

# Procédure d'installation pour utiliser les applications de texte-mining

## **Introduction :**

Ce document s'inscrit dans le périmètre du projet tutoré de l'un des groupes de la classe d'Année Spéciale de l'IUT Informatique LYON 1. Réunissant :

- DUFOUR Alexandre
- FERRAND Gérome
- LAURENT Jeremy
- MOUSSU Nathan
- VATON Juliette

Ce groupe était placé sous le tutorat de M. VIDAL VINCENT.

Vous trouverez sur le dépôt 2 dossiers :

- Application générale (permet de lancer n'importe quelle application depuis cette dernière)
- Applications individuelles (permet d'avoir accès aux codes de chaque application et à la réflexion apporté à ces dernières)
  - Moteur de recherche (affecté à DUFOUR Alexandre)
  - Détecteur de spams (affecté à LAURENT Jérémy)
  - Analyseur d'opinion (affecté à VATON Juliette)
  - Catégorisation de texte (affecté à FERRAND Gérome)

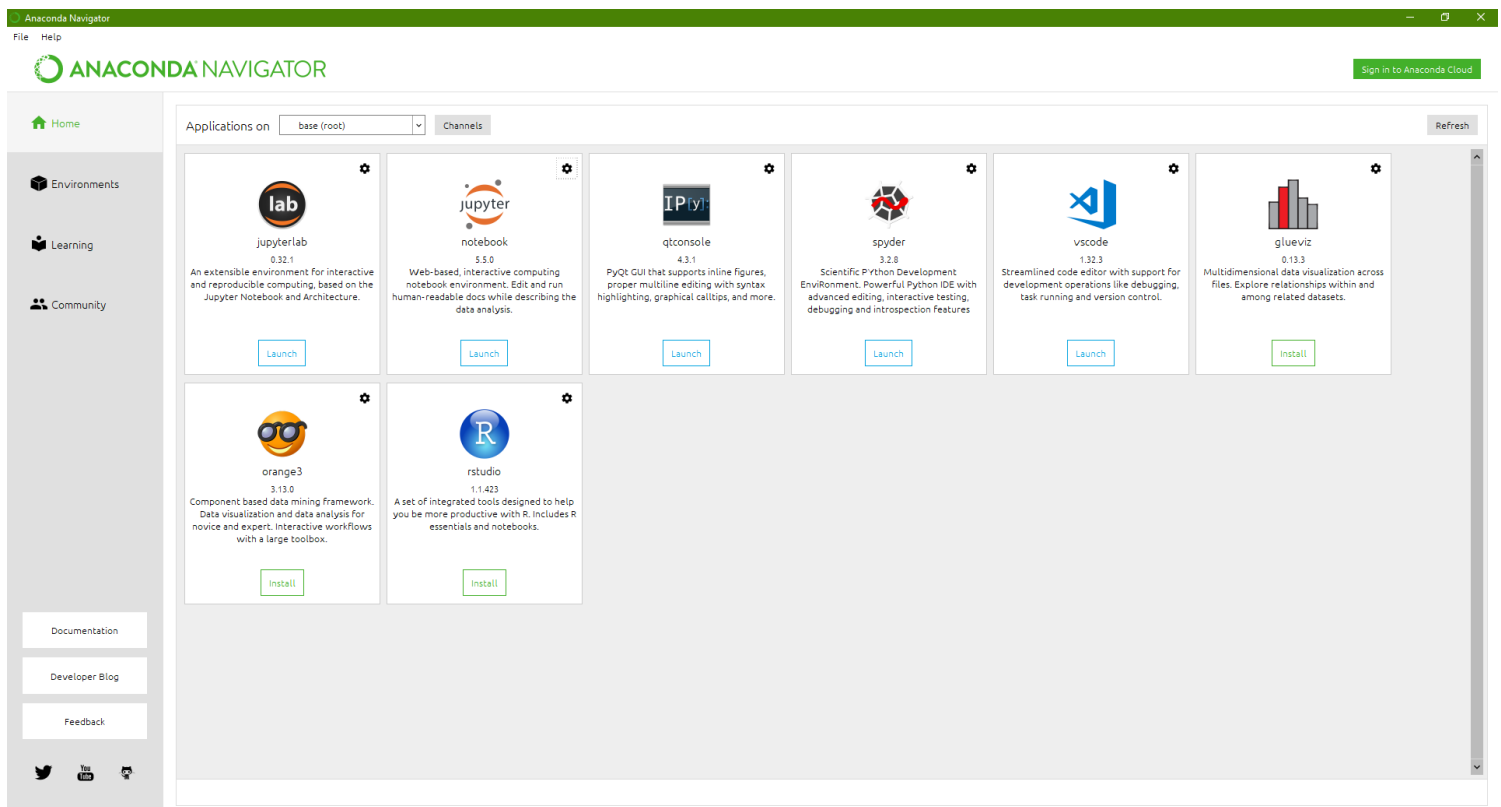
## Étape n°1 : Installation d'Anaconda navigateur

Cette étape est facultative si vous avez déjà un IDE python (exemple : Spyder) et que vous souhaitez juste utiliser l'application sans comprendre le code. Néanmoins elle permet d'avoir accès à tout la suite anaconda en une application (Notebook Jupyter, Spyder etc....)

*Avantages : Anaconda est une distribution Python. A son installation, Anaconda installera Python ainsi qu'une multitude de packages (voir liste de packages anaconda). Cela nous évite de nous ruer dans les problèmes d'incompatibilités entre les différents packages.*

Lien de téléchargement : <https://www.anaconda.com/distribution/>

Après l'installation vous arriverez sur le menu suivant :



## Étape n°2 : Mise à jour des bibliothèques

Après l'installation d'Anaconda, il est toujours utile de s'assurer qu'on a les dernières versions des packages qu'on sera amené à utiliser. Dans notre cas nous aurons besoin des librairies suivantes :

- NumPy
- SciPy
- Matplotlib
- Pandas
- Scikit-learn
- Statsmodels
- NLTK

Avec Conda, il est possible de télécharger la dernière version d'un package donné à l'aide de la commande suivante (windows CMD) :

`conda update PACKAGE_NAME`

Exemple pour NLTK:

```
conda install -c anaconda nltk
```

Puis valider les différentes étapes :

```

Anaconda Prompt
(base) C:\Users\Admin>conda install -c anaconda nltk
Solving environment: done

The following packages will be downloaded:

package | build | size
-----|-----|-----
vs2015_runtime-15.5.2 | 3 | 2.2 MB
vc-14.1 | h21ff451_3 | 5 KB
rstudio-1.0.153 | 1 | 49.6 MB
nltk-3.3.0 | py36_0 | 2.0 MB
fonts-continuum-1 | 0 | 2 KB
certifi-2018.8.24 | py36_1 | 140 KB
-----|-----|-----
Total: | | 54.0 MB

The following NEW packages will be INSTALLED:

fonts-continuum: 1-0 anaconda

The following packages will be UPDATED:

ca-certificates: 2017.08.26-h94faf87_0 --> 2018.03.27-h5f8d65e_0
certifi: 2018.1.18-py36_0 --> 2018.8.24-py36_1
nltk: 3.2.5-py36h76d52bb_0 --> 3.3.0-py36_0
openssl: 1.0.2n-h74b6da3_0 --> 1.0.2p-h74b6da3_0
vc: 14-h0510ff6_3 --> 14.1-h21ff451_3
vs2015_runtime: 14.0.25123-3 --> 15.5.2-3

The following packages will be DOWNGRADED:

rstudio: 1.1.383-vc14h5b9a52b_2 r --> 1.0.153-vc14 [vc14] --> 1.0.153

Proceed ([y]/n)? y

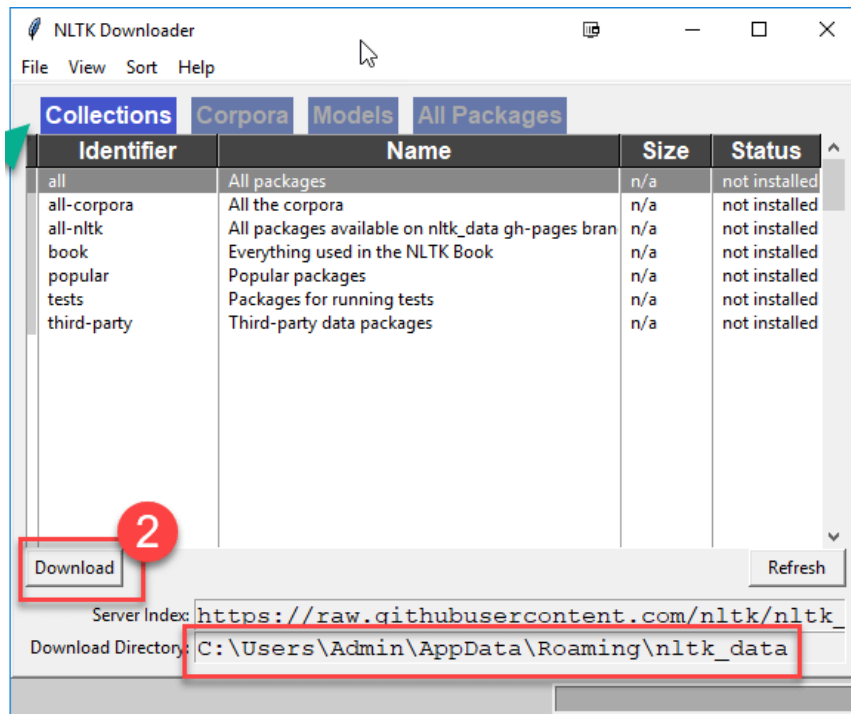
Downloading and Extracting Packages
vs2015_runtime 15.5.2: #####
vc 14.1: #####
rstudio 1.0.153: #####
nltk 3.3.0: #####
fonts-continuum 1: #####
certifi 2018.8.24: #####
Preparing transaction: done
Verifying transaction: done
Executing transaction: | DEBUG menuinst_win32: _init__(199): Menu
DEBUG menuinst_win32:create(323): Shortcut cmd is C:\Users\Admin\
Library\bin\rstudio.exe']
| DEBUG menuinst_win32: _init__(199): Menu: name: 'RStudio Deskt
DEBUG menuinst_win32:create(323): Shortcut cmd is C:\Users\Admin\
Library\bin\rstudio.exe']
done
```

## Étape 3 : NLTK.Downloader (chargement du dataset)

Une fois dans l'IDE Python réaliser les commandes suivantes

- `import nltk`
- `nltk.download()`

La fenêtre suivante va s'ouvrir :



Cliquer sur " All " et téléchargez-les tous !

Vous pouvez tester le bon déroulement de l'installation avec la commande :

```
>>> from nltk.corpus import brown
>>> brown.words()
```

Qui devrait faire apparaître à l'écran :

```
['The', 'Fulton', 'County', 'Grand', 'Jury', 'said', ...]
>>>
```

Vous-pouvez maintenant utiliser les applications de text-mining.