جامعة شعيب الدكالي

ⵜⴰⵙⴷⴰⵡⵉⵜ ⵛⵄⵢⴰⴱ ⴷⴽⴰⵍⵉ

Université Chouaïb Doukkali

Projet : Réalisation d'une application d'indexation sémantique

Module :

Recherche d'information Et

Indexation

BIBDA

BUSINESS INTELLIGENCE AND BIG DATA ANALYTICS

Encadré par :

- dd

Realisé par :

HAJJAJI Nassim

Dans ce projet j'ai essayé de réaliser une application de l'indexation sémantique, et la recherche d'information, à l'aide d'une source externe : « Word net »

Dans ce guide je vais vous donner les outils et les méthodes pour le bon fonctionnement de l'application :

Au début je vous conseille d'utiliser ANACONDA, est une distribution libre et open source du langage de programmation Python. C'est un logiciel qui facilite plusieurs taches au niveau d'installation des packages.

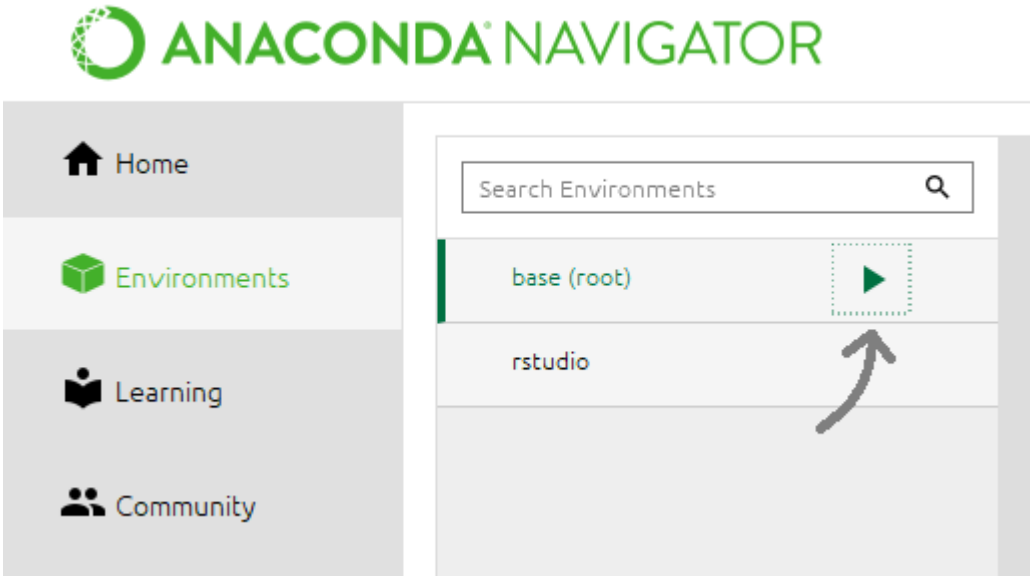
1ére étape : Installation ANACONDA :

Selon votre machine et votre système d'exploitation choisissez laquelle la distribution qui vous convient, ***attention*** vous devez télécharger ***la version 3.7***. Voilà le lien qui vous amène:

https://www.anaconda.com/distribution/#download-section

2éme étape : Installation NLTK :

Pour cela il faut d'abord ouvrir ANACONDA et cliquez sur le bouton Environnement qui se trouve à gauche. Après vous ouvrirez le terminal : en cliquant sur le triangle indiqué dans la figure ci-dessous choisissez « *Open Terminal* ».



Quand le terminal est ouvert coller la commande suivante :

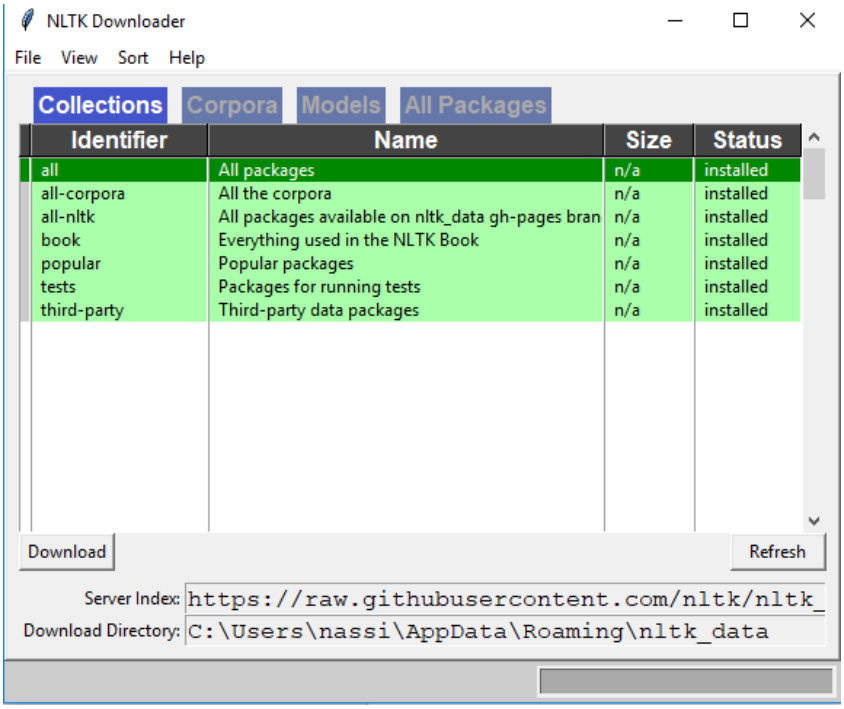
conda install -c anaconda nltk

À un moment le programme va vous afficher une question de vérification de téléchargement " Proceed *([y]/n)*". Vous répondez par "y" (ça veut dire yes).

Maintenant cous cliquer sur le même triangle ci-dessus mais cette fois ci choisisser "Open with Python" quand le terminal est ouvert vous tapez ou collez les commandes suivantes:

```
>>> import nltk
>>> nltk.download()
```

Une fenêtre s'affichera, choisissez "all" et cliquez sur le bouton download qui se trouve en bas à gauche.



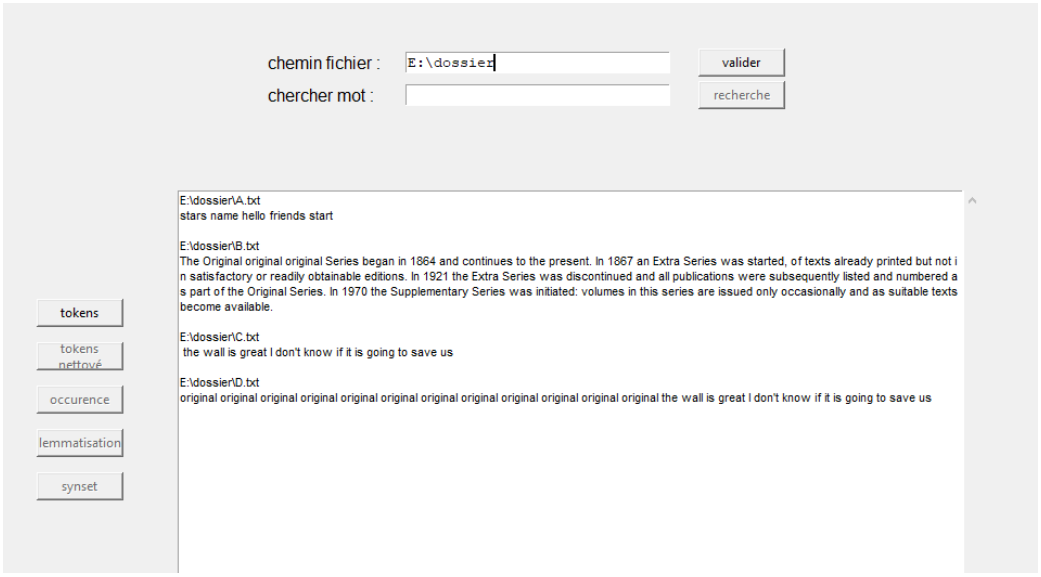
A ce niveau-là vous êtes normalement capable d'exécutez l'application sans aucun problème

3éme étape : exécution de l'application :

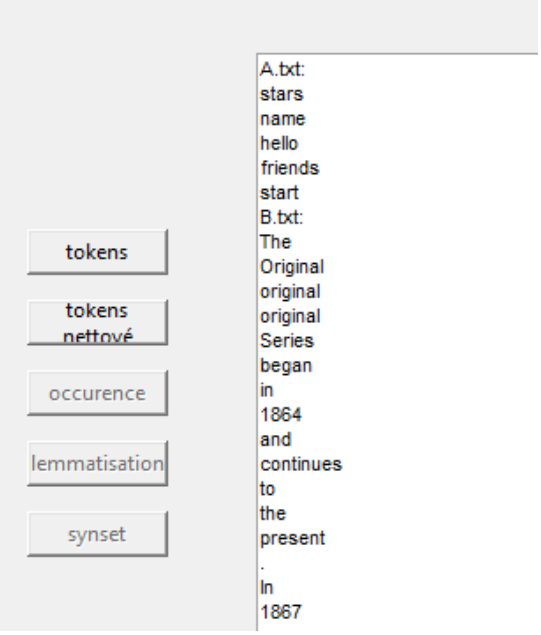
Vous trouverez l'application dans le dossier que je vous ai envoyé, le nom du fichier est "app_rech_sem.py". Ouvrez-le via "Spyder" un outil qui se trouve dans ANACONDA et vous exécutez le programme (sinon vous pouvez exécuter l'application en faisant double clics sur le programme).

Une fenêtre s'affichera avec un champ où vous spécifiez le chemin vers le dossier qui contient votre fichiers ".txt", quand vous validez les bouton d'indexation seront activés avec cet ordre : quand vous cliquez sur le bouton "tokens" il activera le bouton "tokens nettoyés " et ainsi de suite jusqu'à le dernier bouton "synset". Pour chaque bouton cliqué les résultats seront affichées.

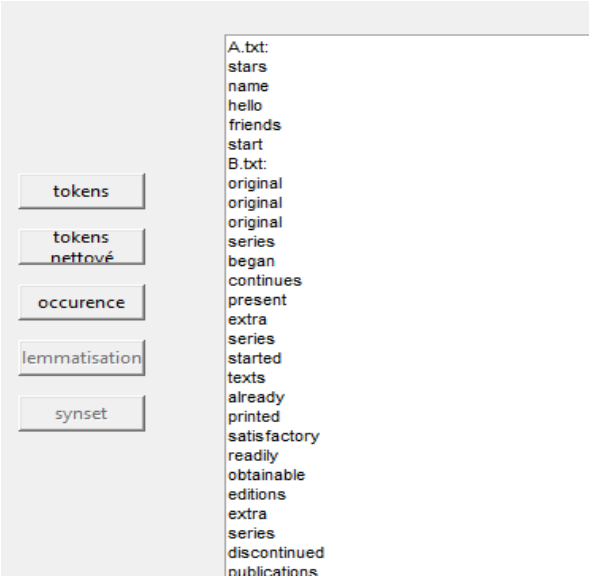
Quand vous valider le chemin le texte de chaque fichier sera affiché:



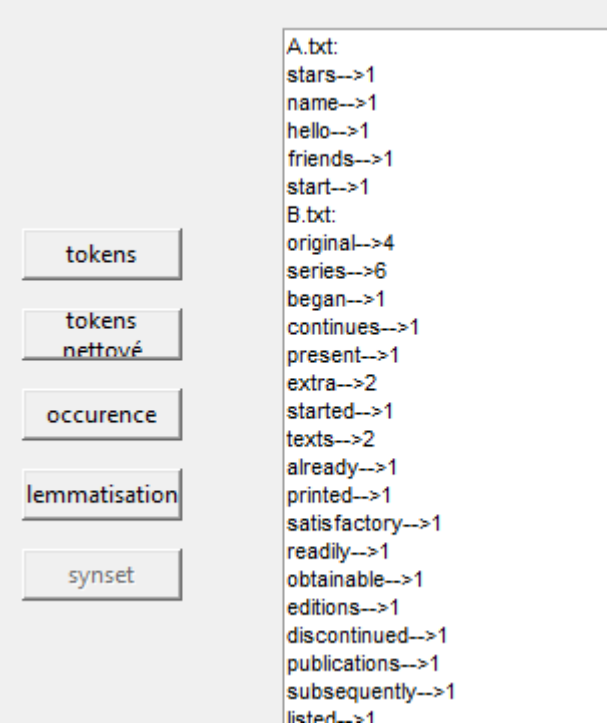
Quand vous cliqué sur toukens:



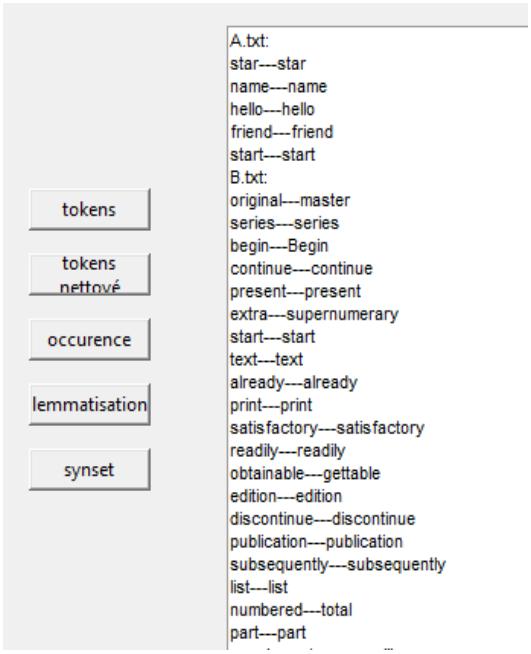
toukens nettoyés:



Occurrences :



Lemmatisation et Synset :



Après, le bouton recherche sera activez, maintenant vous pouvez faire des recherches sur les fichiers du dossier désigné au début.