



## INSTITUT SUPERIEUR POLYTECHNIQUE DE MADAGASCAR

AMBATOMARO - ANTSOBOLO - ANTANANARIVO

### Algorithmique Avancée – Informatique et Télécommunication L3

Programmation Dynamique – Application de la distance de Levenshtein

Langage à utiliser : JAVA ou C#

On donne le code JAVA suivant :

```
import java.util.*;
class LevenshteinDistanceDP {
    static int compute_Levenshtein_distanceDP(String str1, String str2)
    {
        // A 2-D matrix to store previously calculated
        // answers of subproblems in order
        // to obtain the final
        int[][] dp = new int[str1.length() + 1][str2.length() + 1];
        for (int i = 0; i <= str1.length(); i++)
        {
            for (int j = 0; j <= str2.length(); j++) {
                // If str1 is empty, all characters of
                // str2 are inserted into str1, which is of
                // the only possible method of conversion
                // with minimum operations.
                if (i == 0) {
                    dp[i][j] = j;
                }
                // If str2 is empty, all characters of str1
                // are removed, which is the only possible
                // method of conversion with minimum
                // operations.
                else if (j == 0) {
                    dp[i][j] = i;
                }
                else {
                    // find the minimum among three
                    // operations below
                    dp[i][j] = minm_edits(dp[i - 1][j - 1]
                        + NumOfReplacement(str1.charAt(i - 1), str2.charAt(j - 1)), // replace
                        dp[i - 1][j] + 1, // delete
                        dp[i][j - 1] + 1); // insert
                }
            }
        }
        return dp[str1.length()][str2.length()];
    }
    // check for distinct characters
    // in str1 and str2
    static int NumOfReplacement(char c1, char c2)
    {
        return c1 == c2 ? 0 : 1;
    }
    // receives the count of different
    // operations performed and returns the
    // minimum value among them.
    static int minm_edits(int... nums)
    {

```

```

        return Arrays.stream(nums).min().orElse(
            Integer.MAX_VALUE);
    }
    // Driver Code
    public static void main(String args[])
    {
        String s1 = "glomax";
        String s2 = "folmax";
        System.out.println(compute_Levenshtein_distanceDP(s1, s2));
    }
}

```

Source : <https://www.geeksforgeeks.org/java-program-to-implement-levenshtein-distance-computing-algorithm/>

- 1) Donner trois (03) « faits » qui vont ont surprises dans ce programme.  
Expliquer en une phrase chaque « fait ».
- 2) Considérer la liste de mots en français : `liste_francais.txt`.  
(source <http://www.3zsoftware.com/fr/listes.php>)  
Observer le Format de ce fichier.  
Ecrire un code qui permet de charger les mots existants dans ce fichier dans une structure de données en mémoire vive (exemples : `ArrayList<>` et/ou `HashSet<>` en JAVA ou `List<>` et/ou `Dictionary<>` en C#).
- 3) Ecrire un sous-programme qui permet de tester si un mot donné par l'utilisateur existe dans le dictionnaire ou pas.
- 4) Ecrire un sous-programme qui permet de donner une liste de propositions de quelques mots les plus proches, selon la distance de Levenshtein dans le cas où un mot donné n'existe pas dans le dictionnaire.
- 5) Ecrire un sous-programme qui teste tous les mots (séparés par un caractère 'espace' ou autres séparateurs de votre choix) dans un fichier texte donné (par exemple : `texte.txt`) et produit une liste de corrections des éventuels mots erronés dans un autre fichier (par exemple `texte.err`) avec les propositions de remplacements, selon la distance de Levenshtein.

#### **Exemple :**

##### **Texte.txt**

Je suis trs en fomre

##### **Texte.err**

trs - très, tris, tas, trop.  
fomre - forme, forma, formée.

- 6) Utiliser maintenant, le fichier `gutemberg.txt`. (source <http://www.3zsoftware.com/fr/listes.php>)  
Quelle est la différence entre `liste_francais.txt` et `gutemberg.txt` ?

Réviser vos réponses des questions 2) à 5) afin d'en optimiser le temps d'exécution avec le nouveau fichier.