# Porter Stemmer et son implémentation en Snowball

* Introduction :

L'algorithme de Porter Stemmer est un algorithme de racinisation populaire utilisé en traitement automatique du langage naturel. Il a été développé par Martin Porter dans les années 1980 et est largement utilisé pour réduire les mots à leur forme de base ou racine, ce qui facilite le traitement et l'analyse des textes.

* L'algorithme de Porter Stemmer :

L'algorithme de Porter Stemmer suit une approche en mode arrière (backward) et utilise une série d'étapes pour manipuler les mots en fonction de certains motifs et conditions. L'objectif principal de l'algorithme est de réduire les mots à leur forme de base tout en conservant leur compréhensibilité.

Les étapes de l'algorithme de Porter Stemmer comprennent des règles spécifiques pour la suppression, la modification ou le remplacement de certains suffixes ou préfixes, certaines étapes peuvent supprimer les terminaisons plurielles ou convertir les mots en formes verbales ou adverbiales.

* Implémentation en Snowball :

Snowball est un langage de programmation spécialement conçu pour l'implémentation d'algorithme de racinisation et de lemmatisation. Il prend en charge plusieurs langages, y compris l'anglais, le français, l'espagnol, etc.

L'implémentation en Snowball de l'algorithme de Porter Stemmer suit les mêmes principes et étapes que l'algorithme d'origine. Snowball fournit un ensemble de règles pré-définies pour différentes langues, ce qui facilite son utilisation dans des applications réelles.

* Étapes de l'algorithme de Porter Stemmer en Snowball :
* Étape 1 : Suppression des suffixes communs tels que "s", "es", "ed", "ing", etc.

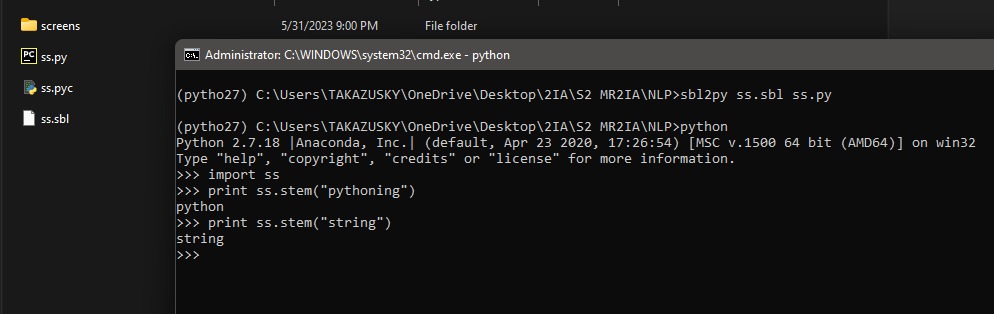
On rappelle que delete est « 🡨 »

* Étape 2 : Traitement des cas particuliers et suppression de certains suffixes ou préfixes spécifiques.
* Étape 3 : Réduction des mots à leur forme de base en supprimant des suffixes tels que "al", "ance", "ence", etc.
* Étape 4 : Réduction des mots à leur forme de base en supprimant d'autres suffixes courants tels que "ion", "er", "able", etc.
* Étape 5 : Application de règles supplémentaires pour supprimer certains suffixes ou préfixes spécifiques.

Il convient de noter que les étapes exactes peuvent varier en fonction de la langue spécifique pour laquelle l'algorithme est implémenté.

* Conclusion :

L'algorithme de Porter Stemmer est un outil puissant pour la racinisation des mots dans le traitement automatique du langage naturel. Son implémentation en Snowball facilite son utilisation dans différents langages de programmation.



-On prendre le scripte de l’algorithme de Porter Stemmer, et en enregistrée le dans un fichier de extension (.sbl).

-On créer enivrement dans Anaconda on prendre notre exemple de (python 2.7) .

-On installer le package sbl2py dans python (enivrement).

-On entrer dans le terminale de cmd, et en changer le chemin de notre répertoire à le chemin de notre fichier de l’algorithme .

-La commande sbl2py ss.sbl ss.py est transforment notre extension de sbl à py.

-Donc on importer le package ss ( snowball) avec python

-On prendre cette exemple de mot de « pythoning »,et ce mot est transformera le mot «  string »