

td1

January 8, 2023

1 TD1 Get started

1.1 Partie 1 Installations

Pour ce TD, vous aurez besoin de votre machine.

En suivant les instructions sur les sites officiels, installer :

- Python3 : <https://python.org/>
- Jupyter notebook : <https://jupyter.org/>. En cas de besoin : https://python.sdv.univ-paris-diderot.fr/18_jupyter/
- NLTK <https://www.nltk.org/>
- Télécharger le corpus DinG et conserver le dans un répertoire dédié de ce cours : <https://gitlab.inria.fr/amblard/ding>

1.2 Partie 2 Premières manipulations

À partir d'un interpréteur python on souhaite manipuler le contenu du fichier `Ding2-1.txt` : <https://gitlab.inria.fr/semagramme-public-projects/resources/ding/>

Exo 1. Ouvrir le fichier et lire ligne par ligne. Pour réussir à lire dans un fichier, vous pouvez suivre les instructions ici : https://python.sdv.univ-paris-diderot.fr/07_fichiers/.

Un exemple, imaginer vous avez un fichier `zoo.txt` qui contient 4 lignes, avec chaque ligne le nom d'un animal

```
girafe
tigre
singe
souris
```

Si vous voulez l'ouvrir et afficher le contenu, vous pouvez utiliser la commande `open()`, `readline()` ou `readlines()`.

ATTENTION : Vous devez d'abord créer ce fichier pour que le bout de code suivant fonctionne.

```
[ ]: >>> filin = open("zoo.txt", "r")
>>> filin
# <_io.TextIOWrapper name='zoo.txt' mode='r' encoding='UTF-8'>
>>> filin.readlines()
# ['girafe\n', 'tigre\n', 'singe\n', 'souris\n']
```

```
>>> filin.close()
```

Oublier pas fermer le fichier ! Utiliser simplement la commande `close()`. Essayer le code suivant, il y a une erreur, pourquoi ?

```
[ ]: >>> filin.readlines()
```

Maintenant, essayer avec le fichier `Ding2-1.txt` téléchargé depuis le corpus `DinG`.

Exo 2. Nettoyage du texte et écriture dans un nouveau fichier.

Pour écrire dans un fichier, vous pouvez tester avec le mot clé `with` et `write`. Voici un exemple :

```
[ ]: >>> animaux2 = ["poisson", "abeille", "chat"]
>>> with open("zoo2.txt", "w") as filout:
...     for animal in animaux2:
...         filout.write(animal)
...
7
7
4
```

Maintenant, travailler sur le fichier `Ding2-1.txt`. Vous conservez uniquement les lignes contenant du texte. Enregistrer/Écrire votre résultat dans un nouveau fichier `Ding2-1-texte.txt`.

Déjà terminé ? Bravo ! Le texte semble-t-il assez propre ? Vous pouvez faire plus de nettoyage.