

## TP5 : Programmation dynamique

Le but de ce TP est d'implanter le calcul de la distance d'édition vu en cours, ainsi que l'alignement de deux mots. Les fichiers suivants sont fournis :

- `edition.h` (à ne pas modifier) et `edition.cpp` (**seul fichier à modifier**) pour l'implantation des algorithmes vus en cours;
- le fichier de test `tests.cpp` et le `Makefile` (à ne pas modifier);
- un dossier `covid19/` pour la question 4.

Le programme principal ne prend pas de paramètre et permet de tester indépendamment chaque fonction. **Testez donc votre code à chaque question** (et pas uniquement avec les EXEMPLES encadrés) !

**Utilisation de la bibliothèque `string`** Les chaînes de caractères sont représentées par le type `string`, donc les principales méthodes nécessaires à ce TP sont fournies ci-dessous.

```
string s = "algo";
int n = s.size();    // n contient 4
s.insert(4,"t");     // insère un "t" en dernière position
s.insert(2,"i");     // insère un "i" en position 2
cout << s << endl; // affiche s (et vive l'aligot !)
```

1. Compléter la fonction `int** matriceDistances(string s1, string s2)` qui calcule la matrice des distances (E dans le cours). *Attention, vous devez allouer la mémoire nécessaire et initialiser la matrice.*

EXEMPLE

```
Entrer le numéro de question à tester (0 pour quitter) : 1
Entrer la première chaîne de caractère : AGORRYTNES
Entrer la seconde chaîne de caractère : ALGORITHMES
Matrice des distances :
0 1 2 3 4 5 6 7 8 9 10 11
1 0 1 2 3 4 5 6 7 8 9 10
2 1 1 1 2 3 4 5 6 7 8 9
3 2 2 2 1 2 3 4 5 6 7 8
4 3 3 3 2 1 2 3 4 5 6 7
5 4 4 4 3 2 2 3 4 5 6 7
6 5 5 5 4 3 3 3 4 5 6 7
7 6 6 6 5 4 4 3 4 5 6 7
8 7 7 7 6 5 5 4 4 5 6 7
9 8 8 8 7 6 6 5 5 5 6
10 9 9 9 8 7 7 6 6 6 6 5
```

2. Compléter la fonction `int alignement(string& s1, string& s2)` qui calcule un alignement des deux chaînes `s1` et `s2`. *La fonction modifie `s1` et `s2` en insérant des caractères `_` (underscore) lorsque nécessaire et elle renvoie la distance entre `s1` et `s2`. Penser à désallouer la mémoire à la fin de la fonction !*

EXEMPLE

```
Entrer le numéro de question à tester (0 pour quitter) : 2
Entrer la première chaîne de caractère : AGORRYTNES
Entrer la seconde chaîne de caractère : ALGORITHMES
Distance entre les deux chaînes : 5
```

```
Alignement : A_GORRYTNES
              ALGORITHMES
```

3. Compléter la fonction `int distanceEdition(string s1, string s2)` qui calcule la distance d'édition entre les deux chaînes de caractères, à l'aide de l'**algorithme qui minimise la mémoire**. Vérifier que les résultats obtenus sont les mêmes qu'avec l'algorithme précédent !

EXEMPLE

```
Entrer le numéro de question à tester (0 pour quitter) : 3
Entrer la première chaîne de caractère : AGORRYTNES
Entrer la seconde chaîne de caractère : ALGORITHMES
Distance entre les deux chaînes : 5
```

4. On va enfin utiliser notre programme de manière utile ! Le dossier covid19 contient les génomes complets de plusieurs coronavirus :

- 2019-nCoV\_WH01.fa, 2019-nCoV\_WH03.fa et 2019-nCoV\_WH04.fa sont des virus prélevés chez des patients chinois lors de l'épidémie actuelle de Covid-19 ;
- bat-SL-CoVZC45.fa et bat-SL-CoVZXC21.fa sont des virus de chauve-souris, séquencés en 2018 ;
- bat-CoV-RaTG13.fa est un virus de chauve-souris, séquencé en février 2020 ;
- MERS-CoV.fa et SARS-CoV.fa sont des virus prélevés chez des patients lors des épidémies de MERS (Moyen-Orient, 2012) et de SRAS (Asie, 2003).

✎ À l'aide de ces fichiers, tenter de déterminer l'origine du virus actuel<sup>1</sup> : mutation à partir d'un virus porté par les chauve-souris, ou mutation à partir d'un virus humain (MERS-CoV ou SARS-CoV).

Note. Vous pourrez comparer vos résultats avec l'étude dont sont tirés ces fichiers, disponible à l'adresse [https://www.thelancet.com/pdfs/journals/lancet/PIIS0140-6736\(20\)30251-8.pdf](https://www.thelancet.com/pdfs/journals/lancet/PIIS0140-6736(20)30251-8.pdf).

EXEMPLE

```
Entrer le numéro de question à tester (0 pour quitter) : 4
Liste des fichiers :
  1. 2019-nCoV_WH01.fa
  2. 2019-nCoV_WH03.fa
  3. 2019-nCoV_WH04.fa
  4. bat-SL-CoVZC45.fa
  5. bat-SL-CoVZXC21.fa
  6. bat-CoV-RaTG13.fa
  7. MERS-CoV.fa
  8. SARS-CoV.fa
Premier choix : 1
Second choix : 2
La séquence 2019-nCoV_WH01.fa est de longueur 29866.
La séquence 2019-nCoV_WH03.fa est de longueur 29868.
Calcul en cours... terminé en 10.0343 s
Distance entre les deux génomes : 9.
```

1. Cette question est principalement pour tester l'efficacité de votre programme : les conclusions *biologiques* ne sont pas forcément très pertinentes. Les véritables études phylogénétiques reposent sur des techniques ressemblantes, mais plus évoluées.