



Regexp

recherche de motifs

Dans ce TD, vous allez apprendre à télécharger certains contenus et à utiliser des expressions régulières en Python après avoir révisé un peu les compréhensions.

Exercice 1. compréhensions

```
1 x = [a for a in [1, 2, 3]]
2
3 y = [a for a in x if a == 2]
4
5 z = [a for a in range(0, 21) if a % 2 == 0]
6
7 t = [a * 2 for a in range(0, 11)]
```

- Décrire les listes précédentes définies par compréhension
- Ecrire une fonction de profil *impair(n)*
- En déduire à l'aide d'une compréhension la liste des nombres impairs inférieurs à 100.

Exercice 2. expressions régulières

```
1 import re
2 phrase = 'le python est un serpent sympathique qui aime la
           salade'
3 re.match('(.*) est (.*) qui (.*)', phrase).groups()
```

- Que fait le code ci-dessus ?
- Comment extraire de la même phrase les groupements principaux à l'aide d'une expression régulière ?
- Essayez de même avec d'autres phrases de votre cru.

Exercice 3. Substitutions hasardeuses

```
1 adresse='157 SOUTH MAJOR ROAD'
2 adresse.replace('ROAD', 'RD')
```

1. Quelle peut-être la difficulté avec la substitution ci-dessus ?
2. Première réponse : utiliser \$ et \
3. encore mieux : les expressions régulières avec *chaines brutes* et *b' limites de mots*.

On peut préférer :

```
1 # il est plus robuste d'utiliser :
2 re.sub(r'\bROAD\b', 'RD.', addr2)
3 # ou r' indique une chaîne brute (raw text)
4 # et \b (boundary) une limite de mot
```

Exercice 4. Recherches itérées de motifs

```
1 import requests
2 url='http://www.gutenberg.org/cache/epub/16/pg16.txt'
3 # si on passe par le proxy univ orleans :
4 proxy={"http":"http://wwwcache.univ-orleans.fr:3128/"}
5 r=requests.get(url,proxies=proxy)
6 # si pas de proxy :
7 # r=requests.get(url)
8 PeterPan = r.text
9 len(r.text)
10 r.encoding
11 it=re.finditer('(The [^.]*)\.',PeterPan)
12 for i in it:
13     print(i.groups())
```

1. Essayez le code ci-dessus. Quel est son résultat ?
2. Y-a-t-il des phrases dans le texte en anglais de *Peter Pan* commençant par The. et finissant par day.
3. Cherchez des phrases dans le texte de *Peter Pan* qui contiennent le mot light au singulier et finissent par night ?