

LO17 : Indexation et Recherche d'information

TD5 : Correcteur Orthographique

Printemps 2025

Vous devez réaliser un code Python qui permettra d'associer un lemme à chacun des mots d'une phrase saisie au clavier. Pour cette première étape, vous définirez manuellement un fichier d'une ou deux dizaines de mots et de lemmes judicieusement choisis pour tester vos algorithmes.

L'analyseur doit :

1. saisir la phrase ;
2. extraire les mots un à un jusqu'à la fin de la phrase et, pour chaque mot, réaliser les opérations suivantes :
 - (a) le convertir en minuscules ;
 - (b) tester si le mot existe dans le lexique et si oui retourner le lemme ;
 - (c) si non, compter le nombre de lettres communes entre le début du mot et le début d'un mot du lexique (algorithme de recherche par préfixe vu en cours) :
 - i. stocker le mot dans une liste de mots candidats ;
 - ii. retourner le lemme du meilleur mot candidat ;
 - (d) si plusieurs mots sont possibles, utiliser l'algorithme de Levenshtein pour choisir le meilleur mot candidat et retourner son lemme.
 - (e) si non, indiquer qu'aucun mot n'a été trouvé.

Lorsque vos algorithmes ont été testés, appliquez votre programme au lexique complet des lemmes du corpus.