

But de la séance et démarche

Le but est de créer un **moteurs de recherche** dans un catalogue de textes. On dispose de plusieurs textes qui sont autant de variations d'une même histoire, mais rédigés avec des styles et des mots différents. Ce sont des extraits de l'ouvrage **Exercices de style**, l'un des plus célèbres de l'écrivain français *Raymond Queneau*.

Vous allez :

1. Lire les différents textes proposés, et relever ce qu'il vous semble être les mots les plus importants, pour chacun des textes (maximum 15). Ces mots peuvent être communs, ou différents des autres textes.
2. Donner une note sur 10 à chaque texte (qualité, ressenti).
3. Compléter ou corriger le schéma XML à l'aide de vos informations.
4. Exécuter la recherche du texte le plus approprié pour un mot donné.
5. Etablir un classement de ces textes à partir du mot recherché.

Travail

2.1 Lire et relever les mots clés

Vous trouverez les textes à l'adresse : allophysique.com/pdf/SNT/internet1_moteur_recherche.pdf

- Relever les mots clés pour chacun des textes
- Ajouter une note sur 10

texte	mots clés	note
Alexandrins	autobus, foutriquet, bouton, chapeau, cou, esclandre, haleine, personnage, place, putride, ruban, sottement, tresse, turban, moche, dandy, voyageur, zèbre, icy	7
—	-----	—
Vulgaire		
—	-----	—
Récit		
—	-----	—
Retrograde		
—	-----	—
Surprise		
—	-----	—
Présent		
—	-----	—
Passé simple		
—	-----	—
Précisions		
—	-----	—
Imparfait		
—	-----	—
Alors		

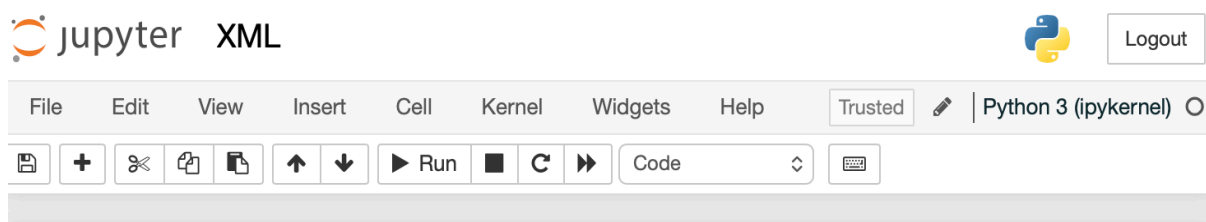
2.2 Notebook

Sur le site allophysique.com, cliquer sur le lien notebook > Capytale et ouvrir le *notebook*. Celui-ci va s'ouvrir dans un onglet de votre navigateur.

Un *notebook* est un éditeur-interpréteur en langage python, constitué de plusieurs cellules, l'une sous l'autre. Chaque cellule contient soit des instructions, soit un programme. Ces programmes construisent des *variables*, dont les valeurs sont conservées, même lorsque l'on *change de cellule*. Sauvegardez régulièrement votre travail lorsque vous progressez dans le notebook (bouton sauvegarde dans la barre des outils, ou CTRL+S)

Suivre les consignes du notebook. Au début de chaque cellule de code :

- Soit il est écrit EXECUTER. Vous devrez alors *Executer* la cellule, à l'aide du bouton > Run dans la barre des outils de la fenêtre. (ou Maj + Entrée), et passer à la suite.
- Soit il est écrit MODIFIER ou COMPLETER. Alors il est demandé d'écrire quelque chose dans la cellule du programme, puis de l'exécuter (> Run ou Maj + Entrée).



Analyse data XML pour SNT

<https://docs.python.org/fr/dev/library/xml.etree.elementtree.html>

In [3]:

```
1 ## EXECUTER LA CELLULE ##
2 # importer les données en lisant un fichier
3 import xml.etree.ElementTree as ET
4 tree = ET.parse('datas/country_data.xml')
5 root = tree.getroot()
```

instructions/
commentaires

Programme

Travail

1. Placer les mots clés que vous avez relevés pour chacun des textes. Attention à bien respecter le format XML
2. Modifier la note du texte si nécessaire
3. Executer la cellule

In [6]:

```
1 ## MODIFIER LA CELLULE ##
2 country_data_as_string = "<?xml version='1.0'?>
3 <data>
4   <country name='Liechtenstein'>
5     <rank>1</rank>
6     <year>2008</year>
7     <gdppc>141100</gdppc>
8     <neighbor name='Austria' direction='E'/>
9     <neighbor name='Switzerland' direction='W'/>
10  </country>
```

instructions/
commentaires

Programme

FIGURE 1 – exemple de notebook

2.3 Compléter le schéma XML

Le schéma XML organise vos informations en les disposant à l'aide de balises, à la manière du langage html. Les noms des balises peuvent être librement choisies, mais doivent suivre une certaine hiérarchie. C'est un format interchangeable (on peut les lire avec différents logiciels ou langages de programmation).

1. Pour compléter les différents mots-clés, placer ceux-ci l'un sous l'autre, en les séparant par des virgules, entre les balises `<motsCles>`, comme sur le schéma suivant :

```
1 <motsCles>autobus ,  
2   foutriquet ,  
3   bouton  
4 </motsCles>
```

Attention à bien placer les mots clés sur le bon texte, dont le nom est indiqué avec l'attribut de la balise *texte* (par exemple `<texte nom="Alexandrins">`).

2. Modifier éventuellement la note entre les balises `<note>`.
3. Exécuter la cellule.
4. Suivre les consignes du notebook et répondre aux questions ci-dessous.

Partie 3

Questions

1. Avec un mot aléatoire, représenter le diagramme classant les textes par pertinence. Indiquer le mot et le texte le plus pertinent.

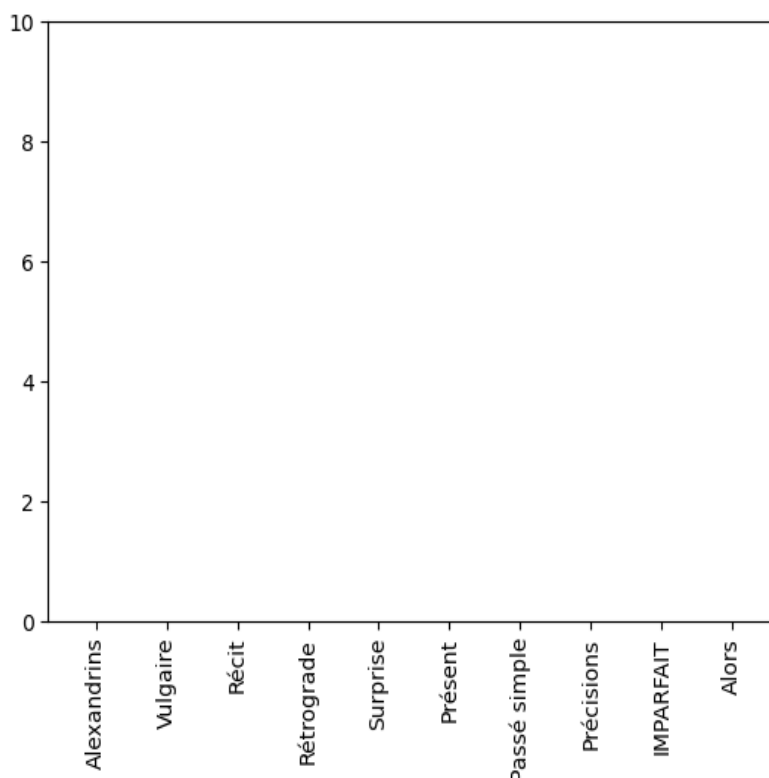


FIGURE 2 – diagramme pour le mot

2. Pour le mot *autobus*, obtenir le diagramme.

a) Comment interprétez vous ce diagramme ?

b) Quels sont les 2 textes les plus recommandés ?

c) Compter le score total obtenu pour le mot *autobus*, en ajoutant toutes les barres du diagramme.

3. Utilisez cette méthode pour rechercher quels sont les 3 mots les plus populaires, et donner leur score total.