στυλιανού Π20004,

Κωνσταντίνος Λοϊζίδης Π20007

Περιεχόμενα

Εκφώνηση

Εργασία Βιοπληροφορικής, εαρινό εξάμηνο ακαδημαϊκού έτους 2022-2023

<u>Ημερομηνία παράδοσης:</u> Ημερομηνία εξέτασης μαθήματος, ώρα 23:59

Η εργασία εκπονείται είναι απαλλακτική και εκπονείται σε ομάδες 1-4 φοιτητών. Μπορεί να παραδοθεί στην εξεταστική Σεπτεμβρίου.

Παραδοτέα μέσω της ενότητας εργασιών στο e-class:

- Η τεκμηρίωση της εργασίας σε ένα αρχείο pdf, στην πρώτη σελίδα της οποίας αναγράφονται τα ονοματεπώνυμα των φοιτητών και οι ΑΜ. Δεν θα βαθμολογηθούν εργασίες που δεν περιέχουν τεκμηρίωση ή που δεν αναφέρουν τα ονόματα των μελών της ομάδας στην τεκμηρίωση.
- Τα αρχεία source code σε ένα συμπιεσμένο αρχείο με όνομα **source2023.zip** (ή .rar ή άλλη σχετική κατάληξη).
- Οποιαδήποτε άλλα συνοδευτικά αρχεία η ομάδα κρίνει απαραίτητα σε ένα συμπιεσμένο αρχείο με το όνομα **auxiliary2023.zip** (ή .rar ή άλλη σχετική κατάληξη).

Θέμα 1 (3.5 βαθμοί): Άσκηση 11.4, βιβλίο "Εισαγωγή στους Αλγορίθμους Βιοπληροφορικής".

Θέμα 2 (3.5 βαθμοί): Υλοποιήστε μία από τις ασκήσεις 6.12, 6.13 του βιβλίου "Εισαγωγή στους Αλγορίθμους Βιοπληροφορικής".

Θέμα 3 (3 βαθμοί): Υλοποιήστε την άσκηση 6.14 του βιβλίου "Εισαγωγή στους Αλγορίθμους Βιοπληροφορικής".

Αποδεκτές γλώσσες υλοποίησης είναι οι Python, Matlab. Κάθε θέμα πρέπει να συνοδεύεται από τεκμηρίωση της λύσης. Τα θέματα 2 και 3 συσχετίζονται με επιπρόσθετα δεδομένα, τις περιγραφές των οποίων θα βρείτε στον σχετικό κατάλογο της ενότητας εγγράφων.

Πρώτο Θέμα

<u>Άσκηση 11.4:</u> Στο Σχήμα 11.7 φαίνεται ένα HMM με δύο καταστάσεις α και β. Όταν το HMM βρίσκεται στην κατάσταση α, έχει μεγαλύτερη πιθανότητα να εκπέμψει πουρίνες (A και G). Όταν βρίσκεται στην κατάσταση β, έχει μεγαλύτερη πιθανότητα να εκπέμψει πυριμιδίνες (C και T). Αποκωδικοποιήστε την πιο πιθανή ακολουθία των καταστάσεων (α/β) για την αλληλουχία GGCT. Χρησιμοποιήστε λογαριθμικές βαθμολογίες αντί για κανονικές βαθμολογίες πιθανοτήτων.

Επεξήγηση κώδικα

Οι αρχικές πιθανότητες για τις καταστάσεις **α** και **β** έχουν οριστεί ως ισοπίθανες, δηλαδή 0.5 η κάθε μια. Έτσι, προκύπτει ότι:

Pα(x) = Pβ(x) = 0.5 * Mk(x)

Όπου x ∈ {A,G,C,T} και Mk(x) η πιθανότητα εκπομπής του x από την κατάσταση k του πίνακα εκπομπών.

Στη συνέχεια, παρουσιάζονται οι πιθανότητες **εκπομπής** (emission) ανά κατάσταση:

State\Nucleotides	A	G	T	C
α	0.4	0.4	0.1	0.1
β	0.2	0.2	0.3	0.3

Ύστερα, παρουσιάζονται και οι πιθανότητες **μετάβασης** (transition) από την τρέχουσα κατάσταση στην επόμενη.

Current State\State to go	α	β
α	0.9	0.1
β	0.1	0.9

Ορίζουμε ότι το $S_{k,i}$ είναι η πιθανότητα της πιο πιθανής διαδρομής για το πρόθεμα $x_1 \dots x_i$ που τελειώνει στην κατάσταση k.Τότε,για οποιαδήποτε κατάσταση l έχουμε πως:

$$S_{l,i+1} = e_l(x_{i+1}) * \max_{k \in Q}\{S_{k,i} * \alpha_{kl}\}$$

Όμως για την υλοποίηση και την αποφυγή υπερχείλισης χρησιμοποιήθηκε η λογαριθμική αναπαράσταση του αλγορίθμου **Viterbi**, έτσι χρησιμοποιήθηκε η συνάρτηση:

$$\mathbf{S_{l,i+1}=log_2 e_l(x_{i+1}) + max_{k\in Q}\{S_{k,i} + log_2(\alpha_{kl})\}}$$

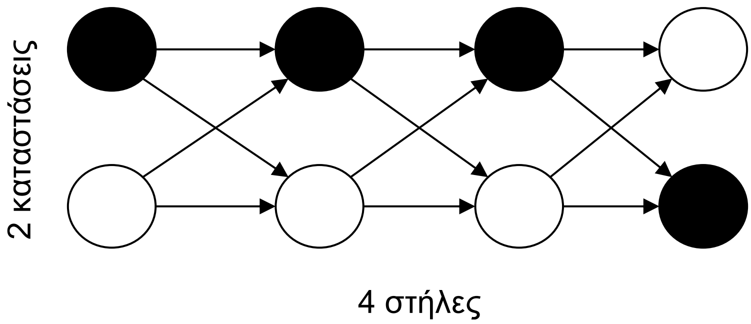
Για την εύρεση της κατάστασης για κάθε ένα από τα επόμενα νουκλεοτίδια της ακολουθίας γίνεται η χρήση του αλγορίθμου **Viterbi**.

Υπολογίζοντας για όλα τα νουκλεοτίδια της ακολουθίας την πιθανότητα εκπομπής ενός συμβόλου από την ακολουθία, εμφανίζουμε το λογαριθμικό αποτέλεσμα τους στον παρακάτω πίνακα.

State\Nucleotides	**A**	**G**	**T**	**C**
α	-2.321928094887362	-3.7958592832197744	-7.269790471552186	-10.743721659884597
β	-3.321928094887362	-5.795859283219775	-7.68482797083103	-9.573796658442285

Διαλέγοντας την μεγαλύτερη πιθανότητα παραγωγής του συγκεκριμένου νουκλεοτιδίου προκύπτει η ακολουθία καταστάσεων “αααβ” για την αλληλουχία GGCT.

Για να οπτικοποιήσουμε την παραπάνω λειτουργία χρησιμοποιούμε ένα πλέγμα Manhattan, σύμφωνα με τον αλγόριθμο Viterbi όπου αποτελείται από 2 καταστάσεις (α, β) και 4 στήλες (G, G, C, T). Διαλέγοντας την μεγαλύτερη πιθανότητα παραγωγής του συγκεκριμένου νουκλεοτιδίου για την κάθε κατάσταση προκύπτει το παρακάτω μονοπάτι:



Για την υλοποίηση της διαδικασίας που παρουσιάζεται στο παραπάνω σχήμα χρησιμοποιείται η μέθοδος **viterbi_algorithm()**. Με την κλήση αυτής από την main, αρχικά βρίσκεται το σκορ της αρχικής παρατήρησης και στην συνέχεια με έναν επαναληπτικό βρόγχο γίνεται η διάτρεξη της ακολουθίας μέχρι την εύρεση της καλύτερης ακολουθίας καταστάσεων, όπως αυτή που αναπαριστάται στο παραπάνω σχήμα. Ανάλογα με το σκορ που υπολογίζεται σε κάθε κατάσταση, διατηρώντας την μεγαλύτερη από τις τιμές τους (max) σχηματίζεται το μονοπάτι μετάβασης (transition path) για την ακολουθία.

<u>Παράδειγμα εκτέλεσης</u>

```
C:\Users\apost\Desktop\bioinformatics_project_2023\venv\Sc
score_a: -2.321928094887362, score_b: -3.321928094887362
score_a: -3.7958592832197744, score_b: -5.795859283219775
score_a: -7.269790471552186, score_b: -7.68482797083103
score_a: -10.743721659884597, score_b: -9.573796658442285
Optimal path through the hidden states: aaab

Process finished with exit code 0
```

Το αποτέλεσμα της ακολουθίας είναι: **αααβ**

Δεύτερο Θέμα

<u>Άσκηση 6.13:</u> Δύο παίκτες παίζουν το ακόλουθο παιχνίδι με δύο αλληλουχίες που έχουν μήκος n και m νουκλεοτίδια αντίστοιχα. Σε κάθε γύρο του παιχνιδιού, ένας παίκτης μπορεί να αφαιρέσει έναν τυχαίο αριθμό νουκλεοτιδίων από τη μία αλληλουχία ή τον ίδιο (αλλά και πάλι τυχαίο) αριθμό νουκλεοτιδίων και από τις δύο αλληλουχίες. Ο παίκτης που αφαιρεί το τελευταίο νουκλεοτίδιο κερδίζει. Ποιος θα κερδίσει; Περιγράψτε τη νικηφόρα στρατηγική για όλες τις τιμές των n και m.

Έστω ότι έχουμε **n** και **m** τα μήκη των δύο αλληλουχιών.

Ανάλογα με την κίνηση του παίκτη κάθε φορά διαγράφεται ένας τυχαίος αριθμός νουκλεοτιδίων από την μία αλληλουχία ή τον ίδιο (αλλά πάλι τυχαίο) αριθμό νουκλεοτιδίων και από τις δύο αλληλουχίες.

Παρατηρούμε ότι οι κινήσεις αυτές παρομοιάζουν κινήσεις βασίλισσας πάνω σε μία σκακιέρα. Έστω λοιπόν ένα παιχνίδι 2 παικτών σε σκακιέρα μεγέθους **n+1** στηλών και **m+1** γραμμών, στην οποία υπάρχει μία βασίλισσα στην κάτω δεξιά γωνία και στόχος του είναι να καταλήξει στην πάνω αριστερή γωνία. Ο παίκτης που θα αφαιρέσει το τελευταίο νουκλεοτίδιο χρίζεται νικητής.

Το **+1** στο μέγεθος της σκακιέρας εξυπηρετεί στη σωστή διαχείριση της περίπτωσης, όπου μία από τις δύο ή και οι δύο αλληλουχίες είναι κενές. Μία αλληλουχία θεωρείται κενή όταν η βασίλισσα βρίσκεται στη θέση μηδέν ενός άξονα μιας αλληλουχίας.

Έστω x ένας τυχαίος αριθμός.

Οι μόνες αποδεκτές κινήσεις της στο παιχνίδι είναι:

- Να κινηθεί οσεσδήποτε θέσεις αριστερά (**n - x**)
- Να κινηθεί οσεσδήποτε θέσεις πάνω (**m - x**)
- Να κινηθεί οσεσδήποτε θέσεις αριστερά (**n - x**) και πάνω (**m - x**)

Κάθε φορά που μετακινείται η βασίλισσα, **διαγράφονται τα αντίστοιχα νουκλεοτίδια** από το μονοπάτι που διένυσε.

Από υπόθεση γνωρίζουμε πως έχουμε 2 παίκτες. Για λόγους προσομοίωσης παραδείγματος ενός παιχνιδιού, ο ένας παίκτης θα κινείται χωρίς να έχει κάποια στρατηγική (παίκτης 2) και ο άλλος παίκτης θα κινείται με την νικηφόρα στρατηγική (παίκτης 1) που θα περιγράψουμε παρακάτω.

Η νικηφόρα στρατηγική χωρίζεται σε δύο κατηγορίες: <u>επιλογή σειράς</u> και <u>επιλογή επόμενης κίνησης</u>.

Νικηφόρα Στρατηγική

Έστω **number** το τυχαίο μήκος νουκλεοτιδίων που θα χρειαστεί να αφαιρέσουμε **ανά περίπτωση** από τις αλληλουχίες.

Επιλογή Σειράς

1) Αν ο πίνακας που σχηματίζεται αρχικά είναι τετραγωνικός τότε η αρχική θέση αποτελεί στοιχείο της διαγωνίου, τότε ο παίκτης 1 παίζει πρώτος.
2) Αντιθέτως, αν ο πίνακας που σχηματίζεται δεν είναι τετραγωνικός, τότε πρώτος παίζει ο παίκτης 2.

Με αυτό τον τρόπο ο παίκτης 1 (που ακολουθεί την νικηφόρα στρατηγική) θα βρίσκεται σε πλεονεκτική θέση στην αρχή του παιχνιδιού.

Ακολουθώντας την παραπάνω στρατηγική ο παίκτης 1 έχει αυξημένες πιθανότητες να κερδίσει τον αντίπαλό του.

Επιλογή Επόμενης Κίνησης

1) Αν οι αλληλουχίες των νουκλεοτιδίων έχουν το ίδιο μήκος (**n == m → τετραγωνικός πίνακας**):
    a) Αν το **number** ισούται με το μήκος των νουκλεοτιδίων των αλληλουχιών (**n - number == 0**), τότε ο παίκτης 1 θα παίξει διαγώνια και το παιχνίδι θα τερματιστεί.
    b) Στις υπόλοιπες περιπτώσεις **αφαιρεί στην τύχη** από μία αλληλουχία το μήκος **number**.
2) Αν οι αλληλουχίες των νουκλεοτιδίων δεν έχουν το ίδιο μήκος (**n != m → ορθογώνιος πίνακας**):
    a) Αν το **number** είναι μεγαλύτερο από το μήκος της μία εκ των δύο αλληλουχιών (**n - number < 0**) ή (**m - number < 0**), τότε ο παίκτης 1 θα αφαιρέσει **number** νουκλεοτίδια από την άλλη αλληλουχία (**m - number**) ή (**n - number**) αντίστοιχα.
    b) Αν τα μήκη των αλληλουχιών μετά την διαγραφή **number** νουκλεοτιδίων και από τις δύο αλληλουχίες δεν είναι ίσα (**n΄ != m΄ → ορθογώνιος πίνακας**), τότε ο παίκτης 1 μπορεί να επιλέξει μία τυχαία κίνηση (αριστερά, διαγώνια ή πάνω).
    c) Αν τα μήκη των αλληλουχιών μετά την διαγραφή **number** νουκλεοτιδίων από μία από τις δύο αλληλουχίες είναι ίσα αλλά άνισα του 0 (**n == m != 0**), τότε υπάρχει περίπτωση δημιουργίας τετραγωνικού πίνακα και επομένως:
        i) Αν (**n - number == m**) και (**n - number != 0**), τότε ο παίκτης 1 αφαιρεί **number** νουκλεοτίδια είτε διαγώνια είτε προς τα πάνω.

ii) Αν (**m - number == n**) και (**m - number != 0**), τότε ο παίκτης 1 αφαιρεί **number** νουκλεοτίδια είτε διαγώνια είτε προς τα αριστερά.

d) Τέλος, εάν (**n - number == 0**) ή (**m - number == 0**), ο παίκτης 1 επιλέγει να παίξει είτε προς τα πάνω είτε προς τα αριστερά, αντίστοιχα.

Ακολουθώντας την παραπάνω στρατηγική ο παίκτης 1 έχει αυξημένες πιθανότητες να κερδίσει τον αντίπαλό του.

Για την καλύτερη οπτικοποίηση των κινήσεων στο παιχνίδι, μετά την εκτέλεση κάθε παραδείγματος, εκτυπώνουμε δυναμικά, σύμφωνα με την νικηφόρα στρατηγική, τον πίνακα του παιχνιδιού.

Δημιουργία πίνακα

Για την εύρεση της στρατηγικής θα κατασκευάσουμε ένα πίνακα που θα ονομάσουμε **strategy_table**. Ο πίνακας θα γεμίζει με τιμές W ή L, όπου W αντιστοιχεί σε θέση νίκης (winning position) και L σε θέση ήττας (losing position) για τον παίκτη που έχει σειρά να παίξει.

Στη θέση (0,0) ο παίκτης χάνει. Τα κελιά της στήλης 0 και της γραμμής 0 που δεν έχουν συμπληρωθεί με L ή W θα συμπληρωθούν με W γιατί πάντα ο παίκτης που βρίσκεται στις θέσεις αυτές μπορεί να οδηγήσει τον αντίπαλο σε L, την (0,0). Το ίδιο ισχύει και για την θέση (1,1).

Αφού συμπληρωθούν οι παραπάνω θέσεις, μπορούμε να συμπληρώσουμε και τις υπόλοιπες κινήσεις που επιτρέπεται να κάνει η βασίλισσα (προσεγγίζοντας αντίστροφα τη διαδικασία που παίζεται το παιχνίδι, δηλαδή από πάνω αριστερά προς τα κάτω δεξιά). Έτσι, κοιτάζοντας πάντα τις οριζόντιες, κάθετες και διαγώνιες προηγούμενες θέσεις (βάσει των κινήσεων της βασίλισσας) καταλήγουμε στις ακόλουθες περιπτώσεις συμπλήρωσης του πίνακα:

- Αν υπάρχει έστω ένα σύμβολο **L** σε κάποια από τις προηγούμενες θέσεις (οριζόντιες, κάθετες ή διαγώνιες), τότε η τρέχουσα θέση συμπληρώνεται με **W**.
- Αν δεν υπάρχει σύμβολο **L** πουθενά στις προηγούμενες θέσεις (οριζόντιες, κάθετες και διαγώνιες), τότε η τρέχουσα θέση συμπληρώνεται με **L**.

Με τον παραπάνω τρόπο συμπλήρωσης, υλοποιείται ένα παράδειγμα πίνακα με μήκη αλληλουχιών **n = 10** και **m = 12**.

	0	1	2	3	4	5	6	7	8	9	10
0	L	W	W	W	W	W	W	W	W	W	W
1	W	W	L	W	W	W	W	W	W	W	W
2	W	L	W	W	W	W	W	W	W	W	W
3	W	W	W	W	W	L	W	W	W	W	W
4	W	W	W	W	W	W	W	L	W	W	W
5	W	W	W	L	W	W	W	W	W	W	W
6	W	W	W	W	W	W	W	W	W	W	L
7	W	W	W	W	L	W	W	W	W	W	W
8	W	W	W	W	W	W	W	W	W	W	W
9	W	W	W	W	W	W	W	W	W	W	W
10	W	W	W	W	W	W	L	W	W	W	W
11	W	W	W	W	W	W	W	W	W	W	W
12	W	W	W	W	W	W	W	W	W	W	W

→ Θέση έναρξης (starting position)

Τρίτο Θέμα

<u>Άσκηση 6.14:</u> Δύο παίκτες παίζουν το ακόλουθο παιχνίδι με δύο αλληλουχίες που έχουν μήκος n και m νουκλεοτίδια αντίστοιχα. Σε κάθε γύρο του παιχνιδιού, ένας παίκτης μπορεί να αφαιρέσει δύο νουκλεοτίδια από τη μία αλληλουχία (είτε την πρώτη είτε τη δεύτερη) και ένα νουκλεοτίδιο από την άλλη. Ο παίκτης που δεν μπορεί να κάνει κίνηση κερδίζει. Ποιος θα κερδίσει; Περιγράψτε τη νικηφόρα στρατηγική για όλες τις τιμές των n και m.

Έστω ότι έχουμε **n** και **m** τα μήκη των δύο αλληλουχιών.

Ανάλογα με την κίνηση του παίκτη κάθε φορά διαγράφετε 1 νουκλεοτίδιο από την μία αλληλουχία και 2 νουκλεοτίδια από την άλλη αλληλουχία.

Παρατηρούμε ότι οι κινήσεις αυτές παρομοιάζουν κινήσεις ίππου πάνω σε μία σκακιέρα. Έστω λοιπόν ένα παιχνίδι 2 παικτών σε σκακιέρα μεγέθους **n+1** στηλών και **m+1** γραμμών, στην οποία υπάρχει ένας ίππος στην κάτω δεξιά γωνία και στόχος του είναι να καταλήξει στην πάνω αριστερή γωνία. Ο παίκτης που δεν μπορεί να κάνει κίνηση κερδίζει.

Το **+1** στο μέγεθος της σκακιέρας εξυπηρετεί στη σωστή διαχείριση της περίπτωσης, όπου μία από τις δύο ή και οι δύο αλληλουχίες είναι κενές. Μία αλληλουχία θεωρείται κενή όταν ο ίππος βρίσκεται στη θέση μηδέν ενός άξονα μιας αλληλουχίας.

Οι μόνες αποδεκτές κινήσεις του ίππου στο παιχνίδι είναι:

- Να κινηθεί 2 θέσεις αριστερά (**n - 2**) και 1 θέση πάνω (**m - 1**)
- Να κινηθεί 2 θέσεις πάνω (**m - 2**) και 1 θέση αριστερά (**n - 1**)

Κάθε φορά που μετακινείται ο ίππος, **διαγράφονται τα αντίστοιχα νουκλεοτίδια** από το μονοπάτι που διένυσε.

Από υπόθεση γνωρίζουμε πως έχουμε 2 παίκτες. Για λόγους προσομοίωσης παραδείγματος ενός παιχνιδιού, ο ένας παίκτης θα κινείται χωρίς να έχει κάποια στρατηγική (παίκτης 2) και ο άλλος παίκτης θα κινείται με την νικηφόρα στρατηγική (παίκτης 1) που θα περιγράψουμε παρακάτω.

Η νικηφόρα στρατηγική χωρίζεται σε δύο κατηγορίες: <u>επιλογή σειράς</u> και <u>επιλογή επόμενης κίνησης</u>.

<u>Νικηφόρα Στρατηγική</u>

Επιλογή Σειράς

Αν η πρώτη κίνηση στο παιχνίδι αποτελεί θέση νίκης (winning position), τότε ο παίκτης 1 παίζει πρώτος. Αντιθέτως, αν είναι θέση ήττας (losing position), τότε ο παίκτης 2 παίζει πρώτος. Με αυτό τον τρόπο ο παίκτης 1 (που ακολουθεί την νικηφόρα στρατηγική) θα βρίσκεται πάντα σε πλεονεκτική θέση στην αρχή του παιχνιδιού (δεν θα ξεκινήσει ποτέ από θέση ήττας). Αυτό εξασφαλίζεται λαμβάνοντας υπόψη τις παρακάτω περιπτώσεις:

1) Αν οι αλληλουχίες νουκλεοτιδίων έχουν το ίδιο μήκος (**n == m → τετραγωνικός πίνακας**) και ο αριθμός της μιας αλληλουχίας διαιρείται με το 3 και αφήνει υπόλοιπο 2 (**n % 3 == 2**), τότε παίζει δεύτερος.
2) Αν το μήκος μιας αλληλουχίας είναι μεγαλύτερος από το μήκος της άλλης αλληλουχίας m (**n > m → ορθογώνιος πίνακας**) και το μήκος m διαιρείται με το 3 και αφήνει υπόλοιπο 1 (**m % 3 == 1**), τότε παίζει δεύτερος.
3) Αν το μήκος n της μιας αλληλουχίας είναι μικρότερος από το μήκος m της άλλης αλληλουχίας (**n < m → ορθογώνιος πίνακας**) και το μήκος n διαιρείται με το 3 και αφήνει υπόλοιπο 1 (**n % 3 == 1**), τότε παίζει δεύτερος.
4) Αν <u>δεν</u> ισχύουν οι περιπτώσεις **1)**, **2)** και **3)**, τότε παίζει πρώτος.

Επιλογή Επόμενης Κίνησης

1) Αν οι αλληλουχίες των νουκλεοτιδίων έχουν το ίδιο μήκος (**n == m → τετραγωνικός πίνακας**), ο παίκτης επιλέγει τυχαία μία αποδεκτή κίνηση.
2) Αν το μήκος n μιας αλληλουχίας, είναι μεγαλύτερη από το μήκος m της άλλης αλληλουχίας (**n > m → ορθογώνιος πίνακας**), τότε υπάρχουν τα ακόλουθα σενάρια:
   a) Το μήκος m να είναι πολλαπλάσιο του 3 (**m % 3 == 0**), τότε ο παίκτης θα κινηθεί 2 θέσεις πάνω (**m - 2**) και 1 θέση αριστερά (**n - 1**).
   b) Το μήκος m να διαιρείται με το 3 και να αφήνει υπόλοιπο 1 (**m % 3 == 1**), τότε ο παίκτης θα επιλέξει τυχαία μία αποδεκτή κίνηση.
   c) Το μήκος m να διαιρείται με το 3 και να αφήνει υπόλοιπο 2 (**m % 3 == 2**), τότε ο παίκτης θα κινηθεί 2 θέσεις αριστερά (**n - 2**) και 1 θέση πάνω (**m - 1**).
3) Αν το μήκος n μιας αλληλουχίας, είναι μικρότερο από το μήκος m της άλλης αλληλουχίας (**n < m → ορθογώνιος πίνακας**), τότε υπάρχουν τα ακόλουθα σενάρια:
   a) Το μήκος n να είναι πολλαπλάσιο του 3 (**n % 3 == 0**), τότε ο παίκτης θα κινηθεί 2 θέσεις αριστερά (**n - 2**) και 1 θέση πάνω (**m - 1**).

b) Το μήκος n να διαιρείται με το 3 και να αφήνει υπόλοιπο 1 (**n % 3 == 1**), τότε ο παίκτης θα επιλέξει τυχαία μία αποδεκτή κίνηση.
c) Το μήκος n να διαιρείται με το 3 και να αφήνει υπόλοιπο 2 (**n % 3 == 2**), τότε ο παίκτης θα κινηθεί 2 θέσεις πάνω (**m - 2**) και 1 θέση αριστερά (**n - 1**).

<u>Ειδικές περιπτώσεις</u>

4) Αν το μήκος n της μίας αλληλουχίας είναι 1 (**n == 1**), τότε ο παίκτης θα κινηθεί υποχρεωτικά 2 θέσεις πάνω (**m - 2**) και 1 θέση αριστερά (**n - 1**).
5) Αν το μήκος m της μίας αλληλουχίας είναι 1 (**m == 1**), τότε ο παίκτης θα κινηθεί υποχρεωτικά 2 θέσεις αριστερά (**n - 2**) και 1 θέση πάνω (**m - 1**).

Ακολουθώντας την παραπάνω στρατηγική ο παίκτης 1 έχει αυξημένες πιθανότητες να κερδίσει τον αντίπαλό του.

Για την καλύτερη οπτικοποίηση των κινήσεων στο παιχνίδι, μετά την εκτέλεση κάθε παραδείγματος, εκτυπώνουμε δυναμικά, σύμφωνα με την νικηφόρα στρατηγική, τον πίνακα του παιχνιδιού.

Δημιουργία πίνακα

Για την εύρεση της στρατηγικής θα κατασκευάσουμε ένα πίνακα που θα ονομάσουμε **strategy_table**. Ο πίνακας θα γεμίζει με τιμές W ή L, όπου W αντιστοιχεί σε θέση νίκης (winning position) και L σε θέση ήττας (losing position) για τον παίκτη που έχει σειρά να παίξει. Επομένως, συμπληρώνουμε τις θέσεις στην πρώτη γραμμή (γραμμή 0) και πρώτη στήλη (στήλη 0) με W, καθώς είναι τελική κατάσταση και δεν μπορεί να γίνει καμία κίνηση. Το ίδιο ισχύει και για την θέση (1,1). Με μια ματιά ο υπολογισμός του πίνακα $\mathbf{strategy\_table}_{n,m}$ για τυχαίες τιμές των **n** και **m** φαίνεται δύσκολος, αλλά χρησιμοποιώντας προηγούμενες τιμές ως βάση μπορούμε να συνεχίσουμε σταδιακά.

Αφού συμπληρωθούν οι παραπάνω θέσεις, μπορούμε να συμπληρώσουμε και τις υπόλοιπες κινήσεις που επιτρέπεται να κάνει ο ίππος (προσεγγίζοντας αντίστροφα τη διαδικασία που παίζεται το παιχνίδι, δηλαδή από πάνω αριστερά προς τα κάτω δεξιά). Έτσι, κοιτάζοντας πάντα τις 2 προηγούμενες θέσεις (βάσει των κινήσεων του ίππου) καταλήγουμε στις ακόλουθες περιπτώσεις συμπλήρωσης του πίνακα:

- Αν υπάρχει και στις 2 προηγούμενες καταχωρημένες θέσεις το σύμβολο **W**, τότε η τρέχουσα θέση συμπληρώνεται με **L**.
- Αν υπάρχει τουλάχιστον ένα σύμβολο **L** σε μια από τις 2 προηγούμενες καταχωρημένες θέσεις, τότε η τρέχουσα θέση συμπληρώνεται με **W**.

Με τον παραπάνω τρόπο συμπλήρωσης, υλοποιείται ένα παράδειγμα πίνακα με μήκη αλληλουχιών **n = 10** και **m = 12**.

	0	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
0	W	W	W	W	W	W	W	W	W	W	W
1	W	W	L	L	L	L	L	L	L	L	L
2	W	L	L	L	W	W	W	W	W	W	W
3	W	L	L	W	W	W	W	W	W	W	W
4	W	L	W	W	W	L	L	L	L	L	L
5	W	L	W	W	L	L	L	W	W	W	W
6	W	L	W	W	L	L	W	W	W	W	W
7	W	L	W	W	L	W	W	W	L	L	L
8	W	L	W	W	L	W	W	L	L	L	W
9	W	L	W	W	L	W	W	L	L	W	W
10	W	L	W	W	L	W	W	L	W	W	W
11	W	L	W	W	L	W	W	L	W	W	L
12	W	L	W	W	L	W	W	L	W	W	L

→ Θέση έναρξης (starting position)

Βιβλιογραφία - Δικτυογραφία

Θέμα 1

Μπένος Β. (2008), *Εισαγωγή στους Αλγορίθμους Βιοπληροφορικής*, Αθήνα: Κλειδάριθμος

Pevsner J. (2015), Bioinformatics and Functional Genomics, 3η Έκδοση, Hoboken: John Wiley & Sons

Helen Yannakoudakis (2018), *9: Viterbi Algorithm for HMM Decoding*, Cambridge: University of Cambridge ανάκτηση από https://www.cl.cam.ac.uk/teaching/1718/MLRD/slides/slides9.pdf στις 30/05/2023

CS378 Lecture Note: Viterbi Algorithm, ανάκτηση από https://www.cs.utexas.edu/~gdurrett/courses/sp2020/viterbi.pdf στις 29/05/2023

HMM : Viterbi algorithm - a toy example, ανάκτηση από https://www.cis.upenn.edu/~cis2620/notes/Example-Viterbi-DNA.pdf στις 1/06/2023

Θέμα 2

Μπένος Β. (2008), *Εισαγωγή στους Αλγορίθμους Βιοπληροφορικής*, Αθήνα: Κλειδάριθμος

CS431 homework 2, ανάκτηση από https://www.cs.bu.edu/faculty/homer/431/homework/answers-hw2.pdf στις 15/06/2023

Θέμα 3

Μπένος Β. (2008), *Εισαγωγή στους Αλγορίθμους Βιοπληροφορικής*, Αθήνα: Κλειδάριθμος

CS431 homework 2, ανάκτηση από https://www.cs.bu.edu/faculty/homer/431/homework/answers-hw2.pdf στις 10/06/2023