Distance d'édition

Soient `ch1` et `ch2` deux chaînes de caractères. On cherche à transformer `ch1` en `ch2` en n'utilisant que trois transformations élémentaires :

— la substitution d'un caractère de `ch1` par un autre caractère de l'alphabet

— l'insertion d'un caractère de l'alphabet à n'importe quel endroit dans `ch1`

— la suppression de n'importe lequel des caractères de `ch1`

La distance de Levenshtein (ou distance d'édition) entre `ch1` et `ch2` est le nombre minimal de transformations nécessaires pour passer de `ch1` à `ch2`.
Par exemple, la distance de Levenshtein de `'distance'` à `'édition'` est 6 :

$$\texttt{'distance'} \overset{\text{supp}}{\to} \texttt{'distanc'} \overset{\text{supp}}{\to} \texttt{'distan'} \overset{\text{ins}}{\to} \texttt{'distaon'} \overset{\text{modif}}{\to} \texttt{'distion'} \overset{\text{supp}}{\to} \texttt{'dition'} \overset{\text{ins}}{\to} \texttt{'édition'}$$

Et il est possible de vérifier que 5 transformations ne sont pas suffisantes.

1. Sous-structure optimale
On note, pour $0 \leqslant i \leqslant n = \texttt{len(ch1)}$ et $0 \leqslant j \leqslant m = \texttt{len(ch2)}$, $d_{i,j}$ la distance d'édition de `ch1[:i]` à `ch2[:j]`.

a) Que valent $d_{0,j}$ et $d_{i,0}$?

b) Si `ch1[i-1] == ch2[j-1]` que vaut $d_{i,j}$ en fonction de $d_{i-1,j-1}$?

c) Sinon, quelle est la valeur de $d_{i,j}$ en fonction de $d_{i-1,j}$, $d_{i,j-1}$ et $d_{i-1,j-1}$?

d) En déduire une fonction `tableauLEV(ch1:str,ch2:str)->list` qui prend en argument deux chaînes de caractères et qui renvoie le tableau `D` (liste de liste) tel que `D[i][j]` $= d_{i,j}$

e) Écrire une fonction `distLEV(ch1:str,ch2:str)->int` qui renvoie la distance d'édition de `ch1` à `ch2`.

f) Évaluer la complexité de cette fonction en fonction de n et m.

2. Écrire une fonction récursive `distLEV_rec(ch1:str,ch2:str)->int` « efficace » qui renvoie la distance d'édition de `ch1` à `ch2`.

3. a) On admet que le tableau renvoyé par `tableauLEV('distance','édition')` est le suivant :

0	1	2	3	4	5	6	7
1	1	1	2	3	4	5	6
2	2	2	1	2	3	4	5
3	3	3	2	2	3	4	5
4	4	4	3	2	3	4	5
5	5	5	4	3	3	4	5
6	6	6	5	4	4	4	4
7	7	7	6	5	5	5	5
8	8	8	7	6	6	6	6

Comment est-il possible de retrouver les transformations à faire pour passer de `'distance'` à `'édition'`

b) Écrire une fonction `transfLEV(ch1:ch2)->list` qui renvoie la liste des « mots » permettant de passer de `ch1` à `ch2`.
`transfLEV('distance','édition')` pourra renvoyer, par exemple (la suite des transformations n'est pas unique), la liste

`['distance','distanc','distan','distaon','distion','dition',édition']`.