

L'accès à internet est autorisé mais l'utilisation d'intelligences artificielle génératives de type ChatGPT est interdite.

À l'issue de l'épreuve, vous devez envoyer votre TP au format pdf à Xavier Dupré :
xavier.dupre@gmail.com

Barème prévisionnel : toutes les questions valent deux points.

Exercice 1 : recherches de motifs

On suppose qu'on a une séquence d'ADN "ttcagttgtg aatgaatgga cgtgccaaat agacgtgccg cgcgcgtcg attcgactt cgtgccaaat". Les caractères a, t, g, c sont appelés bases (nucléiques) et le fragment d'ADN est donc une séquence de bases.

On s'intéresse dans ce sujet à la recherche d'un motif, c'est-à-dire d'une sous-séquence de bases, par exemple ccaa.

On pourra écrire :

```
text = "ttcagttgtg aatgaatgga cgtgccaaat agacgtgccg cgcgcgtcg attcgactt cgtgccaaat".replace(" ", "").
```

Q1: Écrire une fonction qui retourne la première position du motif en utilisant une boucle et sans la méthode index

La fonction retourne la longueur du texte cherché si le motif n'est pas trouvé.

```
def recherche_motif(text, motif):
```

Q2 : Quel est, dans le pire des cas, le coût de l'algorithme ?

Q3 : On souhaite maintenant écrire une fonction qui retourne toutes les positions d'un motif au sein d'un texte.

```
def positions_motif(text, motif):
```

Q4 : Quel est, dans le pire des cas, le coût de l'algorithme ?

Q5 : Comment modifier (ou pas) votre fonction pour cet exemple ?

```
text = "cgtgccaaaccaaacc"
```

```
motif = "ccaaac"
```

```
assert positions_motif(text, motif) == [4, 9]
```

Q6 : On veut accélérer le processus. On considère un couple de bases et non plus une seule base.

Écrire une fonction une fonction qui construit tous les couples de bases à partir des bases a,c,g,t.

Q7 : Transformer cette liste en dictionnaire.

Le dictionnaire devra avoir pour clé un couple, pour valeur un nombre entier unique.

Q8 : Écrire une fonction qui découpe une séquence de bases en couple de bases.

Chaque gène sera représenté par le nombre entier défini à la question précédente. On supposera que la séquence est de longueur paire. On fera de même pour le motif.

Q9 : Appliquer l'une de vos précédentes fonctions de recherche avec ces deux séquences pour trouver les positions.

Q10 : Comparer le coût des deux recherches pour des séquences et motifs de longueur paire ?

- Quelle recherche est plus rapide ? Combien de fois plus rapide ?
- Pourquoi ces deux recherches ne sont pas équivalentes ?

Q11 : Réécrire la fonction de la question 7 en ajoutant un argument n qui représente la longueur des motifs. La fonction doit alors renvoyer un dictionnaire dont les clés sont toutes les chaînes de longueur n .