

Ingénierie linguistique : TD 6-7

Chuyuan Li, Marie Cousin

Feb. 17 - Mars 03, 2023

1 Préambule

Aller sur <https://www.python.org/> et lancer la console interactive de Python en appuyant sur le bouton jaune marqué des signes `>_`. Le lancement prend quelques secondes.

Entrer dans la console les paroles de la chanson *Vive le vent*. Attention : pour rentrer un texte de plusieurs lignes, il faut l'encadrer de `"""`. Vous pouvez mettre ce bout de parole dans la console. Attention avec des majuscules ou minuscules.

```
str = """Vive le vent,  
vive le vent,  
Vive le vent d'Hiver,  
Qui s'en va,  
Sifflant Soufflant,  
dans les grands Sapins verts."""
```

Importer la librairie standard `re` pour pouvoir écrire des expressions régulières en Python. Pour cela, écrire `import re` dans la console interactive.

2 Expressions régulières en Python

2.1 `match`, `search`, `findall`

1. En utilisant la méthode `match`, écrire une expression régulière permettant de trouver automatiquement le premier vers de la chanson.
2. Trouver un mot commençant par une majuscule ne se situant pas au début d'un vers.
3. Trouver tous les commençant par une majuscule ne se situant pas au début d'un vers.
4. Trouver tous les mots contenant la lettre "v".
5. Trouver tous les mots qui riment avec "temps".

2.2 Groupes et divination

1. Compter le nombre de fois où les groupes de mots "le vent " et "le temps " apparaissent dans la chanson.
2. Le mot "vent" apparaît-il dans la chanson sans être précédé de "le " ?

2.3 Caractères spéciaux

1. Trouver tous les groupes de mots contenant une apostrophe.
2. Compter le nombre de lignes de la chanson.

2.4 Substitutions

1. Trouver tous les groupes de mots "le vent " et "le temps "apparaissent dans la chanson. Remplacer chacun de ces groupes par "le TAL".
2. Trouver tous les groupes de trois mots consécutifs dont le premier et le troisième commencent par la même lettre. Remplacer cette lettre par la première lettre du second mot.
3. Mettre au pluriel tous les noms commencent avec majuscule et finissent par la lettre "r, t, d".