PLUS LONGUE SOUS-SÉQUENCE COMMUNE

On appelle sous-séquence commune d'une chaine de caractères `ch` toute chaîne de caractères composées de caractères de `ch`, pris dans l'ordre croissant des indices, mais pas nécessairement contigus. Par exemple `'yon'` est une sous-séquence de la chaîne `'python'`.
On s'intéresse dans ce problème à la détermination d'une plus longue sous-séquence commune (PLSSC) de deux chaînes de caractères `ch1` et `ch2`, qui ne sont pas nécessairement de mêmes longueurs. La longueur d'une PLSSC permet par exemple d'évaluer la proximité de deux individus en déterminant la longueur d'une PLSSC de leurs brins d'ADN. Une PLSSC de `'agtaatatcgta'` et `'gattcatacagt'` est `'atatacgt'`.
Dans toute la suite, on notera n la longueur de la chaîne `ch1` et m la longueur de `ch2`.
Pour toute question sur la complexité des fonctions, on donnera une réponse sous la forme $O(f(n,m))$, où f est une certaine fonction simple.

1. Combien y a-t-il de sous-séquences de `ch1` ?

En admettant qu'il est possible de vérifier si une chaîne `ch` est une sous-séquence de `ch2` avec une complexité $O(m)$, quelle serait, en fonction de m et n, la complexité d'une fonction déterminant une PLSSC de `ch1` et `ch2` en testant, pour toute sous-séquence de `ch1`, si elle appartient à `ch2` avant de déterminer la plus longue d'entre elles ?

Pour $i \in [\![0,n]\!]$ et $j \in [\![0,m]\!]$, on note $\ell_{i,j}$ la longueur d'une PLSSC des chaînes `ch1[0:i]` et `ch2[0:j]`

2. Donner la matrice $(\ell_{i,j})_{(i,j)\in[\![0,n]\!]\times[\![0,m]\!]}$ lorsque `ch1 = 'mpsi'` et `ch2 = 'pcsi'`. Quelle est la longueur d'une PLSSC de `ch1` et `ch2` ?

3. a) Justifier, rapidement, que $\ell_{i,j}$ vérifie les relations suivantes :

$$\ell_{i,j} = \begin{cases} 0 & \text{si } i=0 \text{ ou } j=0 \\ 1+\ell_{i-1,j-1} & \text{si } ij\neq 0 \text{ et } \texttt{ch1[i-1]} = \texttt{ch2[j-1]} \\ \max\{\ell_{i-1,j},\ell_{i,j-1}\} & \text{sinon} \end{cases}$$

b) Pourquoi peut-on dire que le problème de la détermination d'une PLSSC présente la propriété de sous structure optimale ?

c) Écrire une fonction `tableauPLSSC(ch1:str,ch2:str) -> list` qui prend en argument deux chaînes `ch1` et `ch2` et qui renvoie une liste de liste `T` de sorte que `T[i][j]` $= \ell_{i,j}$ pour $i \in [\![0,n]\!]$ et $j \in [\![0,m]\!]$.

Pour simplifier l'écriture de cette fonction, on pourra utiliser la fonction `max(L:list) -> int` qui renvoie le maximum d'une liste d'entiers.

Quelle est, en fonction de n et m la complexité de cette fonction ?

d) En déduire une fonction `longueurPLSSC(ch1:str,ch2:str) -> int` qui prend en argument deux chaînes `ch1` et `ch2` et qui renvoie la longueur d'une PLSSC de `ch1` et `ch2`.

4. a) Si on suppose que `ch1 = 'info'` et que le « tableau » renvoyé par la fonction `tableauPLSSC(ch1,ch2)` est le suivant :

0	0	0	0	0	0	0
0	0	0	1	1	1	1
0	0	0	1	1	2	2
0	0	0	1	1	2	2
0	0	0	1	1	2	3

Donner une PLSSC de `ch1` et `ch2`.

b) Écrire une fonction `PLSSC(ch1:str,ch2:str) -> str` qui prend en argument deux chaînes `ch1` et `ch2` et qui renvoie une PLSSC de `ch1` et `ch2`. Cette fonction devra obligatoirement utiliser la fonction `tableauPLSSC`.

c) Quelle est, en fonction de n et m la complexité de cette fonction ?

5. Proposer une fonction `PLSSC2(ch1:str,ch2:str) -> str` récursive qui renvoie une PLSSC de `ch1` et `ch2` ; afin d'optimiser les calculs, on utilisera la technique de mémoïsation à l'aide d'un dictionnaire.