

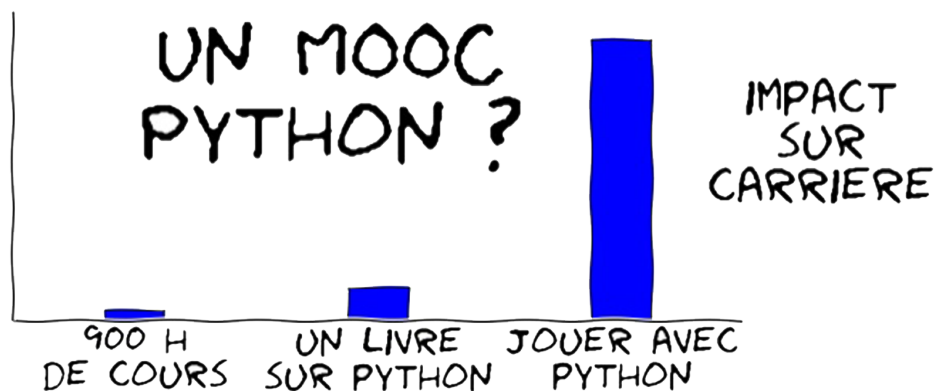


DES FONDAMENTAUX AU CONCEPTS AVANCÉS DU LANGAGE  
SESSION 2 - 17 SEPTEMBRE 2018

---

Thierry PARMENTELAT

Arnaud LEGOUT



<https://www.fun-mooc.fr>

Licence CC BY-NC-ND Thierry Parmentelat et Arnaud Legout



# Table des matières

<b>1</b>	<b>Introduction au MOOC et aux outils Python</b>	<b>7</b>
1.1	Versions de Python	8
1.2	Installer la distribution standard Python	10
1.2.1	Complément - niveau basique	10
1.2.2	Sachez à qui vous parlez	10
1.2.3	Digression - coexistence de Python2 et Python3	11
1.2.4	Installation de base	12
1.3	Un peu de lecture	14
1.3.1	Complément - niveau basique	14
1.3.2	Complément - niveau intermédiaire	15
1.4	“Notebooks” Jupyter comme support de cours	16
1.5	Modes d’exécution	19
1.6	La suite de Fibonacci	22
1.6.1	Complément - niveau basique	22
1.7	La suite de Fibonacci (suite)	24
1.7.1	Complément - niveau intermédiaire	24
1.8	La ligne <i>shebang</i>	26
1.8.1	Complément - niveau avancé	26
1.9	Dessiner un carré	28
1.9.1	Exercice - niveau intermédiaire	28
1.9.2	Exercice - niveau avancé	28
1.10	Noms de variables	29
1.10.1	Complément - niveau basique	29
1.11	Les mots-clés de Python	32
1.12	Un peu de calcul sur les types	34
1.12.1	Complément - niveau basique	34
1.12.2	Complément - niveau avancé	34
1.13	Gestion de la mémoire	36
1.13.1	Complément - niveau basique	36
1.13.2	Complément - niveau intermédiaire	36
1.14	Typages statique et dynamique	38
1.14.1	Complément - niveau intermédiaire	38
1.15	Utiliser Python comme une calculatrice	42
1.16	Affectations et Opérations (à la +=)	48
1.16.1	Complément - niveau intermédiaire	48
1.17	Notions sur la précision des calculs flottants	50
1.17.1	Complément - niveau avancé	50
1.18	Opérations <i>bit à bit</i> ( <i>bitwise</i> )	52

1.18.1	Compléments - niveau avancé	52
1.19	Estimer le plus petit (grand) flottant	54
1.19.1	Exercice - niveau basique	54
1.19.2	Complément - niveau avancé	56
<b>2</b>	<b>Notions de base, premier programme en Python</b>	<b>57</b>
2.1	Caractères accentués	58
2.1.1	Complément - niveau basique	58
2.1.2	Complément - niveau intermédiaire	59
2.2	Les outils de base sur les chaînes de caractères (str)	63
2.2.1	Complément - niveau intermédiaire	63
2.3	Formatage de chaînes de caractères	74
2.3.1	Complément - niveau basique	74
2.3.2	Complément - niveau intermédiaire	77
2.3.3	Complément - niveau avancé	77
2.4	Obtenir une réponse de l'utilisateur	79
2.4.1	Complément - niveau basique	79
2.5	Expressions régulières et le module re	80
2.5.1	Complément - niveau basique	80
2.5.2	Complément - niveau intermédiaire	81
2.6	Expressions régulières	97
2.6.1	Exercice - niveau intermédiaire (1)	97
2.6.2	Exercice - niveau intermédiaire (2)	97
2.6.3	Exercice - niveau intermédiaire (3)	98
2.6.4	Exercice - niveau avancé	98
2.7	Les slices en Python	101
2.7.1	Complément - niveau basique	101
2.7.2	Complément - niveau avancé	103
2.8	Méthodes spécifiques aux listes	105
2.8.1	Complément - niveau basique	105
2.9	Objets mutables et objets immuables	112
2.9.1	Complément - niveau basique	112
2.10	Tris de listes	113
2.10.1	Complément - niveau basique	113
2.11	Indentations en Python	116
2.11.1	Complément - niveau basique	116
2.11.2	Complément - niveau intermédiaire	117
2.11.3	Complément - niveau avancé	118
2.12	Bonnes pratiques de présentation de code	120
2.12.1	Complément - niveau basique	120
2.12.2	Complément - niveau intermédiaire	122
2.13	L'instruction pass	123
2.13.1	Complément - niveau basique	123
2.13.2	Complément - niveau intermédiaire	123
2.14	Fonctions avec ou sans valeur de retour	125
2.14.1	Complément - niveau basique	125
2.15	Formatage des chaînes de caractères	129
2.15.1	Exercice - niveau basique	129
2.16	Séquences	130
2.16.1	Exercice - niveau basique	130
2.16.2	Exercice - niveau intermédiaire	130

2.17	Listes	132
2.17.1	Exercice - niveau basique	132
2.18	Instruction if et fonction def	133
2.18.1	Exercice - niveau basique	133
2.18.2	Exercice - niveau basique	133
2.19	Comptage dans les chaînes	135
2.19.1	Exercice - niveau basique	135
2.19.2	La commande UNIX wc(1)	135
2.20	Compréhensions (1)	136
2.20.1	Exercice - niveau basique	136
2.20.2	Récréation	136
2.21	Compréhensions (2)	137
2.21.1	Exercice - niveau intermédiaire	137
<b>3</b>	<b>Renforcement des notions de base, références partagées</b>	<b>139</b>
3.1	Les fichiers	140
3.1.1	Complément - niveau basique	140
3.1.2	Complément - niveau intermédiaire	141
3.1.3	Complément - niveau avancé	141
3.2	Fichiers et utilitaires	145
3.2.1	Complément - niveau basique	145
3.2.2	Complément - niveau avancé	148
3.3	Formats de fichiers : JSON et autres	149
3.3.1	Compléments - niveau basique	149
3.3.2	Compléments - niveau intermédiaire	150
3.4	Fichiers systèmes	152
3.4.1	Complément - niveau avancé	152
3.5	La construction de tuples	154
3.5.1	Complément - niveau intermédiaire	154
3.6	Sequence unpacking	158
3.6.1	Complément - niveau basique	158
3.6.2	Complément - niveau intermédiaire	160
3.6.3	Pour en savoir plus	162
3.7	Plusieurs variables dans une boucle for	163
3.7.1	Complément - niveau basique	163
3.7.2	Complément - niveau intermédiaire	163
3.8	Fichiers	166
3.8.1	Exercice - niveau basique	166
3.9	Sequence unpacking	168
3.9.1	Exercice - niveau basique	168
3.10	Dictionnaires	169
3.10.1	Complément - niveau basique	169
3.10.2	Complément - niveau intermédiaire	172
3.10.3	Complément - niveau avancé	175
3.11	Clés immuables	178
3.11.1	Complément - niveau intermédiaire	178
3.12	Gérer des enregistrements	180
3.12.1	Complément - niveau intermédiaire	180
3.12.2	Complément - niveau avancé	182
3.13	Dictionnaires et listes	183
3.13.1	Exercice - niveau basique	183

3.14 Fusionner des données . . . . .	184
3.14.1 Exercices . . . . .	184
3.15 Ensembles . . . . .	188
3.15.1 Complément - niveau basique . . . . .	188
3.16 Ensembles . . . . .	193
3.16.1 Exercice - niveau basique . . . . .	193
3.16.2 Deuxième partie - niveau basique . . . . .	193
3.17 Exercice sur les ensembles . . . . .	194
3.17.1 Exercice - niveau intermédiaire . . . . .	194
3.18 try ... else ... finally . . . . .	196
3.18.1 Complément - niveau intermédiaire . . . . .	196
3.19 L'opérateur is . . . . .	199
3.19.1 Complément - niveau basique . . . . .	199
3.19.2 Complément - niveau intermédiaire . . . . .	200
3.20 Listes infinies & références circulaires . . . . .	203
3.20.1 Complément - niveau intermédiaire . . . . .	203
3.21 Les différentes copies . . . . .	206
3.21.1 Complément - niveau basique . . . . .	206
3.21.2 Complément - niveau intermédiaire . . . . .	207
3.22 L'instruction del . . . . .	210
3.22.1 Complément - niveau basique . . . . .	210
3.23 Affectation simultanée . . . . .	212
3.23.1 Complément - niveau basique . . . . .	212
3.24 Les instructions += et autres revisitées . . . . .	213
3.24.1 Complément - niveau intermédiaire . . . . .	213
3.25 Classe . . . . .	216
3.25.1 Exercice - niveau basique . . . . .	216

# Chapitre 1

## Introduction au MOOC et aux outils Python

## 1.1 Versions de Python

### Version de référence : Python-3.6

Comme on l'indique dans la vidéo, la version de Python qui a servi de **référence pour le MOOC est la version 3.6**, c'est notamment avec cette version que l'on a tourné les vidéos.

### Versions plus anciennes

Certaines précautions sont à prendre si vous utilisez une version plus ancienne :

**Python-3.5** Si vous préférez utiliser python-3.5, la différence la plus visible pour vous apparaîtra avec les *f-strings* :

```
In [1]: age = 10
```

```
# un exemple de f-string
f"Jean a {age} ans"
```

```
Out[1]: 'Jean a 10 ans'
```

Cette construction - que nous utilisons très fréquemment - n'a été introduite qu'en Python-3.6, aussi si vous utilisez Python-3.5 vous verrez ceci :

```
>>> age = 10
>>> f"Jean a {age} ans"
File "<stdin>", line 1
    f"Jean a {age} ans"
    ^
```

```
SyntaxError: invalid syntax
```

Dans ce cas vous devrez remplacer ce code avec la méthode `format` - que nous verrons en Semaine 2 avec les chaînes de caractères - et dans le cas présent il faudrait remplacer par ceci :

```
In [2]: age = 10
```

```
"Jean a {} ans".format(age)
```

```
Out[2]: 'Jean a 10 ans'
```

Comme ces *f-strings* sont très présents dans le cours, il est recommandé d'utiliser au moins python-3.6.

**Python-3.4** La remarque vaut donc *a fortiori* pour python-3.4 qui, en outre, ne vous permettra pas de suivre la semaine 8 sur la programmation asynchrone, car les mots-clés `async` et `await` ont été introduits seulement dans Python-3.5.

### Version utilisée dans les notebooks / versions plus récentes

Tout le cours doit pouvoir s'exécuter tel quel avec une version plus récente de Python.

Cela dit, certains compléments illustrent des nouveautés apparues après la 3.6, comme les *dataclasses* qui sont apparues avec python-3.7, et que nous verrons en semaine 6.



Dans tous les cas, nous **signalons systématiquement** les notebooks qui nécessitent une version plus récente que 3.6.

Voici enfin, à toutes fins utiles, un premier fragment de code Python qui affiche la version de Python utilisée dans tous les notebooks de ce cours.

Nous reviendrons en détail sur l'utilisation des notebooks dans une prochaine séquence, dans l'immédiat pour exécuter ce code vous pouvez :

- désigner avec la souris la cellule de code ; vous verrez alors apparaître une petite flèche à coté du mot In, en cliquant cette flèche vous exécutez le code ;
- un autre méthode consiste à sélectionner la cellule de code avec la souris ; une fois que c'est fait vous pouvez cliquer sur le bouton >| Run dans la barre de menu (bleue claire) du notebook.

```
In [3]: # ce premier fragment de code affiche des détails sur la
        # version de python qui exécute tous les notebooks du cours
import sys
print(sys.version_info)
```

```
sys.version_info(major=3, minor=6, micro=3, releaselevel='final', serial=0)
```

Pas de panique si vous n'y arrivez pas, nous consacrons très bientôt une séquence entière à l'utilisation des notebooks :)

## 1.2 Installer la distribution standard Python

### 1.2.1 Complément - niveau basique

Ce complément a pour but de vous donner quelques guides pour l'installation de la distribution standard Python 3.

Notez bien qu'il ne s'agit ici que d'indications, il existe de nombreuses façons de procéder.

En cas de souci, commencez par chercher par vous-même, sur google ou autre, une solution à votre problème; pensez également à utiliser le forum du cours.

Le point important est de **bien vérifier le numéro de version** de votre installation qui doit être **au moins 3.6**

### 1.2.2 Sachez à qui vous parlez

Mais avant qu'on n'avance sur l'installation proprement dite, il nous faut insister sur un point qui déroute parfois les débutants. On a parfois besoin de recourir à l'emploi d'un terminal, surtout justement pendant la phase d'installation.

Lorsque c'est le cas, il est important de bien distinguer :

- les cas où on s'adresse **au terminal** (en jargon, on dit le *shell*),
- et les cas où on s'adresse à **l'interpréteur Python**.

C'est très important car ces deux programmes ne parlent **pas** du tout le **même langage** ! Il peut arriver au début qu'on écrive une commande juste, mais au mauvais interlocuteur, et cela peut être source de frustration. Essayons de bien comprendre ce point.

#### Le terminal

Je peux dire que je parle à mon **terminal** quand l'invite de commande (en jargon on dit le *prompt*) **se termine par un dollar \$** - ou un simple chevron > sur Windows

Par exemple sur un mac :

```
~/git/flotpython/w1 $
```

Ou sur Windows :

```
C:\Users>
```

#### L'interprète Python

À partir du terminal, je peux lancer un **interpréteur Python**, qui se reconnaît car son prompt est fait de **3 chevrons >>>**

```
~/git/flotpython/w1 $ python3
Python 3.7.0 (default, Jun 29 2018, 20:14:27)
[Clang 9.0.0 (clang-900.0.39.2)] on darwin
Type "help", "copyright", "credits" or "license" for more information.
>>>
```

Pour sortir de l'interpréteur Python, et retourner au terminal, j'utilise la fonction Python `exit()` :

```
~/git/flotpython/w1 $ python3
>>> 20 * 60
1200
>>> exit()
~/git/flotpython/w1 $ python3
```

### Les erreurs typiques

Gardez bien cette distinction présente à l'esprit, lorsque vous lisez la suite. Voici quelques symptômes habituels de ce qu'on obtient si on se trompe.

Par exemple, la commande `python3 -V` est une commande qui s'adresse au terminal ; c'est pourquoi nous la faisons précéder d'un dollar `$`.

Si vous essayez de la taper alors que vous êtes déjà dans un interpréteur python - ou sous IDLE d'ailleurs -, vous obtenez un message d'erreur de ce genre :

```
>>> python3 -V
Traceback (most recent call last):
  File "<stdin>", line 1, in <module>
NameError: name 'python3' is not defined
```

Réciproquement, si vous essayez de taper du Python directement dans un terminal, ça se passe mal aussi, forcément. Par exemple sur Mac, avec des fragments Python tout simples :

```
~/git/flotpython/w1 $ import math
-bash: import: command not found
~/git/flotpython/w1 $ 30 * 60
-bash: 30: command not found
~/git/flotpython/w1 $ foo = 30 * 60
-bash: foo: command not found
```

### 1.2.3 Digression - coexistence de Python2 et Python3

Avant l'arrivée de la version 3 de Python, les choses étaient simples, on exécutait un programme Python avec une seule commande `python`. Depuis 2014-2015, maintenant que les deux versions de Python coexistent, il est nécessaire d'adopter une convention qui permette d'installer les deux langages sous des noms qui sont non-ambigus.

C'est pourquoi actuellement, on trouve **le plus souvent** la convention suivante sous Linux et macOS :

- `python3` est pour exécuter les programmes en Python-3; du coup on trouve alors également les commandes comme `idle3` pour lancer IDLE, et par exemple `pip3` pour le gestionnaire de paquets (voir ci-dessous);
- `python2` est pour exécuter les programmes en Python-2, avec typiquement `idle2` et `pip2`;
- enfin selon les systèmes, la commande `python` tout court est un alias pour `python2` ou `python3`. De plus en plus souvent, par défaut `python` désigne `python3`.

à titre d'illustration, voici ce que j'obtiens sur mon mac :

```
$ python3 -V
Python 3.6.2
$ python2 -V
Python 2.7.13
$ python -V
Python 3.6.2
```

Sous Windows, vous avez un lanceur qui s'appelle `py`. Par défaut, il lance la version de Python la plus récente installée, mais vous pouvez spécifier une version spécifique de la manière suivante :

```
C:\> py -2.7
```

pour lancer, par exemple, Python en version 2.7. Vous trouverez toute la documentation nécessaire pour Windows sur cette page (en anglais) : <https://docs.python.org/3/using/windows.html>

Pour éviter d'éventuelles confusions, nous précisons toujours `python3` dans le cours.

### 1.2.4 Installation de base

#### Vous utilisez Windows

La méthode recommandée sur Windows est de partir de la page <https://www.python.org/download> où vous trouverez un programme d'installation qui contient tout ce dont vous aurez besoin pour suivre le cours.

Pour vérifier que vous êtes prêts, il vous faut lancer IDLE (quelque part dans le menu Démarrer) et vérifier le numéro de version.

#### Vous utilisez macOS

Ici encore, la méthode recommandée est de partir de la page [www.python.org/download](http://www.python.org/download) et d'utiliser le programme d'installation.

Sachez aussi, si vous utilisez déjà MacPorts (<https://www.macports.org>), que vous pouvez également utiliser cet outil pour installer, par exemple Python 3.6, avec la commande

```
$ sudo port install python36
```

#### Vous utilisez Linux

Dans ce cas il est très probable que Python-3.x soit déjà disponible sur votre machine. Pour vous en assurer, essayez de lancer la commande `python3` dans un terminal.

##### RHEL / Fedora

Voici par exemple ce qu'on obtient depuis un terminal sur une machine installée en Fedora-20 :

```
$ python3
Python 3.6.2 (default, Jul 20 2017, 12:30:02)
[GCC 6.3.1 20161221 (Red Hat 6.3.1-1)] on linux
Type "help", "copyright", "credits" or "license" for more information.
>>> exit()
```

**Vérifiez bien le numéro de version** qui doit être en 3.x. Si vous obtenez un message du style `python3: command not found` utilisez `dnf` (anciennement connu sous le nom de `yum`) pour installer le rpm `python3` comme ceci :

```
$ sudo dnf install python3
```

S'agissant d'idle, l'éditeur que nous utilisons dans le cours (optionnel si vous êtes familier avec un éditeur de texte), vérifiez sa présence comme ceci :

```
$ type idle3
idle is hashed (/usr/bin/idle3)
```

Ici encore, si la commande n'est pas disponible vous pouvez l'installer avec :

```
$ sudo yum install python3-tools
```

### Debian / Ubuntu

Ici encore, Python-2.7 est sans doute déjà disponible. Procédez comme ci-dessus, voici un exemple recueilli dans un terminal sur une machine installée en Ubuntu-14.04/trusty :

```
$ python3
Python 3.6.2 (default, Jul 20 2017, 12:30:02)
[GCC 6.3.1 20161221 (Red Hat 6.3.1-1)] on linux
Type "help", "copyright", "credits" or "license" for more information.
>>> exit()
```

Pour installer Python :

```
$ sudo apt-get install python3
```

Pour installer idle :

```
$ sudo apt-get install idle3
```

### Installation de bibliothèques complémentaires

Il existe un outil très pratique pour installer des bibliothèques Python, il s'appelle pip3, qui est documenté ici <https://pypi.python.org/pypi/pip>

Sachez aussi, si par ailleurs vous utilisez un gestionnaire de paquets comme rpm sur RHEL, apt-get sur Debian, ou port sur macOS, que de nombreux paquets sont également disponibles au travers de ces outils.

### Anaconda

Sachez qu'il existe beaucoup de distributions alternatives qui incluent Python; parmi elles, la plus populaire est sans aucun doute [Anaconda](#), qui contient un grand nombre de bibliothèques de calcul scientifique, et également d'ailleurs jupyter pour travailler nativement sur des notebooks au format .ipynb.

Anaconda vient avec son propre gestionnaires de paquets pour l'installation de bibliothèques supplémentaires qui s'appelle conda.

## 1.3 Un peu de lecture

### 1.3.1 Complément - niveau basique

#### Mise à jour de Juillet 2018

Le 12 Juillet 2018, Guido van Rossum [a annoncé qu'il quittait la fonction de BDFL](#) qu'il occupait depuis près de trois décennies. Il n'est pas tout à fait clair à ce stade comment va évoluer la gouvernance de Python.

#### Le Zen de Python

Vous pouvez lire le "Zen de Python", qui résume la philosophie du langage, en important le module `this` avec ce code : (pour exécuter ce code, cliquez dans la cellule de code, et faites au clavier "Majuscule/Entrée" ou "Shift/Enter")

```
In [1]: # le Zen de Python
import this
```

The Zen of Python, by Tim Peters

```
Beautiful is better than ugly.
Explicit is better than implicit.
Simple is better than complex.
Complex is better than complicated.
Flat is better than nested.
Sparse is better than dense.
Readability counts.
Special cases aren't special enough to break the rules.
Although practicality beats purity.
Errors should never pass silently.
Unless explicitly silenced.
In the face of ambiguity, refuse the temptation to guess.
There should be one-- and preferably only one --obvious way to do it.
Although that way may not be obvious at first unless you're Dutch.
Now is better than never.
Although never is often better than *right* now.
If the implementation is hard to explain, it's a bad idea.
If the implementation is easy to explain, it may be a good idea.
Namespaces are one honking great idea -- let's do more of those!
```

#### Documentation

- On peut commencer par citer l'[article de Wikipédia sur Python en français](#).
- La [page sur le langage en français](#).
- La [documentation originale](#) de Python 3 - donc, en anglais - est un très bon point d'entrée lorsqu'on cherche un sujet particulier, mais (beaucoup) trop abondante pour être lue d'un seul trait. Pour chercher de la documentation sur un module particulier, le plus simple est encore d'utiliser Google - ou votre moteur de recherche favori - qui vous redirigera, dans la grande majorité des cas, vers la page qui va bien dans, précisément, la documentation de Python.

À titre d'exercice, cherchez la documentation du module `pathlib` en [cherchant sur Google](#) les mots-clé "python module pathlib".

- J'aimerais vous signaler également une initiative pour [traduire la documentation officielle en français](#).

### Historique et survol

- La FAQ officielle de Python (en anglais) sur [les choix de conception et l'historique du langage](#).
- L'article de Wikipédia (en anglais) sur [l'historique du langage](#).
- Sur Wikipédia, un article (en anglais) sur [la syntaxe et la sémantique de Python](#).

### Un peu de folklore

- Le [discours de Guido van Rossum à PyCon 2016](#).
- Sur YouTube, le [sketch des Monty Python](#) d'où proviennent les termes spam, eggs et autres beans que l'on utilise traditionnellement dans les exemples en Python plutôt que `foo` et `bar`.
- L'[article Wikipédia correspondant](#), qui cite le langage Python.

## 1.3.2 Complément - niveau intermédiaire

### Licence

- La [licence d'utilisation](#) est disponible [ici](#).
- La page de la [Python Software Foundation](#), qui est une entité légale similaire à nos associations de 1901, à but non lucratif; elle possède les droits sur le langage.

### Le processus de développement

- Comment les choix d'évolution sont proposés et discutés, au travers des PEP (Python Enhancement Proposals)
  - [http://en.wikipedia.org/wiki/Python\\_\(programming\\_language\)#Development](http://en.wikipedia.org/wiki/Python_(programming_language)#Development)
- Le premier PEP décrit en détail le cycle de vie des PEPs
  - <http://legacy.python.org/dev/peps/pep-0001/>
- Le PEP 8, qui préconise un style de présentation (*style guide*)
  - <http://legacy.python.org/dev/peps/pep-0008/>
- L'index de tous les PEPs
  - <http://legacy.python.org/dev/peps/>

## 1.4 “Notebooks” Jupyter comme support de cours

Pour illustrer les vidéos du MOOC, nous avons choisi d'utiliser Jupyter pour vous rédiger les documents “mixtes” contenant du texte et du code Python, qu'on appelle des “notebooks”, et dont le présent document est un exemple.

Nous allons, dans la suite, utiliser du code Python, pourtant nous n'avons pas encore abordé le langage. Pas d'inquiétude, ce code est uniquement destiné à valider le fonctionnement des notebooks, et nous n'utilisons que des choses très simples.

### Avertissement : réglages du navigateur

Avant toute chose, pour un bon fonctionnement des notebooks, on rappelle qu'il est nécessaire d'avoir **autorisé** dans votre navigateur les **cookies** en provenance du site Internet `nbhosting.inria.fr`, qui héberge l'infrastructure qui héberge tous les notebooks.

### Avantages des notebooks

Comme vous le voyez, ce support permet un format plus lisible que des commentaires dans un fichier de code.

Nous attirons votre attention sur le fait que **les fragments de code peuvent être évalués et modifiés**. Ainsi vous pouvez facilement essayer des variantes autour du notebook original.

Notez bien également que le code Python est interprété **sur une machine distante**, ce qui vous permet de faire vos premiers pas avant même d'avoir procédé à l'installation de Python sur votre propre ordinateur.

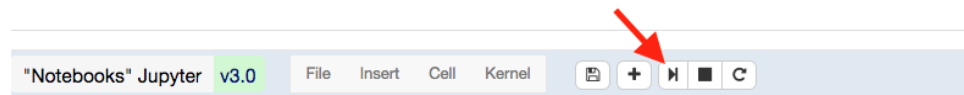
### Comment utiliser les notebooks

En haut du notebook, vous avez une barre de menu (sur fond bleu clair), contenant :

- un titre pour le notebook, avec un numéro de version ;
- une barre de menus avec les entrées File, Insert, Cell, Kernel ;
- et une barre de boutons qui sont des raccourcis vers certains menus fréquemment utilisés. Si vous laissez votre souris au dessus d'un bouton, un petit texte apparaît, indiquant à quelle fonction correspond ce bouton.

Nous avons vu dans la vidéo qu'un notebook est constitué d'une suite de cellules, soit textuelles, soit contenant du code. Les cellules de code sont facilement reconnaissables, elles sont précédées de `In [ ] :`. La cellule qui suit celle que vous êtes en train de lire est une cellule de code.

Pour commencer, sélectionnez cette cellule de code avec votre souris, et appuyez dans la barre de menu - en haut du notebook, donc - sur celui en forme de flèche triangulaire vers la droite (Play) :



```
In [1]: 20 * 30
```

```
Out[1]: 600
```

Comme vous le voyez, la cellule est “exécutée” (on dira plus volontiers évaluée), et on passe à la cellule suivante.



Alternativement, vous pouvez simplement taper au clavier *Shift+Enter*, ou selon les claviers *Maj-Entrée*, pour obtenir le même effet. D'une manière générale, il est important d'apprendre et d'utiliser les raccourcis clavier, cela vous fera gagner beaucoup de temps par la suite.

La façon habituelle d'exécuter l'ensemble du notebook consiste :

- à sélectionner la première cellule,
- et à taper *Shift+Enter* jusqu'à attendre la fin du notebook.

Lorsqu'une cellule de code a été évaluée, Jupyter ajoute sous la cellule In une cellule Out qui donne le résultat du fragment Python, soit ci-dessus 600.

Jupyter ajoute également un nombre entre les crochets pour afficher, par exemple ci-dessus, In [1] :. Ce nombre vous permet de retrouver l'ordre dans lequel les cellules ont été évaluées.

Vous pouvez naturellement modifier ces cellules de code pour faire des essais; ainsi vous pouvez vous servir du modèle ci-dessous pour calculer la racine carrée de 3, ou essayer la fonction sur un nombre négatif et voir comment est signalée l'erreur.

```
In [2]: # math.sqrt (pour square root) calcule la racine carrée
import math
math.sqrt(2)
```

```
Out[2]: 1.4142135623730951
```

On peut également évaluer tout le notebook en une seule fois en utilisant le menu *Cell -> Run All*.

### Attention à bien évaluer les cellules dans l'ordre

Il est important que les cellules de code soient évaluées dans le bon ordre. Si vous ne respectez pas l'ordre dans lequel les cellules de code sont présentées, le résultat peut être inattendu.

En fait, évaluer un programme sous forme de notebook revient à le découper en petits fragments, et si on exécute ces fragments dans le désordre, on obtient naturellement un programme différent.

On le voit sur cet exemple :

```
In [3]: message = "Il faut faire attention à l'ordre dans lequel on évalue les notebooks"
```

```
In [4]: print(message)
```

```
Il faut faire attention à l'ordre dans lequel on évalue les notebooks
```

Si un peu plus loin dans le notebook on fait par exemple :

```
In [5]: # ceci a pour effet d'effacer la variable 'message'
del message
```

qui rend le symbole `message` indéfini, alors bien sûr on ne peut plus évaluer la cellule qui fait `print` puisque la variable `message` n'est plus connue de l'interpréteur.

### Réinitialiser l'interpréteur

Si vous faites trop de modifications, ou perdez le fil de ce que vous avez évalué, il peut être utile de redémarrer votre interpréteur. Le menu *Kernel* → *Restart* vous permet de faire cela, un peu à la manière de IDLE qui repart d'un interpréteur vierge lorsque vous utilisez la fonction F5.

Le menu *Kernel* → *Interrupt* peut être quant à lui utilisé si votre fragment prend trop longtemps à s'exécuter (par exemple vous avez écrit une boucle dont la logique est cassée et qui ne termine pas).

### Vous travaillez sur une copie

Un des avantages principaux des notebooks est de vous permettre de modifier le code que nous avons écrit, et de voir par vous-mêmes comment se comporte le code modifié.

Pour cette raison, chaque élève dispose de sa **propre copie** de chaque notebook, vous pouvez bien sûr apporter toutes les modifications que vous souhaitez à vos notebooks sans affecter les autres étudiants.

### Revenir à la version du cours

Vous pouvez toujours revenir à la version “du cours” grâce au menu *File* → *Reset to original*.

Attention, avec cette fonction vous restaurez **tout le notebook** et donc **vous perdez vos modifications sur ce notebook**.

### Télécharger au format Python

Vous pouvez télécharger un notebook au format Python sur votre ordinateur grâce au menu *File* → *Download as* → *Python*

Les cellules de texte sont préservées dans le résultat sous forme de commentaires Python.

### Partager un notebook en lecture seule

Enfin, avec le menu *File* → *Share static version*, vous pouvez publier une version en lecture seule de votre notebook; vous obtenez une URL que vous pouvez publier, par exemple pour demander de l'aide sur le forum. Ainsi, les autres étudiants peuvent accéder en lecture seule à votre code.

Notez que lorsque vous utilisez cette fonction plusieurs fois, c'est toujours la dernière version publiée que verront vos camarades, l'URL utilisée reste toujours la même pour un étudiant et un notebook donné.

### Ajouter des cellules

Vous pouvez ajouter une cellule n'importe où dans le document avec le bouton + de la barre de boutons.

Aussi, lorsque vous arrivez à la fin du document, une nouvelle cellule est créée chaque fois que vous évaluez la dernière cellule; de cette façon vous disposez d'un brouillon pour vos propres essais.

À vous de jouer.

## 1.5 Modes d'exécution

Nous avons donc à notre disposition plusieurs façons d'exécuter un programme Python. Nous allons les étudier plus en détail :

Quoi	Avec quel outil
fichier complet	python3 <fichier>.py
ligne à ligne	python3 en mode interactif ou sous ipython3 ou avec IDLE
par fragments	dans un notebook

Pour cela nous allons voir le comportement d'un tout petit programme Python lorsqu'on l'exécute sous ces différents environnements.

On veut surtout expliquer une petite différence quant au niveau de détail de ce qui se trouve imprimé.

Essentiellement, lorsqu'on utilise l'interpréteur en mode interactif - ou sous IDLE - à chaque fois que l'on tape une ligne, le résultat est **calculé** (on dit aussi **évalué**) puis **imprimé**.

Par contre, lorsqu'on écrit tout un programme, on ne peut plus imprimer le résultat de toutes les lignes, cela produirait un flot d'impression beaucoup trop important. Par conséquent, si vous ne déclenchez pas une impression avec, par exemple, la fonction `print`, rien ne s'affichera.

Enfin, en ce qui concerne le notebook, le comportement est un peu hybride entre les deux, en ce sens que seul le **dernier résultat** de la cellule est imprimé.

### L'interpréteur Python interactif

Le programme choisi est très simple, c'est le suivant :

```
10 * 10
20 * 20
30 * 30
```

Voici comment se comporte l'interpréteur interactif quand on lui soumet ces instructions :

```
$ python3
Python 3.5.1 (v3.5.1:37a07cee5969, Dec 5 2015, 21:12:44)
[GCC 4.2.1 (Apple Inc. build 5666) (dot 3)] on darwin
Type "help", "copyright", "credits" or "license" for more information.
>>> 10 * 10
100
>>> 20 * 20
400
>>> 30 * 30
900
>>> exit()
$
```

Notez que pour terminer la session, il nous faut “sortir” de l’interpréteur en tapant `exit()`.

On peut aussi taper `Control-D` sous Linux ou macOS.

Comme on le voit ici, l’interpréteur imprime **le résultat de chaque ligne**. On voit bien apparaître toutes les valeurs calculées, 100, 400, puis enfin 900.

### Sous forme de programme constitué

Voyons à présent ce que donne cette même séquence de calculs dans un programme complet. Pour cela, il nous faut tout d’abord fabriquer un fichier avec un suffixe en `.py`, en utilisant par exemple un éditeur de fichier. Le résultat doit ressembler à ceci :

```
$ cat foo.py
10 * 10
20 * 20
30 * 30
$
```

Exécutons à présent ce programme :

```
$ python3 foo.py
$
```

On constate donc que ce programme **ne fait rien** ! En tout cas, selon toute apparence.

En réalité, les 3 valeurs 100, 400 et 900 sont bien calculées, mais comme aucune instruction `print` n’est présente, rien n’est imprimé et le programme se termine sans signe apparent d’avoir réellement fonctionné.

Ce comportement peut paraître un peu déroutant au début, mais comme nous l’avons mentionné c’est tout à fait délibéré. Un programme fonctionnel faisant facilement plusieurs milliers de lignes, voire beaucoup plus, il ne serait pas du tout réaliste que chaque ligne produise une impression, comme c’est le cas en mode interactif.

### Dans un notebook

Voici à présent le même programme dans un notebook :

```
In [1]: 10 * 10
        20 * 20
        30 * 30

Out[1]: 900
```

Lorsqu’on exécute cette cellule (rappel : sélectionner la cellule, et utiliser le bouton en forme de flèche vers la droite, ou entrer “**Shift+Enter**” au clavier), on obtient une seule valeur dans la rubrique `Out []`, 900, qui correspond **au résultat de la dernière ligne**.

### Utiliser print

Ainsi, pour afficher un résultat intermédiaire, on utilise l'instruction `print`. Nous verrons cette instruction en détail dans les semaines qui viennent, mais en guise d'introduction disons seulement que c'est une fonction comme les autres en Python 3.

```
In [2]: a = 10
        b = 20

        print(a, b)
```

10 20

On peut naturellement mélanger des objets de plusieurs types, et donc mélanger des chaînes de caractères et des nombres pour obtenir un résultat un peu plus lisible. En effet, lorsque le programme devient gros, il est important de savoir à quoi correspond une ligne dans le flot de toutes les impressions. Aussi on préférera quelque chose comme :

```
In [3]: print("a =", a, "et b =", b)
```

a = 10 et b = 20

```
In [4]: # ou encore, équivalente mais avec un f-string
        print(f"a = {a} et b = {b}")
```

a = 10 et b = 20

Une pratique courante consiste d'ailleurs à utiliser les commentaires pour laisser dans le code les instructions `print` qui correspondent à du debug (c'est-à-dire qui ont pu être utiles lors de la mise au point et qu'on veut pouvoir réactiver rapidement).

### Utiliser print pour "sous-titrer" une affectation

Remarquons enfin que l'affectation à une variable ne retourne aucun résultat. C'est-à-dire, en pratique, que si on écrit :

```
In [5]: a = 100
```

même une fois l'expression évaluée par l'interpréteur, aucune ligne `Out []` n'est ajoutée.

C'est pourquoi, il nous arrivera parfois d'écrire, notamment lorsque l'expression est complexe et pour rendre explicite la valeur qui vient d'être affectée :

```
In [6]: a = 100; print(a)
```

100

Notez bien que cette technique est uniquement pédagogique, et n'a absolument aucun autre intérêt dans la pratique; il n'est **pas recommandé** de l'utiliser en dehors de ce contexte.

## 1.6 La suite de Fibonacci

### 1.6.1 Complément - niveau basique

Voici un premier exemple de code qui tourne.

Nous allons commencer par le faire tourner dans ce notebook. Nous verrons en fin de séance comment le faire fonctionner localement sur votre ordinateur.

Le but de ce programme est de calculer la [suite de Fibonacci](#), qui est définie comme ceci :

- $u_0 = 1$
- $u_1 = 1$
- $\forall n \geq 2, u_n = u_{n-1} + u_{n-2}$

Ce qui donne pour les premières valeurs :

n	fibonacci(n)
0	1
1	1
2	2
3	3
4	5
5	8
6	13

On commence par définir la fonction `fibonacci` comme il suit. Naturellement vous n'avez pas encore tout le bagage pour lire ce code, ne vous inquiétez pas, nous allons vous expliquer tout ça dans les prochaines semaines. Le but est uniquement de vous montrer un fonctionnement de l'interpréteur Python et de IDLE.

```
In [1]: def fibonacci(n):
        "retourne le nombre de fibonacci pour l'entier n"
        # pour les petites valeurs de n il n'y a rien à calculer
        if n <= 1:
            return 1
        # sinon on initialise f1 pour n-1 et f2 pour n-2
        f2, f1 = 1, 1
        # et on itère n-1 fois pour additionner
        for i in range(2, n + 1):
            f2, f1 = f1, f1 + f2
        # print(i, f2, f1)
        # le résultat est dans f1
        return f1
```

Pour en faire un programme utilisable on va demander à l'utilisateur de rentrer un nombre ; il faut le convertir en entier car `input` renvoie une chaîne de caractères :

```
In [2]: entier = int(input("Entrer un entier "))
```

Entrer un entier 42

On imprime le résultat :

```
In [3]: print(f"fibonacci({entier}) = {fibonacci(entier)}")  
  
fibonacci(42) = 433494437
```

### Exercice

Vous pouvez donc à présent :

- exécuter le code dans ce notebook
- télécharger ce code sur votre disque comme un fichier `fibonacci_prompt.py`
  - utiliser pour cela le menu *"File -> Download as -> Python"*
  - et **renommer le fichier obtenu** au besoin
- l'exécuter sous IDLE
- le modifier, par exemple pour afficher les résultats intermédiaires
  - on a laissé exprès une fonction `print` en commentaire que vous pouvez réactiver simplement
- l'exécuter avec l'interpréteur Python comme ceci :  
\$ `python3 fibonacci_prompt.py`

Ce code est volontairement simple et peu robuste pour ne pas l'alourdir. Par exemple, ce programme se comporte mal si vous entrez un entier négatif.

Nous allons voir tout de suite une version légèrement différente qui va vous permettre de donner la valeur d'entrée sur la ligne de commande.

## 1.7 La suite de Fibonacci (suite)

### 1.7.1 Complément - niveau intermédiaire

Nous reprenons le cas de la fonction `fibonacci` que nous avons déjà vue, mais cette fois nous voulons que l'utilisateur puisse indiquer l'entier en entrée de l'algorithme, non plus en répondant à une question, mais sur la ligne de commande, c'est-à-dire en tapant :

```
$ python3 fibonacci.py 12
```

#### Avertissement :

Attention, cette version-ci **ne fonctionne pas dans ce notebook**, justement car on n'a pas de moyen dans un notebook d'invoquer un programme en lui passant des arguments de cette façon. Ce notebook est rédigé pour vous permettre de vous entraîner avec la fonction de téléchargement au format Python, qu'on a vue dans la vidéo, et de faire tourner ce programme sur votre propre ordinateur.

#### Le module `argparse`

Cette fois nous importons le module `argparse`, c'est lui qui va nous permettre d'interpréter les arguments passés sur la ligne de commande.

```
In [ ]: from argparse import ArgumentParser
```

Puis nous répétons la fonction `fibonacci` :

```
In [ ]: def fibonacci(n):
    "retourne le nombre de fibonacci pour l'entier n"
    # pour les petites valeurs de n il n'y a rien à calculer
    if n <= 1:
        return 1
    # sinon on initialise f1 pour n-1 et f2 pour n-2
    f2, f1 = 1, 1
    # et on itère n-1 fois pour additionner
    for i in range(2, n + 1):
        f2, f1 = f1, f1 + f2
    # print(i, f2, f1)
    # le résultat est dans f1
    return f1
```

#### Remarque :

Certains d'entre vous auront évidemment remarqué que l'on aurait pu éviter de copier-coller la fonction `fibonacci` comme cela ; c'est à ça que servent les modules, mais nous n'en sommes pas là.

#### Un objet parser

À présent, nous utilisons le module `argparse`, pour lui dire qu'on attend exactement un argument sur la ligne de commande, et qu'il doit être un entier. Ici encore, ne vous inquiétez pas si vous ne comprenez pas tout le code. L'objectif est de vous donner un morceau de code utilisable tout de suite, pour jouer avec votre interpréteur Python.



```
In [ ]: # à nouveau : ceci n'est pas conçu pour être exécuté dans le notebook !
        parser = ArgumentParser()
        parser.add_argument(dest="entier", type=int,
                             help="entier d'entrée")
        input_args = parser.parse_args()
        entier = input_args.entier
```

Nous pouvons à présent afficher le résultat :

```
In [ ]: print(f"fibonacci({entier}) = {fibonacci(entier)}")
```

Vous pouvez donc à présent :

- télécharger ce code sur votre disque comme un fichier `fibonacci.py` en utilisant le menu *"File -> Download as -> Python"*
- l'exécuter avec simplement l'interpréteur Python comme ceci :  
\$ python3 fibonacci.py 56

## 1.8 La ligne *shebang*

```
#!/usr/bin/env python3
```

### 1.8.1 Complément - niveau avancé

Ce complément est uniquement valable pour macOS et Linux.

#### Le besoin

Nous avons vu dans la vidéo que, pour lancer un programme Python, on fait depuis le terminal :

```
$ python3 mon_module.py
```

Lorsqu’il s’agit d’un programme que l’on utilise fréquemment, on n’est pas forcément dans le répertoire où se trouve le programme Python. Aussi, dans ce cas, on peut utiliser un chemin “absolu”, c’est-à-dire à partir de la racine des noms de fichiers, comme par exemple :

```
$ python3 /le/chemin/jusqu/a/mon_module.py
```

Sauf que c’est assez malcommode, et cela devient vite pénible à la longue.

#### La solution

Sur Linux et macOS, il existe une astuce utile pour simplifier cela. Voyons comment s’y prendre, avec par exemple le programme `fibonacci.py` que vous pouvez [télécharger ici](#) (nous avons vu ce code en détail dans les deux compléments précédents). Commencez par sauvegarder ce code sur votre ordinateur dans un fichier qui s’appelle, bien entendu, `fibonacci.py`.

On commence par éditer le tout début du fichier pour lui ajouter une **première ligne** :

```
#!/usr/bin/env python3

## La suite de Fibonacci (Suite)
...etc...
```

Cette première ligne s’appelle un **Shebang** dans le jargon Unix. Unix stipule que le Shebang doit être en **première position** dans le fichier.

Ensuite on rajoute au fichier, depuis le terminal, le caractère exécutable comme ceci :

```
$ pwd
/le/chemin/jusqu/a/

$ chmod +x fibonacci.py
```

À partir de là, vous pouvez utiliser le fichier `fibonacci.py` comme une commande, sans avoir à mentionner `python3`, qui sera invoqué au travers du shebang :

```
$ /le/chemin/jusqu/a/fibonacci.py 20
fibonacci(20) = 10946
```

Et donc vous pouvez aussi le déplacer dans un répertoire qui est dans votre variable `PATH`; de cette façon vous les rendez ainsi accessible à partir n'importe quel répertoire en faisant simplement :

```
$ export PATH=/le/chemin/jusqu/a:$PATH
```

```
$ cd /tmp
```

```
$ fibonacci.py 20
```

```
fibonacci(20) = 10946
```

**Remarque :** tout ceci fonctionne très bien tant que votre point d'entrée - ici `fibonacci.py` - n'utilise que des modules standards. Dans le cas où le point d'entrée vient avec au moins un module, il est également nécessaire d'installer ces modules quelque part, et d'indiquer au point d'entrée comment les trouver, nous y reviendrons dans la semaine où nous parlerons des modules.

## 1.9 Dessiner un carré

### 1.9.1 Exercice - niveau intermédiaire

Voici un tout petit programme qui dessine un carré.

Il utilise le module `turtle`, conçu précisément à des fins pédagogiques. Pour des raisons techniques, le module `turtle` n'est **pas disponible** au travers de la plateforme FUN.

**Il est donc inutile d'essayer d'exécuter ce programme depuis le notebook.** L'objectif de cet exercice est plutôt de vous entraîner à télécharger ce programme en utilisant le menu "File -> Download as -> Python", puis à le charger dans votre IDLE pour l'exécuter sur votre machine.

**Attention** également à sauver le programme téléchargé **sous un autre nom** que `turtle.py`, car sinon vous allez empêcher python de trouver le module standard `turtle`; appelez-le par exemple `turtle_basic.py`.

```
In [ ]: # on a besoin du module turtle
import turtle
```

On commence par définir une fonction qui dessine un carré de côté `length` :

```
In [ ]: def square(length):
        "have the turtle draw a square of side <length>"
        for side in range(4):
            turtle.forward(length)
            turtle.left(90)
```

Maintenant on commence par initialiser la tortue :

```
In [ ]: turtle.reset()
```

On peut alors dessiner notre carré :

```
In [ ]: square(200)
```

Et pour finir on attend que l'utilisateur clique dans la fenêtre de la tortue, et alors on termine :

```
In [ ]: turtle.exitonclick()
```

### 1.9.2 Exercice - niveau avancé

Naturellement vous pouvez vous amuser à modifier ce code pour dessiner des choses un peu plus amusantes.

Dans ce cas, commencez par chercher "*module python turtle*" dans votre moteur de recherche favori, pour localiser la documentation du module `turtle`.

Vous trouverez quelques exemples pour commencer ici :

- [turtle\\_multi\\_squares.py](#) pour dessiner des carrés à l'emplacement de la souris en utilisant plusieurs tortues;
- [turtle\\_fractal.py](#) pour dessiner une fractale simple;
- [turtle\\_fractal\\_reglable.py](#) une variation sur la fractale, plus paramétrable.

## 1.10 Noms de variables

### 1.10.1 Complément - niveau basique

Revenons sur les noms de variables autorisés ou non.

Les noms les plus simples sont constitués de lettres. Par exemple :

```
In [1]: factoriel = 1
```

On peut utiliser aussi les majuscules, mais attention cela définit une variable différente. Ainsi :

```
In [2]: Factoriel = 100
        factoriel == Factoriel
```

```
Out[2]: False
```

Le signe == permet de tester si deux variables ont la même valeur. Si les variables ont la même valeur, le test retournera True, et False sinon. On y reviendra bien entendu.

#### Conventions habituelles

En règle générale, on utilise **uniquement des minuscules** pour désigner les variables simples (ainsi d'ailleurs que pour les noms de fonctions), les majuscules sont réservées en principe pour d'autres sortes de variables, comme les noms de classe, que nous verrons ultérieurement.

Notons qu'il s'agit uniquement d'une convention, ceci n'est pas imposé par le langage lui-même.

Pour des raisons de lisibilité, il est également possible d'utiliser le tiret bas \_ dans les noms de variables. On préférera ainsi :

```
In [3]: age_moyen = 75 # oui
```

plutôt que ceci (bien qu'autorisé par le langage) :

```
In [4]: AgeMoyen = 75 # autorisé, mais non
```

On peut également utiliser des chiffres dans les noms de variables comme par exemple :

```
In [5]: age_moyen_dept75 = 80
```

avec la restriction toutefois que le premier caractère ne peut pas être un chiffre, cette affectation est donc refusée :

```
In [6]: 75_age_moyen = 80 # erreur de syntaxe
```

```
File "<ipython-input-6-823fed77034a>", line 1
75_age_moyen = 80 # erreur de syntaxe
^
```

```
SyntaxError: invalid token
```

### Le tiret bas comme premier caractère

Il est par contre, possible de faire commencer un nom de variable par un tiret bas comme premier caractère; toutefois, à ce stade, nous vous déconseillons d'utiliser cette pratique qui est réservée à des conventions de nommage bien spécifiques.

```
In [7]: _autorise_mais_deconseille = 'Voir le PEP 008'
```

Et en tout cas, il est **fortement déconseillé** d'utiliser des noms de la forme `__variable__` qui sont réservés au langage. Nous reviendrons sur ce point dans le futur, mais regardez par exemple cette variable que nous n'avons définie nulle part mais qui pourtant existe bel et bien :

```
In [8]: __name__ # ne définissez pas vous-même de variables de ce genre
```

```
Out[8]: '__main__'
```

### Ponctuation

Dans la plage des caractères ASCII, il n'est **pas possible** d'utiliser d'autres caractères que les caractères alphanumériques et le tiret bas. Notamment le tiret haut - est interprété comme l'opération de soustraction. Attention donc à cette erreur fréquente :

```
In [9]: age-moyen = 75 # erreur : en fait python l'interprète comme 'age - moyen = 75'
```

```
File "<ipython-input-9-78de3a1bfc60>", line 1
age-moyen = 75 # erreur : en fait python l'interprète comme 'age - moyen = 75'
SyntaxError: can't assign to operator
```

### Caractères exotiques

En Python 3, il est maintenant aussi possible d'utiliser des caractères Unicode dans les identificateurs :

```
In [10]: # les caractères accentués sont permis
nom_élève = "Jules Maigret"
```

```
In [11]: # ainsi que l'alphabet grec
from math import cos, pi as  $\pi$ 
 $\theta$  =  $\pi$  / 4
cos( $\theta$ )
```

```
Out[11]: 0.7071067811865476
```

Tous les caractères Unicode ne sont pas permis - heureusement car cela serait source de confusion. Nous citons dans les références les documents qui précisent quels sont exactement les caractères autorisés.

```
In [12]: # ce caractère n'est pas autorisé, car il
# est considéré comme un signe mathématique (produit)
 $\prod$  = 10
```

```
File "<ipython-input-12-4b0589d77bdc>", line 3
Π = 10
^
SyntaxError: invalid character in identifier
```

```
In [13]: # ce caractère est encore différent, c'est aussi
        # un pi grec mais pas le même, cette fois-ci
        # c'est un nom de variable acceptable mais
        # il n'est pas défini
        Π
```

```
-----

NameError                                Traceback (most recent call last)

<ipython-input-13-42996045225e> in <module>()
      3 # c'est un nom de variable acceptable mais
      4 # il n'est pas défini
----> 5 Π

NameError: name 'Π' is not defined
```

**Conseil** Il est très vivement recommandé :

- tout d’abord de coder **en anglais** ;
- ensuite de **ne pas** définir des identificateurs avec des caractères non ASCII, dans toute la mesure du possible , voyez par exemple la confusion que peut créer le fait de nommer un identificateur  $\pi$  ou  $\Pi$  ou  $\Pi$  ;
- enfin si vous utilisez un encodage autre que UTF-8, vous **devez** bien **spécifier l’encodage** utilisé dans votre fichier source ; nous y reviendrons en deuxième semaine.

### Pour en savoir plus

Pour les esprits curieux, Guido van Rossum, le fondateur de Python, est le co-auteur d’un document qui décrit les conventions de codage à utiliser dans la bibliothèque standard Python. Ces règles sont plus restrictives que ce que le langage permet de faire, mais constituent une lecture intéressante si vous projetez d’écrire beaucoup de Python.

Voir dans le PEP 008 [la section consacrée aux règles de nommage - \(en anglais\)](#)

Voir enfin, au sujet des caractères exotiques dans les identificateurs :

- <https://www.python.org/dev/peps/pep-3131/> qui définit les caractères exotiques autorisés, et qui repose à son tour sur
- <http://www.unicode.org/reports/tr31/> (très technique !)

## 1.11 Les mots-clés de Python

### Mots réservés

Il existe en Python certains mots spéciaux, qu'on appelle des mots-clés, ou *keywords* en anglais, qui sont réservés et **ne peuvent pas être utilisés** comme identifiants, c'est-à-dire comme un nom de variable.

C'est le cas par exemple pour l'instruction `if`, que nous verrons prochainement, qui permet bien entendu d'exécuter tel ou tel code selon le résultat d'un test.

```
In [1]: variable = 15
        if variable <= 10:
            print("en dessous de la moyenne")
        else:
            print("au dessus")
```

au dessus

À cause de la présence de cette instruction dans le langage, il n'est pas autorisé d'appeler une variable `if`.

```
In [2]: # interdit, if est un mot-clé
        if = 1
```

```
File "<ipython-input-2-f16082c36546>", line 2
if = 1
^
SyntaxError: invalid syntax
```

### Liste complète

Voici la liste complète des mots-clés :

<b>False</b>	<code>class</code>	<code>finally</code>	<code>is</code>	<code>return</code>
<b>None</b>	<code>continue</code>	<code>for</code>	<code>lambda</code>	<code>try</code>
<b>True</b>	<code>def</code>	<code>from</code>	<b><code>nonlocal</code></b>	<code>while</code>
<code>and</code>	<code>del</code>	<code>global</code>	<code>not</code>	<code>with</code>
<code>as</code>	<code>elif</code>	<code>if</code>	<code>or</code>	<code>yield</code>
<code>assert</code>	<code>else</code>	<code>import</code>	<code>pass</code>	
<code>break</code>	<code>except</code>	<code>in</code>	<code>raise</code>	

Nous avons indiqué **en gras** les nouveautés **par rapport à Python 2** (sachant que réciproquement `exec` et `print` ont perdu leur statut de mot-clé depuis Python 2, ce sont maintenant des fonctions).

Il vous faudra donc y prêter attention, surtout au début, mais avec un tout petit peu d'habitude vous saurez rapidement les éviter.



Vous remarquerez aussi que tous les bons éditeurs de texte supportant du code Python vont colorer les mots-clés différemment des variables. Par exemple, IDLE colorie les mots-clés en orange, vous pouvez donc très facilement vous rendre compte que vous allez, par erreur, en utiliser un comme nom de variable.

Cette fonctionnalité, dite de *coloration syntaxique*, permet d'identifier d'un coup d'œil, grâce à un code de couleur, le rôle des différents éléments de votre code : variables, mots-clés, etc... D'une manière générale, nous vous déconseillons fortement d'utiliser un éditeur de texte qui n'offre pas cette fonctionnalité de coloration syntaxique.

**Pour en savoir plus**

On peut se reporter à cette page :

[https://docs.python.org/3/reference/lexical\\_analysis.html#keywords](https://docs.python.org/3/reference/lexical_analysis.html#keywords)

## 1.12 Un peu de calcul sur les types

### 1.12.1 Complément - niveau basique

#### La fonction `type`

Nous avons vu dans la vidéo que chaque objet possède un type. On peut très simplement accéder au type d'un objet en appelant une fonction *built-in*, c'est-à-dire prédéfinie dans Python, qui s'appelle, eh bien oui, `type`.

On l'utilise tout simplement comme ceci :

```
In [1]: type(1)
```

```
Out[1]: int
```

```
In [2]: type('spam')
```

```
Out[2]: str
```

Cette fonction est assez peu utilisée par les programmeurs expérimentés, mais va nous être utile à bien comprendre le langage, notamment pour manipuler les valeurs numériques.

#### Types, variables et objets

On a vu également que le type est attaché à l'objet et non à la variable.

```
In [3]: x = 1  
        type(x)
```

```
Out[3]: int
```

```
In [4]: # la variable x peut référencer un objet de n'importe quel type
```

```
        x = [1, 2, 3]  
        type(x)
```

```
Out[4]: list
```

### 1.12.2 Complément - niveau avancé

#### La fonction `isinstance`

Une autre fonction prédéfinie, voisine de `type` mais plus utile dans la pratique, est la fonction `isinstance` qui permet de savoir si un objet est d'un type donné. Par exemple :

```
In [5]: isinstance(23, int)
```

```
Out[5]: True
```

À la vue de ce seul exemple, on pourrait penser que `isinstance` est presque identique à `type`; en réalité elle est un peu plus élaborée, notamment pour la programmation objet et l'héritage, nous aurons l'occasion d'y revenir.

On remarque ici en passant que la variable `int` est connue de Python alors que nous ne l'avons pas définie. Il s'agit d'une variable prédéfinie, qui désigne le type des entiers, que nous étudierons très bientôt.

Pour conclure sur `isinstance`, cette fonction est utile en pratique précisément parce que Python est à typage dynamique. Aussi il est souvent utile de s'assurer qu'une variable passée à une fonction est du (ou des) type(s) attendu(s), puisque contrairement à un langage typé statiquement comme C++, on n'a aucune garantie de ce genre à l'exécution. À nouveau, nous aurons l'occasion de revenir sur ce point.

## 1.13 Gestion de la mémoire

### 1.13.1 Complément - niveau basique

L'objet de ce complément est de vous montrer qu'avec Python vous n'avez pas à vous préoccuper de la mémoire. Pour expliquer la notion de gestion de la mémoire, il nous faut donner un certain nombre de détails sur d'autres langages comme C et C++. Si vous souhaitez suivre ce cours à un niveau basique vous pouvez ignorer ce complément et seulement retenir que Python se charge de tout pour vous :)

### 1.13.2 Complément - niveau intermédiaire

#### Langages de bas niveau

Dans un langage traditionnel de bas niveau comme C ou C++, le programmeur est en charge de l'allocation - et donc de la libération - de la mémoire.

Ce qui signifie que, sauf pour les valeurs stockées dans la pile, le programmeur est amené :

- à réclamer de la mémoire au système d'exploitation en appelant explicitement `malloc` (C) ou `new` (C++);
- et réciproquement à rendre cette mémoire au système d'exploitation lorsqu'elle n'est plus utilisée, en appelant `free` (C) ou `delete` (C++).

Avec ce genre de langage, la gestion de la mémoire est un aspect important de la programmation. Ce modèle offre une grande flexibilité, mais au prix d'un coût élevé en matière de vitesse de développement.

En effet, il est assez facile d'oublier de libérer la mémoire après usage, ce qui peut conduire à épuiser les ressources disponibles. À l'inverse, utiliser une zone mémoire non allouée peut conduire à des bugs très difficiles à localiser et à des problèmes de sécurité majeurs. Notons qu'une grande partie des attaques en informatique reposent sur l'exploitation d'erreurs de gestion de la mémoire.

#### Langages de haut niveau

Pour toutes ces raisons, avec un langage de plus haut niveau comme Python, le programmeur est libéré de cet aspect de la programmation.

Pour anticiper un peu sur le cours des semaines suivantes, voici ce que vous pouvez garder en tête s'agissant de la gestion mémoire en Python : \* vous créez vos objets au fur et à mesure de vos besoins ;

- vous n'avez pas besoin de les libérer explicitement, le "*Garbage Collector*" de Python va s'en charger pour recycler la mémoire lorsque c'est possible ;
- Python a tendance à être assez gourmand en mémoire, comparé à un langage de bas niveau, car tout est objet et chaque objet est assorti de *méta-informations* qui occupent une place non négligeable. Par exemple, chaque objet possède au minimum :
  - une référence vers son type - c'est le prix du typage dynamique ;
  - un compteur de références - le nombre d'autres valeurs (variables ou objets) qui pointent vers l'objet, cette information est notamment utilisée, précisément, par le *Garbage Collector* pour déterminer si la mémoire utilisée par un objet peut être libérée ou non.

- un certain nombre de types prédéfinis et non mutables sont implémentés en Python comme des *singletons*, c'est-à-dire qu'un seul objet est créé et partagé, c'est le cas par exemple pour les petits entiers et les chaînes de caractères, on en reparlera ;
- lorsqu'on implémente une classe, il est possible de lui conférer cette caractéristique de singleton, de manière à optimiser la mémoire nécessaire pour exécuter un programme.

## 1.14 Typages statique et dynamique

### 1.14.1 Complément - niveau intermédiaire

Parmi les langages typés, on distingue les langages à typage statique et ceux à typage dynamique. Ce notebook tente d'éclaircir ces notions pour ceux qui n'y sont pas familiers.

#### Typage statique

À une extrémité du spectre, on trouve les langages compilés, dits à typage statique, comme par exemple C ou C++.

En C on écrira, par exemple, une version simpliste de la fonction factoriel comme ceci :

```
int factoriel(int n) {
    int result = 1;
    for (int loop = 1; loop <= n; loop++)
        result *= loop;
    return result;
}
```

Comme vous pouvez le voir - ou le deviner - toutes les **variables** utilisées ici (comme par exemple `n`, `result` et `loop`) sont typées :

- on doit appeler `factoriel` avec un argument `n` qui doit être un entier (`int` est le nom du type entier);
- les variables internes `result` et `loop` sont de type entier;
- `factoriel` retourne une valeur de type entier.

Ces informations de type ont essentiellement trois fonctions :

- en premier lieu, elles sont nécessaires au compilateur. En C si le programmeur ne précisait pas que `result` est de type entier, le compilateur n'aurait pas suffisamment d'éléments pour générer le code assembleur correspondant;
- en contrepartie, le programmeur a un contrôle très fin de l'usage qu'il fait de la mémoire et du matériel. Il peut choisir d'utiliser un entier sur 32 ou 64 bits, signé ou pas, ou construire avec `struct` et `union` un arrangement de ses données;
- enfin, et surtout, ces informations de type permettent de faire un contrôle *a priori* de la validité du programme, par exemple, si à un autre endroit dans le code on trouve :

```
#include <stdio.h>

int main(int argc, char *argv[]) {
    /* le premier argument de la ligne de commande est argv[1] */
    char *input = argv[1];
    /* calculer son factoriel et afficher le résultat */
    printf("Factoriel (%s) = %d\n", input, factoriel(input));
    /*
        * ici on appelle factoriel avec une entrée de type 'chaîne de caractères' */
}
```

alors le compilateur va remarquer qu'on essaie d'appeler `factoriel` avec comme argument `input` qui, pour faire simple, est une chaîne de caractères et comme `factoriel` s'attend à recevoir un entier, ce programme n'a aucune chance de compiler.

On parle alors de **typage statique**, en ce sens que chaque **variable** a exactement un type qui est défini par le programmeur une bonne fois pour toutes.

C'est ce qu'on appelle le **contrôle de type**, ou *type-checking* en anglais. Si on ignore le point sur le contrôle fin de la mémoire, qui n'est pas crucial à notre sujet, ce modèle de contrôle de type présente :

- l'**inconvénient** de demander davantage au programmeur (je fais abstraction, à ce stade et pour simplifier, de **langages à inférence de types** comme ML et Haskell);
- et l'**avantage** de permettre un contrôle étendu, et surtout précoce (avant même de l'exécuter), de la bonne correction du programme.

Cela étant dit, le typage statique en C n'empêche pas le programmeur débutant d'essayer d'écrire dans la mémoire à partir d'un pointeur NULL - et le programme de s'interrompre brutalement. Il faut être conscient des limites du typage statique.

### Typage dynamique

À l'autre bout du spectre, on trouve des langages comme, eh bien, Python.

Pour comprendre cette notion de typage dynamique, regardons la fonction suivante `somme`.

```
In [1]: def somme(*largs):
        "retourne la somme de tous ses arguments"
        if not largs:
            return 0
        result = largs[0]
        for i in range(1, len(largs)):
            result += largs[i]
        return result
```

Naturellement, vous n'êtes pas à ce stade en mesure de comprendre le fonctionnement intime de la fonction. Mais vous pouvez tout de même l'utiliser :

```
In [2]: somme(12, 14, 300)
```

```
Out[2]: 326
```

```
In [3]: liste1 = ['a', 'b', 'c']
        liste2 = [0, 20, 30]
        liste3 = ['spam', 'eggs']
        somme(liste1, liste2, liste3)
```

```
Out[3]: ['a', 'b', 'c', 0, 20, 30, 'spam', 'eggs']
```

Vous pouvez donc constater que `somme` peut fonctionner avec des objets de types différents. En fait, telle qu'elle est écrite, elle va fonctionner s'il est possible de faire + entre ses arguments. Ainsi, par exemple, on pourrait même faire :

```
In [4]: # Python sait faire + entre deux chaînes de caractères
        somme('abc', 'def')
```

```
Out[4]: 'abcdef'
```

Mais par contre on ne pourrait pas faire

```
In [5]: # ceci va déclencher une exception à l'exécution
        somme(12, [1, 2, 3])
```

```
-----

TypeError                                Traceback (most recent call last)

<ipython-input-5-1b5269e9e129> in <module>()
      1 # ceci va déclencher une exception à l'exécution
----> 2 somme(12, [1, 2, 3])

<ipython-input-1-29005c25d5bb> in somme(*larges)
      5     result = larges[0]
      6     for i in range(1, len(larges)):
----> 7         result += larges[i]
      8     return result

TypeError: unsupported operand type(s) for +=: 'int' and 'list'
```

Il est utile de remarquer que le typage de Python, qui existe bel et bien comme on le verra, est qualifié de dynamique parce que le type est attaché à **un objet** et non à la variable qui le référence. On aura bien entendu l'occasion d'approfondir tout ça dans le cours.

En Python, on fait souvent référence au typage sous l'appellation *duck typing*, de manière imagée :

If it looks like a duck and quacks like a duck, it's a duck.

On voit qu'on se trouve dans une situation très différente de celle du programmeur C/C++ en ce sens que :

- à l'écriture du programme, il n'y a aucun des surcoûts qu'on trouve avec C ou C++ en matière de définition de type ;
- aucun contrôle de type n'est effectué *a priori* par le langage au moment de la définition de la fonction `somme` ;
- par contre au moment de l'exécution, s'il s'avère qu'on tente de faire une somme entre deux types qui ne peuvent pas être additionnés, comme ci-dessus avec un entier et une liste, le programme ne pourra pas se dérouler correctement.

Il y a deux points de vue vis-à-vis de la question du typage.

Les gens habitués au *typage statique* se plaignent du typage dynamique en disant qu'on peut écrire des programmes faux et qu'on s'en rend compte trop tard - à l'exécution.

À l'inverse les gens habitués au *typage dynamique* font valoir que le typage statique est très partiel, par exemple, en C si on essaie d'écrire dans un pointeur NULL, le système d'exploitation ne le permet pas et le programme sort tout aussi brutalement.



Bref, selon le point de vue, le typage dynamique est vécu comme un inconvénient (pas assez de bonnes propriétés détectées par le langage) ou comme un avantage (pas besoin de passer du temps à déclarer le type des variables, ni à faire des conversions pour satisfaire le compilateur).

Vous remarquerez cependant à l'usage, qu'en matière de vitesse de développement, les inconvénients du typage dynamique sont très largement compensés par ses avantages.

### *Type hints*

Signalons enfin que depuis python-3.5, il est **possible** d'ajouter des annotations de type, pour expliciter les suppositions qui sont faites par le programmeur pour le bon fonctionnement du code.

Nous aurons là encore l'occasion de détailler ce point dans le cours, signalons simplement que ces annotations sont totalement optionnelles, et que même lorsqu'elles sont présentes elles ne sont pas utilisées à l'exécution par l'interpréteur. L'idée est plutôt de permettre à des outils externes, [comme par exemple mypy](#), d'effectuer des contrôles plus poussés concernant la correction du programme.

## 1.15 Utiliser Python comme une calculette

Lorsque vous démarrez l'interprète Python, vous disposez en fait d'une calculette, par exemple, vous pouvez taper :

```
In [1]: 20 * 60
```

```
Out[1]: 1200
```

Les règles de **priorité** entre les opérateurs sont habituelles, les produits et divisions sont évalués en premier, ensuite les sommes et soustractions :

```
In [2]: 2 * 30 + 10 * 5
```

```
Out[2]: 110
```

De manière générale, il est recommandé de bien parenthéser ses expressions. De plus, les parenthèses facilitent la lecture d'expressions complexes.

Par exemple, il vaut mieux écrire ce qui suit, qui est équivalent mais plus lisible :

```
In [3]: (2 * 30) + (10 * 5)
```

```
Out[3]: 110
```

Attention, en Python la division / est une division naturelle :

```
In [4]: 48 / 5
```

```
Out[4]: 9.6
```

Rappelez-vous des opérateurs suivants qui sont très pratiques :

code	opération
//	quotient
%	modulo
**	puissance

```
In [5]: # calculer un quotient  
48 // 5
```

```
Out[5]: 9
```

```
In [6]: # modulo (le reste de la division par)  
48 % 5
```

```
Out[6]: 3
```

```
In [7]: # puissance  
2 ** 10
```

```
Out[7]: 1024
```

Vous pouvez facilement faire aussi des calculs sur les complexes. Souvenez-vous seulement que la constante complexe que nous notons  $i$  en français se note  $j$  en Python, ce choix a été fait par [le BDFL](#) - alias Guido van Rossum - pour des raisons de lisibilité :

```
In [8]: # multiplication de deux nombres complexes
        (2 + 3j) * 2.5j
```

```
Out[8]: (-7.5+5j)
```

Aussi, pour entrer ce nombre complexe  $j$ , il faut toujours le faire précéder d'un nombre, donc ne pas entrer simplement  $j$  (qui serait compris comme un nom de variable, nous allons voir ça tout de suite) mais plutôt  $1j$  ou encore  $1.j$ , comme ceci :

```
In [9]: 1j * 1.j
```

```
Out[9]: (-1+0j)
```

### Utiliser des variables

Il peut être utile de stocker un résultat qui sera utilisé plus tard, ou de définir une valeur constante. Pour cela on utilise tout simplement une affectation comme ceci :

```
In [10]: # pour définir une variable il suffit de lui assigner une valeur
         largeur = 5
```

```
In [11]: # une fois la variable définie, on peut l'utiliser, ici comme un nombre
         largeur * 20
```

```
Out[11]: 100
```

```
In [12]: # après quoi bien sûr la variable reste inchangée
         largeur * 10
```

```
Out[12]: 50
```

Pour les symboles mathématiques, on peut utiliser la même technique :

```
In [13]: # pour définir un réel, on utilise le point au lieu d'une virgule en français
         pi = 3.14159
         2 * pi * 10
```

```
Out[13]: 62.8318
```

Pour les valeurs spéciales comme  $\pi$ , on peut utiliser les valeurs prédéfinies par la bibliothèque mathématique de Python. En anticipant un peu sur la notion d'importation que nous approfondirons plus tard, on peut écrire :

```
In [14]: from math import e, pi
```

Et ainsi imprimer les racines troisièmes de l'unité par la formule :

$$r_n = e^{2i\pi\frac{n}{3}}, \text{ pour } n \in \{0, 1, 2\}$$

```
In [15]: n = 0
          print("n=", n, "racine = ", e**((2.j*pi*n)/3))
          n = 1
          print("n=", n, "racine = ", e**((2.j*pi*n)/3))
          n = 2
          print("n=", n, "racine = ", e**((2.j*pi*n)/3))

n= 0 racine = (1+0j)
n= 1 racine = (-0.4999999999999998+0.8660254037844387j)
n= 2 racine = (-0.5000000000000004-0.8660254037844384j)
```

**Remarque :** bien entendu il sera possible de faire ceci plus simplement lorsque nous aurons vu les boucles for.

### Les types

Ce qui change par rapport à une calculatrice standard est le fait que les valeurs sont typées. Pour illustrer les trois types de nombres que nous avons vus jusqu'ici :

```
In [16]: # le type entier s'appelle 'int'
          type(3)

Out[16]: int

In [17]: # le type flottant s'appelle 'float'
          type(3.5)

Out[17]: float

In [18]: # le type complexe s'appelle 'complex'
          type(1j)

Out[18]: complex
```

### Chaînes de caractères

On a également rapidement besoin de chaînes de caractères, on les étudiera bientôt en détail, mais en guise d'avant-goût :

```
In [19]: chaine = "Bonjour le monde !"
          print(chaine)
```

Bonjour le monde !

### Conversions

Il est parfois nécessaire de convertir une donnée d'un type dans un autre. Par exemple on peut demander à l'utilisateur d'entrer une valeur au clavier grâce à la fonction input, comme ceci :

```
In [20]: reponse = input("quel est votre âge ? ")
```

quel est votre âge ? 42

```
In [21]: # vous avez entré la chaîne suivante
         print(reponse)
```

42

```
In [22]: # ici reponse est une variable, et son contenu est de type chaîne de caractères
         type(reponse)
```

Out[22]: str

Maintenant je veux faire des calculs sur votre âge, par exemple le multiplier par 2. Si je m'y prends naïvement, ça donne ceci :

```
In [23]: # multiplier une chaîne de caractères par deux ne fait pas ce que l'on veut,
         # nous verrons plus tard que ça fait une concaténation
         2 * reponse
```

Out[23]: '4242'

C'est pourquoi il me faut ici d'abord **convertir** la (valeur de la) variable `reponse` en un entier, que je peux ensuite doubler (assurez-vous d'avoir bien entré ci-dessus une valeur qui correspond à un nombre entier)

```
In [24]: # reponse est une chaine
         # je la convertis en entier en appelant la fonction int()
         age = int(reponse)
         type(age)
```

Out[24]: int

```
In [25]: # que je peux maintenant multiplier par 2
         2 * age
```

Out[25]: 84

Ou si on préfère, en une seule fois :

```
In [26]: print("le double de votre age est", 2*int(reponse))
```

le double de votre age est 84

## Conversions - suite

De manière plus générale, pour convertir un objet en un entier, un flottant, ou une chaîne de caractères, on peut simplement appeler une fonction *built-in* qui porte le même nom que le type cible :

Type	Fonction
Entier	<code>int</code>
Flottant	<code>float</code>
Complexe	<code>complex</code>
Chaîne	<code>str</code>

Ainsi dans l'exemple précédent, `int(reponse)` représente la conversion de `reponse` en entier.

On a illustré cette même technique dans les exemples suivants :

```
In [27]: # dans l'autre sens, si j'ai un entier
a = 2345
```

```
In [28]: # je peux facilement le traduire en chaîne de caractères
str(2345)
```

```
Out[28]: '2345'
```

```
In [29]: # ou en complexe
complex(2345)
```

```
Out[29]: (2345+0j)
```

Nous verrons plus tard que ceci se généralise à tous les types de Python, pour convertir un objet `x` en un type bidule, on appelle `bidule(x)`. On y reviendra, bien entendu.

### Grands nombres

Comme les entiers sont de précision illimitée, on peut améliorer leur lisibilité en insérant des caractères `_` qui sont simplement ignorés à l'exécution.

```
In [30]: tres_grand_nombre = 23_456_789_012_345

tres_grand_nombre
```

```
Out[30]: 23456789012345
```

```
In [31]: # ça marche aussi avec les flottants
123_456.789_012
```

```
Out[31]: 123456.789012
```

### Entiers et bases

Les calculettes scientifiques permettent habituellement d'entrer les entiers dans d'autres bases que la base 10.

En Python, on peut aussi entrer un entier sous forme binaire comme ceci :

```
In [32]: deux_cents = 0b11001000
print(deux_cents)
```

200

Ou encore sous forme octale (en base 8) comme ceci :

```
In [33]: deux_cents = 0o310  
         print(deux_cents)
```

200

Ou enfin encore en hexadécimal (base 16) comme ceci :

```
In [34]: deux_cents = 0xc8  
         print(deux_cents)
```

200

Pour d'autres bases, on peut utiliser la fonction de conversion `int` en lui passant un argument supplémentaire :

```
In [35]: deux_cents = int('3020', 4)  
         print(deux_cents)
```

200

## Fonctions mathématiques

Python fournit naturellement un ensemble très complet d'opérateurs mathématiques pour les fonctions exponentielles, trigonométriques et autres, mais leur utilisation ne nous est pas encore accessible à ce stade et nous les verrons ultérieurement.

## 1.16 Affectations et Opérations (à la +=)

### 1.16.1 Complément - niveau intermédiaire

Il existe en Python toute une famille d'opérateurs dérivés de l'affectation qui permettent de faire en une fois une opération et une affectation. En voici quelques exemples.

#### Incrémentation

On peut facilement augmenter la valeur d'une variable numérique comme ceci :

```
In [1]: entier = 10

        entier += 2
        print('entier', entier)

entier 12
```

Comme on le devine peut-être, ceci est équivalent à :

```
In [2]: entier = 10

        entier = entier + 2
        print('entier', entier)

entier 12
```

#### Autres opérateurs courants

Cette forme, qui combine opération sur une variable et réaffectation du résultat à la même variable, est disponible avec tous les opérateurs courants :

```
In [3]: entier -= 4
        print('après décrémentation', entier)
        entier *= 2
        print('après doublement', entier)
        entier /= 2
        print('mis à moitié', entier)

après décrémentation 8
après doublement 16
mis à moitié 8.0
```

#### Types non numériques

En réalité cette construction est disponible sur tous les types qui supportent l'opérateur en question. Par exemple, les listes (que nous verrons bientôt) peuvent être additionnées entre elles :



```
In [4]: liste = [0, 3, 5]
        print('liste', liste)

        liste += ['a', 'b']
        print('après ajout', liste)
```

```
liste [0, 3, 5]
après ajout [0, 3, 5, 'a', 'b']
```

Beaucoup de types supportent l'opérateur +, qui est sans doute de loin celui qui est le plus utilisé avec cette construction.

### Opérateurs plus abscons

Signalons enfin que l'on trouve aussi cette construction avec d'autres opérateurs moins fréquents, par exemple :

```
In [5]: entier = 2
        print('entier:', entier)
        entier **= 10
        print('à la puissance dix:', entier)
        entier %= 5
        print('modulo 5:', entier)
```

```
entier: 2
à la puissance dix: 1024
modulo 5: 4
```

Et pour ceux qui connaissent déjà un peu Python, on peut même le faire avec des opérateurs de décalage, que nous verrons très bientôt :

```
In [6]: entier <<= 2
        print('double décalage gauche:', entier)
```

```
double décalage gauche: 16
```

## 1.17 Notions sur la précision des calculs flottants

### 1.17.1 Complément - niveau avancé

#### Le problème

Comme pour les entiers, les calculs sur les flottants sont, naturellement, réalisés par le processeur. Cependant contrairement au cas des entiers où les calculs sont toujours exacts, les flottants posent un problème de précision. Cela n'est pas propre au langage Python, mais est dû à la technique de codage des nombres flottants sous forme binaire.

Voyons tout d'abord comment se matérialise le problème :

```
In [1]: 0.2 + 0.4
```

```
Out[1]: 0.6000000000000001
```

Il faut retenir que lorsqu'on écrit un nombre flottant sous forme décimale, la valeur utilisée en mémoire pour représenter ce nombre, parce que cette valeur est codée en binaire, ne représente **pas toujours exactement** le nombre entré.

```
In [2]: # du coup cette expression est fausse, à cause de l'erreur d'arrondi
0.3 - 0.1 == 0.2
```

```
Out[2]: False
```

Aussi, comme on le voit, les différentes erreurs d'arrondi qui se produisent à chaque étape du calcul s'accumulent et produisent un résultat parfois surprenant. De nouveau, ce problème n'est pas spécifique à Python, il existe pour tous les langages, et il est bien connu des numériciens.

Dans une grande majorité des cas, ces erreurs d'arrondi ne sont pas pénalisantes. Il faut toutefois en être conscient car cela peut expliquer des comportements curieux.

#### Une solution : penser en termes de nombres rationnels

Tout d'abord si votre problème se pose bien en termes de nombres rationnels, il est alors tout à fait possible de le résoudre avec exactitude.

Alors qu'il n'est pas possible d'écrire exactement  $3/10$  en base 2, ni d'ailleurs  $1/3$  en base 10, on peut représenter **exactement** ces nombres dès lors qu'on les considère comme des fractions et qu'on les encode avec deux nombres entiers.

Python fournit en standard le module `fractions` qui permet de résoudre le problème. Voici comment on pourrait l'utiliser pour vérifier, cette fois avec succès, que  $0.3 - 0.1$  vaut bien  $0.2$ . Ce code anticipe sur l'utilisation des modules et des classes en Python, ici nous créons des objets de type `Fraction` :

```
In [3]: # on importe le module fractions, qui lui-même définit le symbole Fraction
from fractions import Fraction

# et cette fois, les calculs sont exacts, et l'expression retourne bien True
Fraction(3, 10) - Fraction(1, 10) == Fraction(2, 10)
```

```
Out[3]: True
```

Ou encore d'ailleurs, équivalent et plus lisible :

```
In [4]: Fraction('0.3') - Fraction('0.1') == Fraction('2/10')
```

```
Out[4]: True
```

### Une autre solution : le module decimal

Si par contre vous ne manipulez pas des nombres rationnels et que du coup la représentation sous forme de fractions ne peut pas convenir dans votre cas, signalons l'existence du module standard `decimal` qui offre des fonctionnalités très voisines du type `float`, tout en éliminant la plupart des inconvénients, au prix naturellement d'une consommation mémoire supérieure.

Pour reprendre l'exemple de départ, mais en utilisant le module `decimal`, on écrirait alors :

```
In [5]: from decimal import Decimal
```

```
Decimal('0.3') - Decimal('0.1') == Decimal('0.2')
```

```
Out[5]: True
```

### Pour aller plus loin

Tous ces documents sont en anglais :

- un [tutoriel sur les nombres flottants](#);
- la [documentation sur la classe Fraction](#);
- la [documentation sur la classe Decimal](#).

## 1.18 Opérations *bit à bit* (*bitwise*)

### 1.18.1 Compléments - niveau avancé

Les compléments ci-dessous expliquent des fonctions évoluées sur les entiers. Les débutants en programmation peuvent sans souci sauter cette partie en cas de difficultés.

#### Opérations logiques : ET &, OU | et OU exclusif ^

Il est possible aussi de faire des opérations *bit à bit* sur les nombres entiers. Le plus simple est de penser à l'écriture du nombre en base 2.

Considérons par exemple deux entiers constants dans cet exercice

```
In [1]: x49 = 49
        y81 = 81
```

Ce qui nous donne comme décomposition binaire :

$$\begin{aligned} x_{49} &= 49 = 32 + 16 + 1 \rightarrow (0, 1, 1, 0, 0, 0, 1) \\ y_{81} &= 81 = 64 + 16 + 1 \rightarrow (1, 0, 1, 0, 0, 0, 1) \end{aligned}$$

Pour comprendre comment passer de  $32 + 16 + 1$  à  $(0, 1, 1, 0, 0, 0, 1)$  il suffit d'observer que :

$$32 + 16 + 1 = 0 * 2^6 + 1 * 2^5 + 1 * 2^4 + 0 * 2^3 + 0 * 2^2 + 0 * 2^1 + 1 * 2^0$$

#### Et logique : opérateur &

L'opération logique & va faire un *et* logique bit à bit entre les opérandes, ainsi

```
In [2]: x49 & y81
```

```
Out [2]: 17
```

Et en effet :

$$\begin{aligned} x_{49} &\rightarrow (0, 1, 1, 0, 0, 0, 1) \\ y_{81} &\rightarrow (1, 0, 1, 0, 0, 0, 1) \\ x_{49} \&y_{81} &\rightarrow (0, 0, 1, 0, 0, 0, 1) \rightarrow 16 + 1 \rightarrow 17 \end{aligned}$$

#### Ou logique : opérateur |

De même, l'opérateur logique | fait simplement un *ou* logique, comme ceci :

```
In [3]: x49 | y81
```

```
Out [3]: 113
```

On s'y retrouve parce que :

$$\begin{aligned} x_{49} &\rightarrow (0, 1, 1, 0, 0, 0, 1) \\ y_{81} &\rightarrow (1, 0, 1, 0, 0, 0, 1) \\ x_{49} | y_{81} &\rightarrow (1, 1, 1, 0, 0, 0, 1) \rightarrow 64 + 32 + 16 + 1 \rightarrow 113 \end{aligned}$$

**Ou exclusif : opérateur ^**

Enfin, on peut également faire la même opération à base de *ou exclusif* avec l'opérateur ^ :

```
In [4]: x49 ^ y81
```

```
Out[4]: 96
```

Je vous laisse le soin de décortiquer le calcul à titre d'exercice (le ou exclusif de deux bits est vrai si et seulement si exactement une des deux entrées est vraie).

**Décalages**

Un décalage à **gauche** de, par exemple, 4 positions, revient à décaler tout le champ de bits de 4 cases à gauche (les 4 nouveaux bits insérés sont toujours des 0). C'est donc équivalent à une **multiplication** par  $2^4 = 16$  :

```
In [5]: x49 << 4
```

```
Out[5]: 784
```

$$\begin{aligned} x49 &\rightarrow (0, 1, 1, 0, 0, 0, 1) \\ x49 << 4 &\rightarrow (0, 1, 1, 0, 0, 0, 1, 0, 0, 0, 0) \rightarrow 512 + 256 + 16 \rightarrow 784 \end{aligned}$$

De la même manière, le décalage à **droite** de  $n$  revient à une **division** par  $2^n$  (plus précisément, le quotient de la division) :

```
In [6]: x49 >> 4
```

```
Out[6]: 3
```

$$\begin{aligned} x49 &\rightarrow (0, 1, 1, 0, 0, 0, 1) \\ x49 >> 4 &\rightarrow (0, 0, 0, 0, 0, 1, 1) \rightarrow 2 + 1 \rightarrow 3 \end{aligned}$$

**Une astuce**

On peut utiliser la fonction *built-in* `bin` pour calculer la représentation binaire d'un entier. Attention, la valeur de retour est une chaîne de caractères de type `str` :

```
In [7]: bin(x49)
```

```
Out[7]: '0b110001'
```

Dans l'autre sens, on peut aussi entrer un entier directement en base 2 comme ceci :

```
In [8]: x49bis = 0b110001
        x49bis == x49
```

```
Out[8]: True
```

Ici, comme on le voit, `x49bis` est bien un entier.

**Pour en savoir plus**

[Section de la documentation Python.](#)

## 1.19 Estimer le plus petit (grand) flottant

### 1.19.1 Exercice - niveau basique

#### Le plus petit flottant

En corollaire de la discussion sur la précision des flottants, il faut savoir que le système de codage en mémoire impose aussi une limite. Les réels très petits, ou très grands, ne peuvent plus être représentés de cette manière.

C'est notamment très gênant si vous implémentez un logiciel probabiliste, comme des graphes de Markov, où les probabilités d'occurrence de séquences très longues tendent très rapidement vers des valeurs extrêmement petites.

Le but de cet exercice est d'estimer la valeur du plus petit flottant qui peut être représenté comme un flottant. Pour vous aider, voici deux valeurs :

```
In [1]: 10**-320
```

```
Out[1]: 1e-320
```

```
In [2]: 10**-330
```

```
Out[2]: 0.0
```

Comme on le voit,  $10^{-320}$  est correctement imprimé, alors que  $10^{-330}$  est, de manière erronée, rapporté comme étant nul.

#### Notes :

- À ce stade du cours, pour estimer le plus petit flottant, procédez simplement par approximations successives.
- Sans utiliser de boucle, la précision que vous pourrez obtenir n'est que fonction de votre patience, ne dépassez pas 4 à 5 itérations successives :)
- Il est par contre pertinent d'utiliser une approche rationnelle pour déterminer l'itération suivante (par opposition à une approche "au petit bonheur"). Pour ceux qui ne connaissent pas, nous vous recommandons de vous documenter sur l'algorithme de [dichotomie](#).

```
In [3]: 10**-325
```

```
Out[3]: 0.0
```

Voici quelques cellules de code vides ; vous pouvez en créer d'autres si nécessaire, le plus simple étant de taper `Alt+Enter`, ou d'utiliser le menu *"Insert -> Insert Cell Below"*

```
In [ ]: # vos essais successifs ici
```

```
In [10]: .24*10**-323
```

```
Out[10]: 0.0
```

## Le plus grand flottant

La même limitation s'applique sur les grands nombres. Toutefois, cela est un peu moins évident, car comme toujours il faut faire attention aux types :

```
In [4]: 10**450
```

[illegible]

Ce qui passe très bien car j'ai utilisé un `int` pour l'exposant, dans ce premier cas Python calcule le résultat comme un `int`, qui est un type qui n'a pas de limitation de précision (Python utilise intelligemment autant de bits que nécessaire pour ce genre de calculs).

Par contre, si j'essaie de faire le même calcul avec un exposant flottant, Python essaie cette fois de faire son calcul avec un flottant, et là on obtient une erreur :

```
In [5]: 10**450.0
```

```
-----  
OverflowError                                Traceback (most recent call last)  
  
  <ipython-input-5-d63f7e475389> in <module>()  
----> 1 10**450.0  
  
OverflowError: (34, 'Numerical result out of range')
```

On peut d'ailleurs remarquer que le comportement ici n'est pas extrêmement cohérent, car avec les petits nombres Python nous a silencieusement transformé  $10^{-330}$  en 0, alors que pour les grands nombres, il lève une exception (nous verrons les exceptions plus tard, mais vous pouvez dès maintenant remarquer que le comportement est différent dans les deux cas).

Quoi qu'il en soit, la limite pour les grands nombres se situe entre les deux valeurs  $10^{300}$  et  $10^{310}$ . On vous demande à nouveau d'estimer comme ci-dessus une valeur approchée du plus grand nombre qu'il soit possible de représenter comme un flottant.

```
In [6]: 10**300.
```

Out[6]: 1e+300

```
In [7]: 10**310.
```

---

```

OverflowError                                Traceback (most recent call last)

<ipython-input-7-5d701e1fa38c> in <module>()
----> 1 10**310.

OverflowError: (34, 'Numerical result out of range')
```

In [ ]: *# vos essais successifs ici*

### 1.19.2 Complément - niveau avancé

En fait, on peut accéder à ces valeurs minimales et maximales pour les flottants comme ceci

```

In [8]: import sys
        print(sys.float_info)

sys.float_info(max=1.7976931348623157e+308, max_exp=1024, max_10_exp=308,
min=2.2250738585072014e-308, min_exp=-1021, min_10_exp=-307, dig=15, mant_dig=53,
epsilon=2.220446049250313e-16, radix=2, rounds=1)
```

Et notamment, [comme expliqué ici](#).

```

In [9]: print("Flottant minimum", sys.float_info.min)
        print("Flottant maximum", sys.float_info.max)

Flottant minimum 2.2250738585072014e-308
Flottant maximum 1.7976931348623157e+308
```

**Sauf que** vous devez avoir trouvé un maximum voisin de cette valeur, mais le minimum observé expérimentalement ne correspond pas bien à cette valeur.

Pour ceux que cela intéresse, l'explication à cette apparente contradiction réside dans l'utilisation de [nombres dénormaux](#).



# Chapitre 2

## Notions de base, premier programme en Python

## 2.1 Caractères accentués

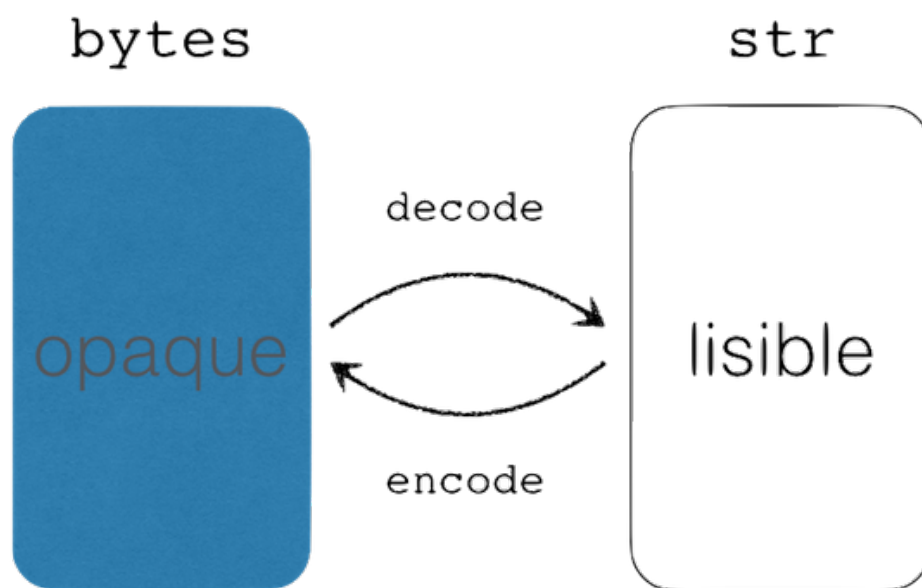
Ce complément expose quelques bases concernant les caractères accentués, et notamment les précautions à prendre pour pouvoir en insérer dans un programme Python. Nous allons voir que cette question, assez scabreuse, dépasse très largement le cadre de Python.

### 2.1.1 Complément - niveau basique

#### Un caractère n'est pas un octet

Avec Unicode, on a cassé le modèle *un caractère == un octet*. Aussi en Python 3, lorsqu'il s'agit de manipuler des données provenant de diverses sources de données :

- le type `byte` est approprié si vous voulez charger en mémoire les données binaires brutes, sous forme d'octets donc ;
- le type `str` est approprié pour représenter une chaîne de caractères - qui, à nouveau ne sont pas forcément des octets ;
- on passe de l'un à l'autre de ces types par des opérations d'encodage et décodage, comme illustré ci-dessous ;
- et pour **toutes** les opérations d'encodage et décodage, il est nécessaire de connaître l'encodage utilisé.



les types bytes et str

On peut appeler les méthodes `encode` et `decode` sans préciser l'encodage (dans ce cas Python choisit l'encodage par défaut sur votre système). Cela dit, il est de loin préférable d'être explicite et de choisir son encodage. En cas de doute, il est recommandé de **spécifier explicitement** `utf-8`, qui se généralise au détriment d'encodages anciens comme `cp1242` (Windows) et `iso8859-*`, que de laisser le système hôte choisir pour vous.

#### Utilisation des accents et autres cédilles

Python 3 supporte Unicode par défaut. Vous pouvez donc, maintenant, utiliser sans

aucun risque des accents ou des cédilles dans vos chaînes de caractères. Il faut cependant faire attention à deux choses :

- Python supporte Unicode, donc tous les caractères du monde, mais les ordinateurs n'ont pas forcément les polices de caractères nécessaires pour afficher ces caractères ;
- Python permet d'utiliser des caractères Unicode pour les noms de variables, mais nous vous recommandons dans toute la mesure du possible d'écrire votre code en anglais, comme c'est le cas pour la quasi-totalité du code que vous serez amenés à utiliser sous forme de bibliothèques. Ceci est particulièrement important pour les noms de lignes et de colonnes dans un dataset afin de faciliter les transferts entre logiciels, la majorité des logiciels n'acceptant pas les accents et cédilles dans les noms de variables.

Ainsi, il faut bien distinguer les chaînes de caractères qui doivent par nature être adaptées au langage des utilisateurs du programme, et le code source qui lui est destiné aux programmeurs et qui doit donc éviter d'utiliser autre chose que de l'anglais.

### 2.1.2 Complément - niveau intermédiaire

#### Où peut-on mettre des accents ?

Cela étant dit, si vous devez vraiment mettre des accents dans vos sources, voici ce qu'il faut savoir.

#### Noms de variables

- S'il n'était **pas possible en Python 2** d'utiliser un caractère accentué dans un **nom de variable** (ou d'un identificateur au sens large), cela est à présent **permis en Python 3** :

```
In [ ]: # pas recommandé, mais autorisé par le langage
        nb_élèves = 12
```

- On peut même utiliser des symboles, comme par exemple

```
In [ ]: from math import cos, pi as Π
        θ = Π / 4
        cos(θ)
```

- Je vous recommande toutefois de **ne pas utiliser** cette possibilité, si vous n'êtes pas extrêmement familier avec les caractères Unicode.
- Enfin, pour être exhaustif, sachez que seule une partie des caractères Unicode sont autorisés dans ce cadre, c'est heureux parce que les caractères comme, par exemple, **l'espace non-sécable** pourraient, s'ils étaient autorisés, être la cause de milliers d'heures de debugging à frustration garantie :)

Pour les curieux, vous pouvez en savoir plus [à cet endroit de la documentation officielle \(en anglais\)](#).

#### Chaînes de caractères

- Vous pouvez naturellement mettre des accents dans les chaînes de caractères. Cela dit, les données manipulées par un programme proviennent pour l'essentiel de sources externes, comme une base de données ou un formulaire Web, et donc le plus souvent pas directement du code source. Les chaînes de caractères présentes dans du vrai code sont bien souvent limitées à des messages de logging, et le plus souvent d'ailleurs en anglais, donc sans accent.

- Lorsque votre programme doit interagir avec les utilisateurs et qu’il doit donc parler leur langue, c’est une bonne pratique de créer un fichier spécifique, que l’on appelle fichier de ressources, qui contient toutes les chaînes de caractères spécifiques à une langue. Ainsi, la traduction de votre programme consistera à simplement traduire ce fichier de ressources.

```
message = "on peut mettre un caractère accentué dans une chaîne"
```

### Commentaires

- Enfin on peut aussi bien sûr mettre dans les commentaires n’importe quel caractère Unicode, et donc notamment des caractères accentués si on choisit malgré tout d’écrire le code en français.

```
# on peut mettre un caractère accentué dans un commentaire
# ainsi que  $\cos(\Theta) \forall x \in \int f(t)dt \iiint$  vous voyez l'idée générale
```

### Qu’est-ce qu’un encodage ?

Comme vous le savez, la mémoire - ou le disque - d’un ordinateur ne permet que de stocker des représentations binaires. Il n’y a donc pas de façon “naturelle” de représenter un caractère comme ‘A’, un guillemet ou un point-virgule.

On utilise pour cela un encodage, par exemple le code US-ASCII - <http://www.asciitable.com/> - stipule, pour faire simple, qu’un ‘A’ est représenté par l’octet 65 qui s’écrit en binaire 01000001. Il se trouve qu’il existe plusieurs encodages, bien sûr incompatibles, selon les systèmes et les langues. Vous trouverez plus de détails ci-dessous.

Le point important est que pour pouvoir ouvrir un fichier “proprement”, il faut bien entendu disposer du **contenu** du fichier, mais il faut aussi connaître l’**encodage** qui a été utilisé pour l’écrire.

### Précautions à prendre pour l’encodage de votre code source

L’encodage ne concerne pas simplement les objets chaîne de caractères, mais également votre code source. **Python 3** considère que votre code source utilise **par défaut l’encodage UTF-8**. Nous vous conseillons de conserver cet encodage qui est celui qui vous offrira le plus de flexibilité.

Vous pouvez malgré tout changer l’encodage **de votre code source** en faisant figurer dans vos fichiers, **en première ou deuxième ligne**, une déclaration comme ceci :

```
# -*- coding: <nom_de_l_encodage> -*-
```

ou plus simplement, comme ceci :

```
# coding: <nom_de_l_encodage>
```

Notons que la première option est également interprétée par l’éditeur de texte *Emacs* pour utiliser le même encodage. En dehors de l’utilisation d’Emacs, la deuxième option, plus simple et donc plus pythonique, est à préférer.

Le nom UTF-8 fait référence à **Unicode** (ou pour être précis, à l’encodage le plus répandu parmi ceux qui sont définis dans la norme Unicode, comme nous le verrons plus bas). Sur

certaines systèmes plus anciens vous pourrez être amenés à utiliser un autre encodage. Pour déterminer la valeur à utiliser dans votre cas précis vous pouvez faire dans l’interpréteur interactif :

```
# ceci doit être exécuté sur votre machine
import sys
print(sys.getdefaultencoding())
```

Par exemple avec d’anciennes versions de Windows (en principe de plus en plus rares) vous pouvez être amenés à écrire :

```
# coding: cp1252
```

La syntaxe de la ligne coding est précisée dans [cette documentation](#) et dans le [PEP 263](#).

### Le grand malentendu

Si je vous envoie un fichier contenant du français encodé avec, disons, ISO/IEC 8859-15 (Latin-9) – [http://en.wikipedia.org/wiki/ISO/IEC\\_8859-15](http://en.wikipedia.org/wiki/ISO/IEC_8859-15) – vous pouvez voir dans la table qu’un caractère ‘€’ va être matérialisé dans mon fichier par un octet ‘0xA4’, soit 164.

Imaginez maintenant que vous essayez d’ouvrir ce même fichier depuis un vieil ordinateur Windows configuré pour le français. Si on ne lui donne aucune indication sur l’encodage, le programme qui va lire ce fichier sur Windows va utiliser l’encodage par défaut du système, c’est-à-dire CP1252 – <http://en.wikipedia.org/wiki/Windows-1252>. Comme vous le voyez dans cette table, l’octet ‘0xA4’ correspond au caractère ☒ et c’est ça que vous allez voir à la place de €.

C’est à cela que sert la balise `# coding: <nom_de_l_encodage>`.

De cette manière, Python lorsqu’il lit votre code source, a les moyens d’interpréter correctement son contenu car il sait quel encodage utiliser. On vous rappelle que si vous ne spécifiez aucun encodage pour votre code source, Python utilisera l’encodage UTF-8, ce qui est souvent le meilleur choix.

### Pourquoi ça marche en local ?

Lorsque le producteur (le programme qui écrit le fichier) et le consommateur (le programme qui le lit) tournent dans le même ordinateur, tout fonctionne bien - en général - parce que les deux programmes se ramènent à l’encodage défini comme l’encodage par défaut. On a vu pourquoi il vaut mieux toutefois être explicite, et spécifier la balise `# coding: ...`.

Il y a toutefois une limite, si vous utilisez un Linux configuré de manière minimale, il se peut qu’il utilise par défaut l’encodage US-ASCII - voir plus bas - qui étant très ancien ne “connaît” pas un simple é, ni a fortiori €. Pour écrire du français, il faut donc au minimum que l’encodage par défaut de votre ordinateur contienne les caractères français, comme par exemple :

- ISO 8859-1 (Latin-1)
- ISO 8859-15 (Latin-9)
- UTF-8
- CP1252

## Un peu d'histoire sur les encodages

### Le code US-ASCII

Jusque dans les années 1980, les ordinateurs ne parlaient pour l'essentiel que l'anglais. La première vague de standardisation avait créé l'encodage dit ASCII, ou encore US-ASCII - voir par exemple <http://www.asciitable.com>, ou en version longue <http://en.wikipedia.org/wiki/ASCII>.

Le code US-ASCII s'étend sur 128 valeurs, soit 7 bits, mais est le plus souvent implémenté sur un octet pour préserver l'alignement, le dernier bit pouvant être utilisé par exemple pour ajouter un code correcteur d'erreur - ce qui à l'époque des modems n'était pas superflu. Bref, la pratique courante était alors de manipuler une chaîne de caractères comme un tableau d'octets.

### Les encodages ISO8859-\* (Latin\*)

Dans les années 1990, pour satisfaire les besoins des pays européens, ont été définis plusieurs encodages alternatifs, connus sous le nom de [ISO/IEC 8859-\\*](#), nommés aussi Latin-\*. Idéalement, on aurait pu et **certainement dû** définir un seul encodage pour représenter tous les nouveaux caractères, mais entre toutes les langues européennes, le nombre de caractères à ajouter était substantiel, et cet encodage unifié aurait largement dépassé 256 caractères différents, il n'aurait donc **pas été possible** de tout faire tenir sur un octet.

On a préféré préserver la "bonne propriété" du modèle *un caractère == un octet*, ceci afin de préserver le code existant qui aurait sinon dû être retouché ou réécrit.

Dès lors il n'y avait pas d'autre choix que de définir **plusieurs** encodages distincts, par exemple, pour le français on a utilisé à l'époque [ISO/IEC 8859-1 \(Latin-1\)](#), pour le russe [ISO/IEC 5589-5 \(Latin/Cyrillic\)](#).

À ce stade, le ver était dans le fruit. Depuis cette époque pour ouvrir un fichier il faut connaître son encodage.

**Unicode** Lorsque l'on a ensuite cherché à manipuler aussi les langues asiatiques, il a de toute façon fallu définir de nouveaux encodages beaucoup plus larges. C'est ce qui a été fait par le standard [Unicode](#) qui définit 3 nouveaux encodages :

- [UTF-8](#) : un encodage à taille variable, à base d'octets, qui maximise la compatibilité avec US-ASCII;
- [UTF-16](#) : un encodage à taille variable, à base de mots de 16 bits;
- [UTF-32](#) : un encodage à taille fixe, à base de mots de 32 bits;

Ces 3 standards couvrent le même jeu de caractères (113 021 tout de même dans la dernière version). Parmi ceux-ci le plus utilisé est certainement UTF-8. Un texte ne contenant que des caractères du code US-ASCII initial peut être lu avec l'encodage UTF-8.

Pour être enfin tout à fait exhaustif, si on sait qu'un fichier est au format Unicode, on peut déterminer quel est l'encodage qu'il utilise, en se basant sur les 4 premiers octets du document. Ainsi dans ce cas particulier (lorsqu'on est sûr qu'un document utilise un des trois encodages Unicode) il n'est plus nécessaire de connaître son encodage de manière "externe".

## 2.2 Les outils de base sur les chaînes de caractères (str)

### 2.2.1 Complément - niveau intermédiaire

#### Lire la documentation

Même après des années de pratique, il est difficile de se souvenir de toutes les méthodes travaillant sur les chaînes de caractères. Aussi il est toujours utile de recourir à la documentation embarquée

```
In [1]: help(str)
```

Help on class str in module builtins:

```
class str(object)
|   str(object='') -> str
|   str(bytes_or_buffer[, encoding[, errors]]) -> str
|
|   Create a new string object from the given object. If encoding or
|   errors is specified, then the object must expose a data buffer
|   that will be decoded using the given encoding and error handler.
|   Otherwise, returns the result of object.__str__() (if defined)
|   or repr(object).
|   encoding defaults to sys.getdefaultencoding().
|   errors defaults to 'strict'.
|
|   Methods defined here:
|
|   __add__(self, value, /)
|       Return self+value.
|
|   __contains__(self, key, /)
|       Return key in self.
|
|   __eq__(self, value, /)
|       Return self==value.
|
|   __format__(...)
|       S.__format__(format_spec) -> str
|
|       Return a formatted version of S as described by format_spec.
|
|   __ge__(self, value, /)
|       Return self>=value.
|
|   __getattr__(self, name, /)
|       Return getattr(self, name).
|
|   __getitem__(self, key, /)
|       Return self[key].
|
|   __getnewargs__(...)
```

```

|  __gt__(self, value, /)
|      Return self>value.
|
|  __hash__(self, /)
|      Return hash(self).
|
|  __iter__(self, /)
|      Implement iter(self).
|
|  __le__(self, value, /)
|      Return self<=value.
|
|  __len__(self, /)
|      Return len(self).
|
|  __lt__(self, value, /)
|      Return self<value.
|
|  __mod__(self, value, /)
|      Return self%value.
|
|  __mul__(self, value, /)
|      Return self*value.n
|
|  __ne__(self, value, /)
|      Return self!=value.
|
|  __new__(*args, **kwargs) from builtins.type
|      Create and return a new object.  See help(type) for accurate signature.
|
|  __repr__(self, /)
|      Return repr(self).
|
|  __rmod__(self, value, /)
|      Return value%self.
|
|  __rmul__(self, value, /)
|      Return self*value.
|
|  __sizeof__(...)
|      S.__sizeof__() -> size of S in memory, in bytes
|
|  __str__(self, /)
|      Return str(self).
|
|  capitalize(...)
|      S.capitalize() -> str
|
|      Return a capitalized version of S, i.e. make the first character
|      have upper case and the rest lower case.
|

```



```
| casefold(...)
|     S.casefold() -> str
|
|     Return a version of S suitable for caseless comparisons.
|
| center(...)
|     S.center(width[, fillchar]) -> str
|
|     Return S centered in a string of length width. Padding is
|     done using the specified fill character (default is a space)
|
| count(...)
|     S.count(sub[, start[, end]]) -> int
|
|     Return the number of non-overlapping occurrences of substring sub in
|     string S[start:end]. Optional arguments start and end are
|     interpreted as in slice notation.
|
| encode(...)
|     S.encode(encoding='utf-8', errors='strict') -> bytes
|
|     Encode S using the codec registered for encoding. Default encoding
|     is 'utf-8'. errors may be given to set a different error
|     handling scheme. Default is 'strict' meaning that encoding errors raise
|     a UnicodeEncodeError. Other possible values are 'ignore', 'replace' and
|     'xmlcharrefreplace' as well as any other name registered with
|     codecs.register_error that can handle UnicodeEncodeErrors.
|
| endswith(...)
|     S.endswith(suffix[, start[, end]]) -> bool
|
|     Return True if S ends with the specified suffix, False otherwise.
|     With optional start, test S beginning at that position.
|     With optional end, stop comparing S at that position.
|     suffix can also be a tuple of strings to try.
|
| expandtabs(...)
|     S.expandtabs(tabsize=8) -> str
|
|     Return a copy of S where all tab characters are expanded using spaces.
|     If tabsize is not given, a tab size of 8 characters is assumed.
|
| find(...)
|     S.find(sub[, start[, end]]) -> int
|
|     Return the lowest index in S where substring sub is found,
|     such that sub is contained within S[start:end]. Optional
|     arguments start and end are interpreted as in slice notation.
|
|     Return -1 on failure.
```

```
| format(...)
|     S.format(*args, **kwargs) -> str
|
|     Return a formatted version of S, using substitutions from args and kwargs.
|     The substitutions are identified by braces ('{' and '}').
|
| format_map(...)
|     S.format_map(mapping) -> str
|
|     Return a formatted version of S, using substitutions from mapping.
|     The substitutions are identified by braces ('{' and '}').
|
| index(...)
|     S.index(sub[, start[, end]]) -> int
|
|     Return the lowest index in S where substring sub is found,
|     such that sub is contained within S[start:end]. Optional
|     arguments start and end are interpreted as in slice notation.
|
|     Raises ValueError when the substring is not found.
|
| isalnum(...)
|     S.isalnum() -> bool
|
|     Return True if all characters in S are alphanumeric
|     and there is at least one character in S, False otherwise.
|
| isalpha(...)
|     S.isalpha() -> bool
|
|     Return True if all characters in S are alphabetic
|     and there is at least one character in S, False otherwise.
|
| isdecimal(...)
|     S.isdecimal() -> bool
|
|     Return True if there are only decimal characters in S,
|     False otherwise.
|
| isdigit(...)
|     S.isdigit() -> bool
|
|     Return True if all characters in S are digits
|     and there is at least one character in S, False otherwise.
|
| isidentifier(...)
|     S.isidentifier() -> bool
|
|     Return True if S is a valid identifier according
|     to the language definition.
```

```
|      Use keyword.iskeyword() to test for reserved identifiers
|      such as "def" and "class".
|
| islower(...)
|     S.islower() -> bool
|
|      Return True if all cased characters in S are lowercase and there is
|      at least one cased character in S, False otherwise.
|
| isnumeric(...)
|     S.isnumeric() -> bool
|
|      Return True if there are only numeric characters in S,
|      False otherwise.
|
| isprintable(...)
|     S.isprintable() -> bool
|
|      Return True if all characters in S are considered
|      printable in repr() or S is empty, False otherwise.
|
| isspace(...)
|     S.isspace() -> bool
|
|      Return True if all characters in S are whitespace
|      and there is at least one character in S, False otherwise.
|
| istitle(...)
|     S.istitle() -> bool
|
|      Return True if S is a titlecased string and there is at least one
|      character in S, i.e. upper- and titlecase characters may only
|      follow uncased characters and lowercase characters only cased ones.
|      Return False otherwise.
|
| isupper(...)
|     S.isupper() -> bool
|
|      Return True if all cased characters in S are uppercase and there is
|      at least one cased character in S, False otherwise.
|
| join(...)
|     S.join(iterable) -> str
|
|      Return a string which is the concatenation of the strings in the
|      iterable. The separator between elements is S.
|
| ljust(...)
|     S.ljust(width[, fillchar]) -> str
|
|      Return S left-justified in a Unicode string of length width. Padding is
```

```
|     done using the specified fill character (default is a space).
|
| lower(...)
|     S.lower() -> str
|
|     Return a copy of the string S converted to lowercase.
|
| lstrip(...)
|     S.lstrip([chars]) -> str
|
|     Return a copy of the string S with leading whitespace removed.
|     If chars is given and not None, remove characters in chars instead.
|
| partition(...)
|     S.partition(sep) -> (head, sep, tail)
|
|     Search for the separator sep in S, and return the part before it,
|     the separator itself, and the part after it. If the separator is not
|     found, return S and two empty strings.
|
| replace(...)
|     S.replace(old, new[, count]) -> str
|
|     Return a copy of S with all occurrences of substring
|     old replaced by new. If the optional argument count is
|     given, only the first count occurrences are replaced.
|
| rfind(...)
|     S.rfind(sub[, start[, end]]) -> int
|
|     Return the highest index in S where substring sub is found,
|     such that sub is contained within S[start:end]. Optional
|     arguments start and end are interpreted as in slice notation.
|
|     Return -1 on failure.
|
| rindex(...)
|     S.rindex(sub[, start[, end]]) -> int
|
|     Return the highest index in S where substring sub is found,
|     such that sub is contained within S[start:end]. Optional
|     arguments start and end are interpreted as in slice notation.
|
|     Raises ValueError when the substring is not found.
|
| rjust(...)
|     S.rjust(width[, fillchar]) -> str
|
|     Return S right-justified in a string of length width. Padding is
|     done using the specified fill character (default is a space).
```

```
| rpartition(...)
|     S.rpartition(sep) -> (head, sep, tail)
|
|     Search for the separator sep in S, starting at the end of S, and return
|     the part before it, the separator itself, and the part after it. If the
|     separator is not found, return two empty strings and S.
|
| rsplit(...)
|     S.rsplit(sep=None, maxsplit=-1) -> list of strings
|
|     Return a list of the words in S, using sep as the
|     delimiter string, starting at the end of the string and
|     working to the front. If maxsplit is given, at most maxsplit
|     splits are done. If sep is not specified, any whitespace string
|     is a separator.
|
| rstrip(...)
|     S.rstrip([chars]) -> str
|
|     Return a copy of the string S with trailing whitespace removed.
|     If chars is given and not None, remove characters in chars instead.
|
| split(...)
|     S.split(sep=None, maxsplit=-1) -> list of strings
|
|     Return a list of the words in S, using sep as the
|     delimiter string. If maxsplit is given, at most maxsplit
|     splits are done. If sep is not specified or is None, any
|     whitespace string is a separator and empty strings are
|     removed from the result.
|
| splitlines(...)
|     S.splitlines([keepends]) -> list of strings
|
|     Return a list of the lines in S, breaking at line boundaries.
|     Line breaks are not included in the resulting list unless keepends
|     is given and true.
|
| startswith(...)
|     S.startswith(prefix[, start[, end]]) -> bool
|
|     Return True if S starts with the specified prefix, False otherwise.
|     With optional start, test S beginning at that position.
|     With optional end, stop comparing S at that position.
|     prefix can also be a tuple of strings to try.
|
| strip(...)
|     S.strip([chars]) -> str
|
|     Return a copy of the string S with leading and trailing
|     whitespace removed.
```

```

|         If chars is given and not None, remove characters in chars instead.
|
| swapcase(...)
|     S.swapcase() -> str
|
|     Return a copy of S with uppercase characters converted to lowercase
|     and vice versa.
|
| title(...)
|     S.title() -> str
|
|     Return a titlecased version of S, i.e. words start with title case
|     characters, all remaining cased characters have lower case.
|
| translate(...)
|     S.translate(table) -> str
|
|     Return a copy of the string S in which each character has been mapped
|     through the given translation table. The table must implement
|     lookup/indexing via __getitem__, for instance a dictionary or list,
|     mapping Unicode ordinals to Unicode ordinals, strings, or None. If
|     this operation raises LookupError, the character is left untouched.
|     Characters mapped to None are deleted.
|
| upper(...)
|     S.upper() -> str
|
|     Return a copy of S converted to uppercase.
|
| zfill(...)
|     S.zfill(width) -> str
|
|     Pad a numeric string S with zeros on the left, to fill a field
|     of the specified width. The string S is never truncated.
|
| -----
| Static methods defined here:
|
| maketrans(x, y=None, z=None, /)
|     Return a translation table usable for str.translate().
|
|     If there is only one argument, it must be a dictionary mapping Unicode
|     ordinals (integers) or characters to Unicode ordinals, strings or None.
|     Character keys will be then converted to ordinals.
|     If there are two arguments, they must be strings of equal length, and
|     in the resulting dictionary, each character in x will be mapped to the
|     character at the same position in y. If there is a third argument, it
|     must be a string, whose characters will be mapped to None in the result.

```

Nous allons tenter ici de citer les méthodes les plus utilisées. Nous n'avons le temps que de les utiliser de manière très simple, mais bien souvent il est possible de passer en argument des options permettant de ne travailler que sur une sous-chaîne, ou sur la première ou dernière occurrence d'une sous-chaîne. Nous vous renvoyons à la documentation pour obtenir toutes les précisions utiles.

### Découpage - assemblage : `split` et `join`

Les méthodes `split` et `join` permettent de découper une chaîne selon un séparateur pour obtenir une liste, et à l'inverse de reconstruire une chaîne à partir d'une liste.

`split` permet donc de découper :

```
In [2]: 'abc==def==ghi==jkl'.split('==')
```

```
Out[2]: ['abc', 'def', 'ghi', 'jkl']
```

Et à l'inverse :

```
In [3]: "===".join(['abc', 'def', 'ghi', 'jkl'])
```

```
Out[3]: 'abc==def==ghi==jkl'
```

Attention toutefois si le séparateur est un terminateur, la liste résultat contient alors une dernière chaîne vide. En pratique, on utilisera la méthode `strip`, que nous allons voir ci-dessous, avant la méthode `split` pour éviter ce problème.

```
In [4]: 'abc;def;ghi;jkl;'.split(';')
```

```
Out[4]: ['abc', 'def', 'ghi', 'jkl', '']
```

Qui s'inverse correctement cependant :

```
In [5]: ";".join(['abc', 'def', 'ghi', 'jkl', ''])
```

```
Out[5]: 'abc;def;ghi;jkl;'
```

### Remplacement : `replace`

`replace` est très pratique pour remplacer une sous-chaîne par une autre, avec une limite éventuelle sur le nombre de remplacements :

```
In [6]: "abcdefabcdefabcdef".replace("abc", "zoo")
```

```
Out[6]: 'zoodefzoodefzoodef'
```

```
In [7]: "abcdefabcdefabcdef".replace("abc", "zoo", 2)
```

```
Out[7]: 'zoodefzoodefabcdef'
```

Plusieurs appels à `replace` peuvent être chaînés comme ceci :

```
In [8]: "les [x] qui disent [y]".replace("[x]", "chevaliers").replace("[y]", "Ni")
```

```
Out[8]: 'les chevaliers qui disent Ni'
```

**Nettoyage : strip**

On pourrait par exemple utiliser `replace` pour enlever les espaces dans une chaîne, ce qui peut être utile pour “nettoyer” comme ceci :

```
In [9]: " abc:def:ghi ".replace(" ", "")
```

```
Out[9]: 'abc:def:ghi'
```

Toutefois bien souvent on préfère utiliser `strip` qui ne s’occupe que du début et de la fin de la chaîne, et gère aussi les tabulations et autres retour à la ligne :

```
In [10]: "\tune chaîne avec des trucs qui dépassent \n".strip()
```

```
Out[10]: 'une chaîne avec des trucs qui dépassent'
```

On peut appliquer `strip` avant `split` pour éviter le problème du dernier élément vide :

```
In [11]: 'abc;def;ghi;jkl;'.strip(';').split(';')
```

```
Out[11]: ['abc', 'def', 'ghi', 'jkl']
```

**Rechercher une sous-chaîne**

Plusieurs outils permettent de chercher une sous-chaîne. Il existe `find` qui renvoie le plus petit index où on trouve la sous-chaîne :

```
In [12]: # l'indice du début de la première occurrence
         "abcdefcdefghefghijkl".find("def")
```

```
Out[12]: 3
```

```
In [13]: # ou -1 si la chaîne n'est pas présente
         "abcdefcdefghefghijkl".find("zoo")
```

```
Out[13]: -1
```

`rfind` fonctionne comme `find` mais en partant de la fin de la chaîne :

```
In [14]: # en partant de la fin
         "abcdefcdefghefghijkl".rfind("fgh")
```

```
Out[14]: 13
```

```
In [15]: # notez que le résultat correspond
         # tout de même toujours au début de la chaîne
         "abcdefcdefghefghijkl"[13]
```

```
Out[15]: 'f'
```

La méthode `index` se comporte comme `find`, mais en cas d’absence elle lève une **exception** (nous verrons ce concept plus tard) plutôt que de renvoyer `-1` :

```
In [16]: "abcdefcdefghefghijkl".index("def")
```

```
Out[16]: 3
```



```
In [17]: try:
          "abcdefcdefghefghijk".index("zoo")
        except Exception as e:
            print("OOPS", type(e), e)

OOPS <class 'ValueError'> substring not found
```

Mais le plus simple pour chercher si une sous-chaîne est dans une autre chaîne est d'utiliser l'instruction `in` sur laquelle nous reviendrons lorsque nous parlerons des séquences :

```
In [18]: "def" in "abcdefcdefghefghijk"

Out[18]: True
```

La méthode `count` compte le nombre d'occurrences d'une sous-chaîne :

```
In [19]: "abcdefcdefghefghijk".count("ef")

Out[19]: 3
```

Signalons enfin les méthodes de commodité suivantes :

```
In [20]: "abcdefcdefghefghijk".startswith("abcd")

Out[20]: True

In [21]: "abcdefcdefghefghijk".endswith("ghijk")

Out[21]: True
```

S'agissant des deux dernières, remarquons que :

```
chaîne.startswith(sous_chaîne)  $\iff$  chaîne.find(sous_chaîne) == 0
chaîne.endswith(sous_chaîne)  $\iff$  chaîne.rfind(sous_chaîne) == (len(chaine) - len(sous_chaîne))
```

On remarque ici la supériorité en terme d'expressivité des méthodes pythoniques `startswith` et `endswith`.

### Changement de casse

Voici pour conclure quelques méthodes utiles qui parlent d'elles-mêmes :

```
In [22]: "monty PYTHON".upper()

Out[22]: 'MONTY PYTHON'

In [23]: "monty PYTHON".lower()

Out[23]: 'monty python'

In [24]: "monty PYTHON".swapcase()

Out[24]: 'MONTY python'

In [25]: "monty PYTHON".capitalize()

Out[25]: 'Monty python'

In [26]: "monty PYTHON".title()

Out[26]: 'Monty Python'
```

### Pour en savoir plus

Tous ces outils sont [documentés en détail ici \(en anglais\)](#).

## 2.3 Formatage de chaînes de caractères

### 2.3.1 Complément - niveau basique

On désigne par formatage les outils qui permettent d'obtenir une présentation fine des résultats, que ce soit pour améliorer la lisibilité lorsqu'on s'adresse à des humains, ou pour respecter la syntaxe d'un outil auquel on veut passer les données pour un traitement ultérieur.

#### La fonction `print`

Nous avons jusqu'à maintenant presque toujours utilisé la fonction `print` pour afficher nos résultats. Comme on l'a vu, celle-ci réalise un formatage sommaire : elle insère une espace entre les valeurs qui lui sont passées.

```
In [1]: print(1, 'a', 12 + 4j)

1 a (12+4j)
```

La seule subtilité notable concernant `print` est que, par défaut, elle ajoute un saut de ligne à la fin. Pour éviter ce comportement, on peut passer à la fonction un argument `end`, qui sera inséré *au lieu* du saut de ligne. Ainsi par exemple :

```
In [2]: # une première ligne
        print("une", "seule", "ligne")

une seule ligne
```

```
In [3]: # une deuxième ligne en deux appels à print
        print("une", "autre", end=' ')
        print("ligne")

une autre ligne
```

Il faut remarquer aussi que `print` est capable d'imprimer **n'importe quel objet**. Nous l'avons déjà fait avec les listes et les tuples, voici par exemple un module :

```
In [4]: # on peut imprimer par exemple un objet 'module'
        import math

        print('le module math est', math)

le module math est <module 'math' (built-in)>
```

En anticipant un peu, voici comment `print` présente les instances de classe (ne vous inquiétez pas, nous apprendrons dans une semaine ultérieure ce que sont les classes et les instances).

```
In [5]: # pour définir la classe Personne
        class Personne:
            pass

        # et pour créer une instance de cette classe
        personne = Personne()
```

```
In [6]: # voilà comment s'affiche une instance de classe
        print(personne)

<__main__.Personne object at 0x0502DB30>
```

On rencontre assez vite les limites de `print` :

- d’une part, il peut être nécessaire de formater une chaîne de caractères sans nécessairement vouloir l’imprimer, ou en tout cas pas immédiatement ;
- d’autre part, les espaces ajoutées peuvent être plus néfastes qu’utiles ;
- enfin, on peut avoir besoin de préciser un nombre de chiffres significatifs, ou de choisir comment présenter une date.

C’est pourquoi il est plus courant de **formater** les chaînes - c’est-à-dire de calculer des chaînes en mémoire, sans nécessairement les imprimer de suite, et c’est ce que nous allons étudier dans ce complément.

### Les *f-strings*

Depuis la version 3.6 de Python, on peut utiliser les *f-strings*, le premier mécanisme de formatage que nous étudierons. C’est le mécanisme de formatage le plus simple et le plus agréable à utiliser.

Je vous recommande tout de même de lire les sections à propos de `format` et de `%`, qui sont encore massivement utilisées dans le code existant (surtout `%` d’ailleurs, bien que essentiellement obsolète).

Mais définissons d’abord quelques données à afficher :

```
In [7]: # donnons-nous quelques variables
        prenom, nom, age = 'Jean', 'Dupont', 35

In [8]: # mon premier f-string
        f"{prenom} {nom} a {age} ans"

Out[8]: 'Jean Dupont a 35 ans'
```

Vous remarquez d’abord que le string commence par `f"`, c’est bien sûr pour cela qu’on l’appelle un *f-string*.

On peut bien entendu ajouter le `f` devant toutes les formes de strings, qu’ils commencent par `'` ou `"` ou `'''` ou `"""`.

Ensuite vous remarquez que les zones délimitées entre `{}` sont remplacées. La logique d’un *f-string*, c’est tout simplement de considérer l’intérieur d’un `{}` comme du code Python (une expression pour être précis), de l’évaluer, et d’utiliser le résultat pour remplir le `{}`.

Ça veut dire, en clair, que je peux faire des calculs à l’intérieur des `{}`.

```
In [9]: # toutes les expressions sont autorisées à l'intérieur d'un {}
        f"dans 10 ans {prenom} aura {age + 10} ans"

Out[9]: 'dans 10 ans Jean aura 45 ans'
```

```
In [10]: # on peut donc aussi mettre des appels de fonction
        notes = [12, 15, 19]
        f"nous avons pour l'instant {len(notes)} notes"
```

```
Out[10]: "nous avons pour l'instant 3 notes"
```

Nous allons en rester là pour la partie en niveau basique. Il nous reste à étudier comment chaque {} est formaté (par exemple comment choisir le nombre de chiffres significatifs sur un flottant), voyez plus bas pour plus de détails sur ce point.

Comme vous le voyez, les *f-strings* fournissent une méthode très simple et expressive pour formater des données dans des chaînes de caractère. Redisons-le pour être bien clair : un *f-string* **ne réalise pas d'impression**, il faut donc le passer à `print` si l'impression est souhaitée.

### La méthode `format`

Avant l'introduction des *f-strings*, la technique recommandée pour faire du formatage était d'utiliser la méthode `format` qui est définie sur les objets `str` et qui s'utilise comme ceci :

```
In [11]: "{} {} a {} ans".format(prenom, nom, age)
```

```
Out[11]: 'Jean Dupont a 35 ans'
```

Dans cet exemple le plus simple, les données sont affichées en lieu et place des {}, dans l'ordre où elles sont fournies.

Cela convient bien lorsqu'on a peu de données. Si par la suite on veut changer l'ordre par exemple des nom et prénom, on peut bien sûr échanger l'ordre des arguments passés à `format`, ou encore utiliser la **liaison par position**, comme ceci :

```
In [12]: "{1} {0} a {2} ans".format(prenom, nom, age)
```

```
Out[12]: 'Dupont Jean a 35 ans'
```

Dans la pratique toutefois, cette forme est assez peu utile, on lui préfère souvent la **liaison par nom** qui se présente comme ceci :

```
In [13]: "{le_prenom} {le_nom} a {l_age} ans".format(le_nom=nom, le_prenom=prenom, l_age=age)
```

```
Out[13]: 'Jean Dupont a 35 ans'
```

Dans ce premier exemple de liaison par nom, nous avons délibérément utilisé des noms différents pour les données externes et pour les noms apparaissant dans le format, pour bien illustrer comment la liaison est résolue, mais on peut aussi bien faire tout simplement :

```
In [14]: "{prenom} {nom} a {age} ans".format(nom=nom, prenom=prenom, age=age)
```

```
Out[14]: 'Jean Dupont a 35 ans'
```

Voici qui conclut notre courte introduction à la méthode `format`.

### 2.3.2 Complément - niveau intermédiaire

#### La toute première version du formatage : l'opérateur %

format a été en fait introduite assez tard dans Python, pour remplacer la technique que nous allons présenter maintenant.

Étant donné le volume de code qui a été écrit avec l'opérateur %, il nous a semblé important d'introduire brièvement cette construction ici. Vous ne devez cependant pas utiliser cet opérateur dans du code moderne, la manière pythonique de formater les chaînes de caractères est le f-string.

Le principe de l'opérateur % est le suivant. On élabore comme ci-dessus un "format" c'est-à-dire le patron de ce qui doit être rendu, auquel on passe des arguments pour "remplir" les trous. Voyons les exemples de tout à l'heure rendus avec l'opérateur % :

```
In [15]: # l'ancienne façon de formater les chaînes avec %
        # est souvent moins lisible
        "%s %s a %s ans" % (prenom, nom, age)
```

```
Out[15]: 'Jean Dupont a 35 ans'
```

On pouvait également avec cet opérateur recourir à un mécanisme de liaison par nommage, en passant par un dictionnaire. Pour anticiper un tout petit peu sur cette notion que nous verrons très bientôt, voici comment

```
In [16]: variables = {'le_nom': nom, 'le_prenom': prenom, 'l_age': age}
        "%(le_nom)s, %(le_prenom)s, %(l_age)s ans" % variables
```

```
Out[16]: 'Dupont, Jean, 35 ans'
```

### 2.3.3 Complément - niveau avancé

De retour aux *f-strings* et à la fonction `format`, il arrive qu'on ait besoin de spécifier plus finement la façon dont une valeur doit être affichée.

#### Précision des arrondis

C'est typiquement le cas avec les valeurs flottantes pour lesquelles la précision de l'affichage vient au détriment de la lisibilité. Voici deux formes équivalentes pour obtenir une valeur de pi arrondie :

```
In [17]: from math import pi
```

```
In [18]: # un f-string
        f"pi avec seulement 2 chiffres apres la virgule {pi:.2f}"
```

```
Out[18]: 'pi avec seulement 2 chiffres apres la virgule 3.14'
```

```
In [19]: # avec format() et liaison par nom
        "pi avec seulement 2 chiffres apres la virgule {flottant:.2f}".format(flottant=pi)
```

```
Out[19]: 'pi avec seulement 2 chiffres apres la virgule 3.14'
```

Dans ces deux exemples, la partie à l'intérieur des {} et à droite du : s'appelle le format, ici `.2f` ; vous remarquez que c'est le même pour les *f-strings* et pour `format`, et c'est toujours le cas. C'est pourquoi on ne verra plus à partir d'ici que des exemples avec les *f-strings*.

### 0 en début de nombre

Pour forcer un petit entier à s’afficher sur 4 caractères, avec des 0 ajoutés au début si nécessaire :

```
In [20]: x = 15
```

```
f"{x:04d}"
```

```
Out[20]: '0015'
```

Ici on utilise le format d (toutes ces lettres d, f, g viennent des formats ancestraux de la libc comme printf). Ici avec 04d on précise qu’on veut une sortie sur 4 caractères et qu’il faut remplir à gauche si nécessaire avec des 0.

### Largeur fixe

Dans certains cas, on a besoin d’afficher des données en colonnes de largeur fixe, on utilise pour cela les formats < ^ et > pour afficher à gauche, au centre, ou à droite d’une zone de largeur fixe :

```
In [21]: # les données à afficher
```

```
comptes = [
    ('Apollin', 'Dupont', 127),
    ('Myrtille', 'Lamartine', 25432),
    ('Prune', 'Soc', 827465),
]
```

```
for prenom, nom, solde in comptes:
    print(f"{prenom:<10} -- {nom:^12} -- {solde:>8} €")
```

```
Apollin    --      Dupont    --      127 €
Myrtille   --   Lamartine   --    25432 €
Prune      --        Soc     --   827465 €
```

### Voir aussi

Nous vous invitons à vous reporter à la documentation de format pour plus de détails [sur les formats disponibles](#), et notamment aux [nombreux exemples](#) qui y figurent.

## 2.4 Obtenir une réponse de l'utilisateur

### 2.4.1 Complément - niveau basique

Occasionnellement, il peut être utile de poser une question à l'utilisateur.

#### La fonction `input`

C'est le propos de la fonction `input`. Par exemple :

```
In [1]: nom_ville = input("Entrez le nom de la ville : ")
        print(f"nom_ville={nom_ville}")
```

```
Entrez le nom de la ville : Metropolis
nom_ville=Metropolis
```

#### Attention à bien vérifier/convertir

Notez bien que `input` renvoie **toujours une chaîne de caractère** (`str`). C'est assez évident, mais il est très facile de l'oublier et de passer cette chaîne directement à une fonction qui s'attend à recevoir, par exemple, un nombre entier, auquel cas les choses se passent mal :

```
>>> input("nombre de lignes ? ") + 3
nombre de lignes ? 12
Traceback (most recent call last):
  File "<stdin>", line 1, in <module>
TypeError: must be str, not int
```

Dans ce cas il faut appeler la fonction `int` pour convertir le résultat en un entier :

```
In [2]: int(input("Nombre de lignes ? ")) + 3
```

```
Nombre de lignes ? 5
```

```
Out[2]: 8
```

#### Limitations

Cette fonction peut être utile pour vos premiers pas en Python.

En pratique toutefois, on utilise assez peu cette fonction, car les applications "réelles" viennent avec leur propre interface utilisateur, souvent graphique, et disposent donc d'autres moyens que celui-ci pour interagir avec l'utilisateur.

Les applications destinées à fonctionner dans un terminal, quant à elles, reçoivent traditionnellement leurs données de la ligne de commande. C'est le propos du module `argparse` que nous avons déjà rencontré en première semaine.

## 2.5 Expressions régulières et le module `re`

### 2.5.1 Complément - niveau basique

#### Avertissement

Après avoir joué ce cours plusieurs années de suite, l'expérience nous montre qu'il est difficile de trouver le bon moment pour appréhender les expressions régulières.

D'un côté il s'agit de manipulations de chaînes de caractères, mais d'un autre cela nécessite de créer des instances de classes, et donc d'avoir vu la programmation orientée objet. Du coup, les premières années nous les avons étudiées tout à la fin du cours, ce qui avait pu créer une certaine frustration.

C'est pourquoi nous avons décidé à présent de les étudier très tôt, dans cette séquence consacrée aux chaînes de caractères. Les étudiants qui seraient décontenancés par ce contenu sont invités à y retourner après la semaine 6, consacrée à la programmation objet.

Il nous semble important de savoir que ces fonctionnalités existent dans le langage, le détail de leur utilisation n'est toutefois pas critique, et on peut parfaitement faire l'impasse sur ce complément en première lecture.

Une expression régulière est un objet mathématique permettant de décrire un ensemble de textes qui possèdent des propriétés communes. Par exemple, s'il vous arrive d'utiliser un terminal, et que vous tapez

```
$ dir *.txt
```

(ou `ls *.txt` sur linux ou mac), vous utilisez l'expression régulière `*.txt` qui désigne tous les fichiers dont le nom se termine par `.txt`. On dit que l'expression régulière *filtre* toutes les chaînes qui se terminent par `.txt` (l'expression anglaise consacrée est le *pattern matching*).

Le langage Perl a été le premier à populariser l'utilisation des expressions régulières en les supportant nativement dans le langage, et non au travers d'une librairie. En python, les expressions régulières sont disponibles de manière plus traditionnelle, via le module `re` (regular expressions) de la librairie standard. Le propos de ce complément est de vous en donner une première introduction.

```
In [1]: import re
```

#### Survol

Pour ceux qui ne souhaitent pas approfondir, voici un premier exemple ; on cherche à savoir si un objet chaîne est ou non de la forme `*-*.txt`, et si oui, à calculer la partie de la chaîne qui remplace le `*` :

```
In [2]: # un objet 'expression régulière' - on dit aussi "pattern"
        regexp = "(.*)-(.*)\.txt"
```

```
In [3]: # la chaîne de départ
        chaine = "abcdef.txt"
```

```
In [4]: # la fonction qui calcule si la chaîne "matche" le pattern
        match = re.match(regexp, chaine)
        match is None
```



```
Out [4]: True
```

Le fait que l'objet `match` vaut `None` indique que la chaîne n'est pas de la bonne forme (il manque un `-` dans le nom); avec une autre chaîne par contre :

```
In [5]: # la chaîne de départ
        chaîne = "abc-def.txt"
```

```
In [6]: match = re.match(regex, chaîne)
        match is None
```

```
Out [6]: False
```

Ici `match` est un objet, qui nous permet ensuite d'extraire les différentes parties, comme ceci :

```
In [7]: match[1]
```

```
Out [7]: 'abc'
```

```
In [8]: match[2]
```

```
Out [8]: 'def'
```

Bien sûr on peut faire des choses beaucoup plus élaborées avec `re`, mais en première lecture cette introduction doit vous suffire pour avoir une idée de ce qu'on peut faire avec les expressions régulières.

## 2.5.2 Complément - niveau intermédiaire

Approfondissons à présent :

Dans un terminal, `*.txt` est une expression régulière très simple. Le module `re` fournit le moyen de construire des expressions régulières très élaborées et plus puissantes que ce que supporte le terminal. C'est pourquoi la syntaxe des regexps de `re` est un peu différente. Par exemple comme on vient de le voir, pour filtrer la même famille de chaînes que `*-*.txt` avec le module `re`, il nous a fallu écrire l'expression régulière sous une forme légèrement différente.

Je vous conseille d'avoir sous la main la [documentation du module `re`](#) pendant que vous lisez ce complément.

### Avertissement

Dans ce complément nous serons amenés à utiliser des traits qui dépendent du `LOCALE`, c'est-à-dire, pour faire simple, de la configuration de l'ordinateur vis-à-vis de la langue.

Tant que vous exécutez ceci dans le notebook sur la plateforme, en principe tout le monde verra exactement la même chose. Par contre, si vous faites tourner le même code sur votre ordinateur, il se peut que vous obteniez des résultats légèrement différents.

## Un exemple simple

findall On se donne deux exemples de chaînes

```
In [9]: sentences = ['Lacus a donec, vitae gravida proin sociis.',
                    'Neque ipsum! rhoncus cras quam.']
```

On peut **chercher tous** les mots se terminant par a ou m dans une chaîne avec `findall`

```
In [10]: for sentence in sentences:
          print(f"---- dans >{sentence}<")
          print(re.findall(r"\w*[am]\W", sentence))

---- dans >Lacus a donec, vitae gravida proin sociis.<
['a ', 'gravida ']
---- dans >Neque ipsum! rhoncus cras quam.<
['ipsum!', 'quam.']
```

Ce code permet de chercher toutes (`findall`) les occurrences de l'expression régulière, qui ici est définie par le *raw-string*

```
r"\w*[am]\W"
```

Nous verrons tout à l'heure comment fabriquer des expressions régulières plus en détail, mais pour démystifier au moins celle-ci, on a mis bout à bout les morceaux suivants.

- `\w*` : on veut trouver une sous-chaîne qui commence par un nombre quelconque, y compris nul (\*) de caractères alphanumériques (`\w`). Ceci est défini en fonction de votre LOCALE, on y reviendra.
- `[am]` : immédiatement après, il nous faut trouver un caractère a ou m.
- `\W` : et enfin, il nous faut un caractère qui ne soit **pas** alphanumérique. Ceci est important puisqu'on cherche les mots qui **se terminent** par un a ou un m, si on ne le mettait pas on obtiendrait ceci

```
In [11]: # le \W final est important
          # voici ce qu'on obtient si on l'omet
          for sentence in sentences:
              print(f"---- dans >{sentence}<")
              print(re.findall(r"\w*[am]", sentence))

---- dans >Lacus a donec, vitae gravida proin sociis.<
['La', 'a', 'vita', 'gravida']
---- dans >Neque ipsum! rhoncus cras quam.<
['ipsum', 'cra', 'quam']
```

`split`

Une autre forme simple d'utilisation des regexps est `re.split`, qui fournit une fonctionnalité voisine de `str.split`, mais où les séparateurs sont exprimés comme une expression régulière

```
In [12]: for sentence in sentences:
          print(f"---- dans >{sentence}<")
          print(re.split(r"\W+", sentence))
          print()

---- dans >Lacus a donec, vitae grvida proin sociis.<
['Lacus', 'a', 'donec', 'vitae', 'grvida', 'proin', 'sociis', '']

---- dans >Neque ipsum! rhoncus cras quam.<
['Neque', 'ipsum', 'rhoncus', 'cras', 'quam', '']
```

Ici l'expression régulière, qui bien sûr décrit le séparateur, est simplement `\W+` c'est-à-dire toute suite d'au moins un caractère non alphanumérique.

Nous avons donc là un moyen simple, et plus puissant que `str.split`, de couper un texte en mots.

sub

Une troisième méthode utilitaire est `re.sub` qui permet de remplacer les occurrences d'une *regex*, comme par exemple

```
In [13]: for sentence in sentences:
          print(f"---- dans >{sentence}<")
          print(re.sub(r"(\w+)", r"X\1Y", sentence))
          print()

---- dans >Lacus a donec, vitae grvida proin sociis.<
XLacusY XaY XdonecY, XvitaeY XgrvidaY XproinY XsociisY.

---- dans >Neque ipsum! rhoncus cras quam.<
XNequeY XipsumY! XrhoncusY XcrasY XquamY.
```

Ici, l'expression régulière (le premier argument) contient un **groupe** : on a utilisé des parenthèses autour du `\w+`. Le second argument est la chaîne de remplacement, dans laquelle on a fait **référence au groupe** en écrivant `\1`, qui veut dire tout simplement "le premier groupe".

Donc au final, l'effet de cet appel est d'entourer toutes les suites de caractères alphanumériques par X et Y.

### Pourquoi un *raw-string* ?

En guise de digression, il n'y a aucune obligation à utiliser un *raw-string*, d'ailleurs on rappelle qu'il n'y a pas de différence de nature entre un *raw-string* et une chaîne usuelle

```
In [14]: raw = r'abc'
          regular = 'abc'
```

```
# comme on a pris une 'petite' chaîne ce sont les mêmes objets
print(f"both compared with is → {raw is regular}")
# et donc a fortiori
print(f"both compared with == → {raw == regular}")
```

both compared with is → True

both compared with == → True

Il se trouve que le *backslash* \ à l'intérieur des expressions régulières est d'un usage assez courant - on l'a vu déjà plusieurs fois. C'est pourquoi on **utilise fréquemment un *raw-string*** pour décrire une expression régulière, et en général à chaque fois qu'elle comporte un *backslash*. On rappelle que le *raw-string* désactive l'interprétation des \ à l'intérieur de la chaîne, par exemple, \t est interprété comme un caractère de tabulation. Sans *raw-string*, il faut doubler tous les \ pour qu'il n'y ait pas d'interprétation.

### Un deuxième exemple

Nous allons maintenant voir comment on peut d'abord vérifier si une chaîne est conforme au critère défini par l'expression régulière, mais aussi *extraire* les morceaux de la chaîne qui correspondent aux différentes parties de l'expression.

Pour cela, supposons qu'on s'intéresse aux chaînes qui comportent 5 parties, une suite de chiffres, une suite de lettres, des chiffres à nouveau, des lettres et enfin de nouveau des chiffres.

Pour cela on considère ces trois chaînes en entrée

```
In [15]: samples = ['890hj000nnm890',    # cette entrée convient
                    '123abc456def789',    # celle-ci aussi
                    '8090abababab879',    # celle-ci non
                    ]
```

match Pour commencer, voyons que l'on peut facilement **vérifier si une chaîne vérifie** ou non le critère.

```
In [16]: regexp1 = "[0-9]+[A-Za-z]+[0-9]+[A-Za-z]+[0-9]+"
```

Si on applique cette expression régulière à toutes nos entrées

```
In [17]: for sample in samples:
          match = re.match(regexp1, sample)
          print(f"{sample:16s} → {match}")

890hj000nnm890    → <_sre.SRE_Match object; span=(0, 14), match='890hj000nnm890'>
123abc456def789   → <_sre.SRE_Match object; span=(0, 15), match='123abc456def789'>
8090abababab879   → None
```

Pour rendre ce résultat un peu plus lisible nous nous définissons une petite fonction de confort.

```
In [18]: # pour simplement visualiser si on a un match ou pas
def nice(match):
    # le retour de re.match est soit None, soit un objet match
    return "no" if match is None else "Match!"
```

Avec quoi on peut refaire l'essai sur toutes nos entrées.

```
In [19]: # la même chose mais un peu moins encombrant
print(f"REGEXP={regex1}\n")
for sample in samples:
    match = re.match(regex1, sample)
    print(f"{sample:>16s} → {nice(match)}")
```

```
REGEXP=[0-9]+[A-Za-z]+[0-9]+[A-Za-z]+[0-9]+
```

```
890hj000nnm890 → Match!
123abc456def789 → Match!
8090abababab879 → no
```

Ici plutôt que d'utiliser les raccourcis comme `\w` j'ai préféré écrire explicitement les ensembles de caractères en jeu. De cette façon, on rend son code indépendant du LOCALE si c'est ce qu'on veut faire. Il y a deux morceaux qui interviennent tour à tour :

- `[0-9]+` signifie une suite de au moins un caractère dans l'intervalle `[0-9]`,
- `[A-Za-z]+` pour une suite d'au moins un caractère dans l'intervalle `[A-Z]` ou dans l'intervalle `[a-z]`.

Et comme tout à l'heure on a simplement juxtaposé les morceaux dans le bon ordre pour construire l'expression régulière complète.

### Nommer un morceau (un groupe)

```
In [20]: # on se concentre sur une entrée correcte
haystack = samples[1]
haystack
```

```
Out[20]: '123abc456def789'
```

Maintenant, on va même pouvoir **donner un nom** à un morceau de la regexp, ici on désigne par `needle` le groupe de chiffres du milieu.

```
In [21]: # la même regexp, mais on donne un nom au groupe de chiffres central
regex2 = "[0-9]+[A-Za-z]+(?P<needle>[0-9]+)[A-Za-z]+[0-9]+"
```

Et une fois que c'est fait, on peut demander à l'outil de nous **retrouver la partie correspondante** dans la chaîne initiale :

```
In [22]: print(re.match(regex2, haystack).group('needle'))
```

```
456
```

Dans cette expression on a utilisé un **groupe nommé** `(?P<needle>[0-9]+)`, dans lequel :

- les parenthèses définissent un groupe,
- `?P<needle>` spécifie que ce groupe pourra être référencé sous le nom `needle` (cette syntaxe très absconse est héritée semble-t-il de perl).

### Un troisième exemple

Enfin, et c'est un trait qui n'est pas présent dans tous les langages, on peut restreindre un morceau de chaîne à être identique à un groupe déjà vu plus tôt dans la chaîne. Dans l'exemple ci-dessus, on pourrait ajouter comme contrainte que le premier et le dernier groupes de chiffres soient identiques, comme ceci

```
In [23]: regexp3 = "(?P<id>[0-9]+)[A-Za-z]+(?P<needle>[0-9]+)[A-Za-z]+(?P=id)"
```

Si bien que maintenant, avec les mêmes entrées que tout à l'heure

```
In [24]: print(f"REGEXP={regexp3}\n")
        for sample in samples:
            match = re.match(regexp3, sample)
            print(f"{sample:>16s} → {nice(match)}")
```

```
REGEXP=(?P<id>[0-9]+)[A-Za-z]+(?P<needle>[0-9]+)[A-Za-z]+(?P=id)
```

```
890hj000nnm890 → Match!
123abc456def789 → no
8090abababab879 → no
```

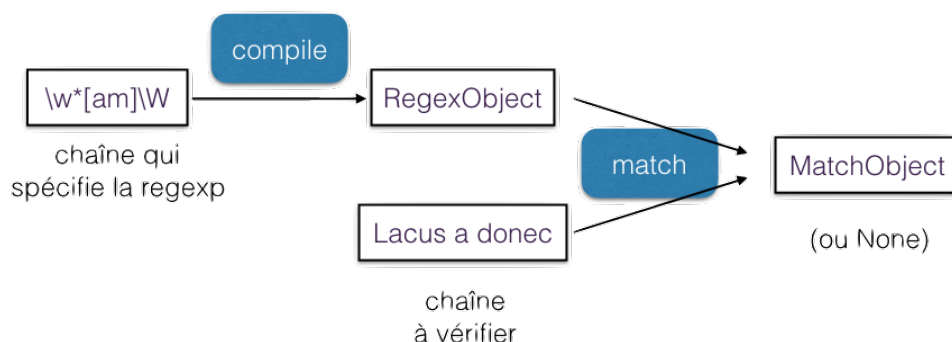
Comme précédemment on a défini le groupe nommé `id` comme étant la première suite de chiffres. La nouveauté ici est la **contrainte** qu'on a imposée sur le dernier groupe avec `(?P=id)`. Comme vous le voyez, on n'obtient un *match* qu'avec les entrées dans lesquelles le dernier groupe de chiffres est identique au premier.

### Comment utiliser la librairie

Avant d'apprendre à écrire une expression régulière, disons quelques mots du mode d'emploi de la librairie.

#### Fonctions de commodité et *workflow*

Comme vous le savez peut-être, une expression régulière décrite sous forme de chaîne, comme par exemple `"\w*[am]\W"`, peut être traduite dans un **automate fini** qui permet de faire le filtrage avec une chaîne. C'est ce qui explique le *workflow* que nous avons résumé dans cette figure.



La méthode recommandée pour utiliser la librairie, lorsque vous avez le même *pattern* à appliquer à un grand nombre de chaînes, est de :

- compiler **une seule fois** votre chaîne en un automate, qui est matérialisé par un objet de la classe `re.RegexObject`, en utilisant `re.compile`,
- puis d'**utiliser directement cet objet** autant de fois que vous avez de chaînes.

Nous avons utilisé dans les exemples plus haut (et nous continuerons plus bas pour une meilleure lisibilité) des **fonctions de commodité** du module, qui sont pratiques, par exemple, pour mettre au point une expression régulière en mode interactif, mais qui ne **sont pas forcément** adaptées dans tous les cas.

Ces fonctions de commodité fonctionnent toutes sur le même principe :

```
re.match(regex, sample)  $\iff$  re.compile(regex).match(sample)
```

Donc à chaque fois qu'on utilise une fonction de commodité, on recompile la chaîne en automate, ce qui, dès qu'on a plus d'une chaîne à traiter, représente un surcoût.

In [25]: # au lieu de faire comme ci-dessus:

```
# imaginez 10**6 chaînes dans samples
for sample in samples:
    match = re.match(regex3, sample)
    print(f"{sample:>16s} → {nice(match)}")

890hj000nnm890 → Match!
123abc456def789 → no
8090abababab879 → no
```

In [26]: # dans du vrai code on fera plutôt:

```
# on compile la chaîne en automate une seule fois
re_obj3 = re.compile(regex3)

# ensuite on part directement de l'automate
for sample in samples:
    match = re_obj3.match(sample)
    print(f"{sample:>16s} → {nice(match)}")

890hj000nnm890 → Match!
123abc456def789 → no
8090abababab879 → no
```

Cette deuxième version ne compile qu'une fois la chaîne en automate, et donc est plus efficace.

**Les méthodes sur la classe `RegexObject`** Les objets de la classe `RegexObject` représentent donc l'automate à état fini qui est le résultat de la compilation de l'expression régulière. Pour résumer ce qu'on a déjà vu, les méthodes les plus utiles sur un objet `RegexObject` sont :

- `match` et `search`, qui cherchent un *match* soit uniquement au début (`match`) ou n'importe où dans la chaîne (`search`),
- `findall` et `split` pour chercher toutes les occurrences (`findall`) ou leur négatif (`split`),
- `sub` (qui aurait pu sans doute s'appeler `replace`, mais c'est comme ça) pour remplacer les occurrences de pattern.

### Exploiter le résultat

Les **méthodes** disponibles sur la classe `re.MatchObject` sont [documentées en détail ici](#). On en a déjà rencontré quelques-unes, en voici à nouveau un aperçu rapide.

```
In [27]: # exemple
        sample = "    Isaac Newton, physicist"
        match = re.search(r"(\w+) (?P<name>\w+)", sample)
```

`re` et `string` pour retrouver les données d'entrée du `match`.

```
In [28]: match.string
```

```
Out[28]: '    Isaac Newton, physicist'
```

```
In [29]: match.re
```

```
Out[29]: re.compile(r'(\w+) (?P<name>\w+)', re.UNICODE)
```

`group`, `groups`, `groupdict` pour retrouver les morceaux de la chaîne d'entrée qui correspondent aux **groupes** de la regexp. On peut y accéder par rang, ou par nom (comme on l'a vu plus haut avec `needle`).

```
In [30]: match.groups()
```

```
Out[30]: ('Isaac', 'Newton')
```

```
In [31]: match.group(1)
```

```
Out[31]: 'Isaac'
```

```
In [32]: match.group('name')
```

```
Out[32]: 'Newton'
```

```
In [33]: match.group(2)
```

```
Out[33]: 'Newton'
```

```
In [34]: match.groupdict()
```

```
Out[34]: {'name': 'Newton'}
```

Comme on le voit pour l'accès par rang **les indices commencent à 1** pour des raisons historiques (on peut déjà référencer `\1` en `sed` depuis la fin des années 70).

On peut aussi accéder au **groupe 0** comme étant la partie de la chaîne de départ qui a effectivement été filtrée par l'expression régulière, et qui peut tout à fait être au beau milieu de la chaîne de départ, comme dans notre exemple

```
In [35]: match.group(0)
```

```
Out[35]: 'Isaac Newton'
```

`expand` permet de faire une espèce de `str.format` avec les valeurs des groupes.

```
In [36]: match.expand(r"last_name \g<name> first_name \1")
```

```
Out[36]: 'last_name Newton first_name Isaac'
```

`span` pour connaître les index dans la chaîne d'entrée pour un groupe donné.

```
In [37]: begin, end = match.span('name')
        sample[begin:end]
```

```
Out[37]: 'Newton'
```



**Les différents modes (flags)** Enfin il faut noter qu'on peut passer à `re.compile` un certain nombre de *flags* qui modifient globalement l'interprétation de la chaîne, et qui peuvent rendre service.

Vous trouverez [une liste exhaustive de ces flags ici](#). Ils ont en général un nom long et parlant, et un alias court sur un seul caractère. Les plus utiles sont sans doute :

- `IGNORECASE` (*alias* `I`) pour, eh bien, ne pas faire la différence entre minuscules et majuscules,
- `UNICODE` (*alias* `U`) pour rendre les séquences `\w` et autres basées sur les propriétés des caractères dans la norme Unicode,
- `LOCALE` (*alias* `L`) cette fois `\w` dépend du `locale` courant,
- `MULTILINE` (*alias* `M`), et
- `DOTALL` (*alias* `S`) pour ces deux flags voir la discussion à la fin du complément.

Comme c'est souvent le cas, on doit passer à `re.compile` un **ou logique** (caractère `|`) des différents flags que l'on veut utiliser, c'est-à-dire qu'on fera par exemple

```
In [38]: regexp = "a*b+"
         re_obj = re.compile(regexp, flags=re.IGNORECASE | re.DEBUG)
```

```
MAX_REPEAT 0 MAXREPEAT
LITERAL 97
MAX_REPEAT 1 MAXREPEAT
LITERAL 98
```

```
In [39]: # on ignore la casse des caractères
         print(regexp, "->", nice(re_obj.match("AabB")))
```

```
a*b+ -> Match!
```

## Comment construire une expression régulière

Nous pouvons à présent voir comment construire une expression régulière, en essayant de rester synthétique (la [documentation du module re](#) en donne une version exhaustive).

### La brique de base : le caractère

Au commencement il faut spécifier des caractères.

- **un seul** caractère :
  - vous le citez tel quel, en le précédant d'un backslash `\` s'il a par ailleurs un sens spécial dans le micro-langage de regexps (comme `+`, `*`, `[`, etc.);
- **l'attrape-tout** (*wildcard*) :
  - un point `.` signifie "n'importe quel caractère";
- **un ensemble** de caractères avec la notation `[...]` qui permet de décrire par exemple :
  - `[a1=]` un ensemble in extenso, ici un caractère parmi `a`, `1`, ou `=`,
  - `[a-z]` un intervalle de caractères, ici de `a` à `z`,
  - `[15e-g]` un mélange des deux, ici un ensemble qui contiendrait `1`, `5`, `e`, `f` et `g`,

- `[^15e-g]` une **négation**, qui a `^` comme premier caractère dans les `[]`, ici tout sauf l'ensemble précédent;
- un **ensemble prédéfini** de caractères, qui peuvent alors dépendre de l'environnement (UNICODE et LOCALE) avec entre autres les notations :
  - `\w` les caractères alphanumériques, et `\W` (les autres),
  - `\s` les caractères "blancs" - espace, tabulation, saut de ligne, etc., et `\S` (les autres),
  - `\d` pour les chiffres, et `\D` (les autres).

```
In [40]: sample = "abcd"
```

```
for regexp in ['abcd', 'ab[cd][cd]', 'ab[a-z]d', r'abc.', r'abc\\.']:
    match = re.match(regexp, sample)
    print(f"{sample} / {regexp:<10s} → {nice(match)}")
```

```
abcd / abcd          → Match!
abcd / ab[cd][cd]    → Match!
abcd / ab[a-z]d      → Match!
abcd / abc.          → Match!
abcd / abc\.         → no
```

Pour ce dernier exemple, comme on a backslashé le `.` il faut que la chaîne en entrée contienne vraiment un `.`

```
In [41]: print(nice(re.match(r"abc\\.", "abc.")))
```

```
Match!
```

**En série ou en parallèle** Si je fais une analogie avec les montages électriques, jusqu'ici on a vu le montage en série, on met des expressions régulières bout à bout qui filtrent (`match`) la chaîne en entrée séquentiellement du début à la fin. On a *un peu* de marge pour spécifier des alternatives, lorsqu'on fait par exemple

```
"ab[cd]ef"
```

mais c'est limité à **un seul** caractère. Si on veut reconnaître deux mots qui n'ont pas grand-chose à voir comme `abc` ou `def`, il faut en quelque sorte mettre deux regexps en parallèle, et c'est ce que permet l'opérateur `|`

```
In [42]: regexp = "abc|def"
```

```
for sample in ['abc', 'def', 'aef']:
    match = re.match(regexp, sample)
    print(f"{sample} / {regexp} → {nice(match)}")
```

```
abc / abc|def → Match!
def / abc|def → Match!
aef / abc|def → no
```

### Fin(s) de chaîne

Selon que vous utilisez `match` ou `search`, vous précisez si vous vous intéressez uniquement à un match en début (`match`) ou n'importe où (`search`) dans la chaîne.

Mais indépendamment de cela, il peut être intéressant de “coller” l'expression en début ou en fin de ligne, et pour ça il existe des caractères spéciaux :

- `^` lorsqu'il est utilisé comme un caractère (c'est à dire pas en début de `[]`) signifie un début de chaîne;
- `\A` a le même sens (sauf en mode MULTILINE), et je le recommande de préférence à `^` qui est déjà pas mal surchargé;
- `$` matche une fin de ligne;
- `\Z` est voisin mais pas tout à fait identique.

Reportez-vous à la documentation pour le détails des différences. Attention aussi à entrer le `^` correctement, il vous faut le caractère ASCII et non un voisin dans la ménagerie Unicode.

```
In [43]: sample = 'abcd'
```

```
for regexp in [ r'bc', r'\Aabc', r'^abc',
                r'\Abc', r'^bc', r'bcd\Z',
                r'bcd$', r'bc\Z', r'bc$' ]:
    match = re.match(regexp, sample)
    search = re.search(regexp, sample)
    print(f"{sample} / {regexp:5s} match → {nice(match):6s} search → {nice(search):6s}")
```

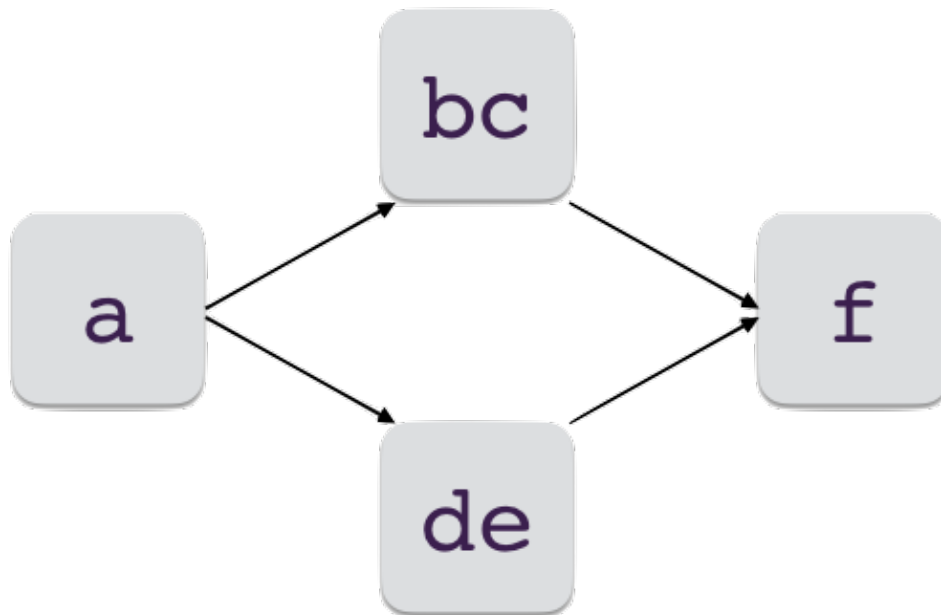
abcd / bc	match → no	search → Match!
abcd / \Aabc	match → Match!	search → Match!
abcd / ^abc	match → Match!	search → Match!
abcd / \Abc	match → no	search → no
abcd / ^bc	match → no	search → no
abcd / bcd\Z	match → no	search → Match!
abcd / bcd\$	match → no	search → Match!
abcd / bc\Z	match → no	search → no
abcd / bc\$	match → no	search → no

On a en effet bien le pattern `bc` dans la chaîne en entrée, mais il n'est ni au début ni à la fin.

### Parenthéser - (grouper)

Pour pouvoir faire des montages élaborés, il faut pouvoir parenthéser.

```
In [44]: # une parenthèse dans une RE
         # pour mettre en ligne:
         # un début 'a',
         # un milieu 'bc' ou 'de'
         # et une fin 'f'
         regexp = "a(bc|de)f"
```



```
In [45]: for sample in ['abcf', 'adef', 'abef', 'abf']:
        match = re.match(regex, sample)
        print(f"{sample:>4s} → {nice(match)}")
```

```
abcf → Match!
adef → Match!
abef → no
abf  → no
```

Les parenthèses jouent un rôle additionnel de **groupe**, ce qui signifie qu'on **peut retrouver** le texte correspondant à l'expression régulière comprise dans les (). Par exemple, pour le premier match

```
In [46]: sample = 'abcf'
        match = re.match(regex, sample)
        print(f"{sample}, {regex} → {match.groups()}")
```

```
abcf, a(bc|de)f → ('bc',)
```

dans cet exemple, on n'a utilisé qu'un seul groupe (), et le morceau de chaîne qui correspond à ce groupe se trouve donc être le seul groupe retourné par `MatchObject.group`.

### Compter les répétitions

Vous disposez des opérateurs suivants :

- \* l'étoile qui signifie n'importe quel nombre, même nul, d'occurrences - par exemple, (ab)\* pour indiquer '' ou 'ab' ou 'abab' ou etc.,
- + le plus qui signifie au moins une occurrence - e.g. (ab)+ pour ab ou abab ou ababab ou etc.,

- ? qui indique une option, c'est-à-dire 0 ou 1 occurrence - autrement dit (ab)? matche '' ou ab,
- {n} pour exactement n occurrences de (ab) - e.g. (ab){3} qui serait exactement équivalent à ababab,
- {m,n} entre m et n fois inclusivement.

```
In [47]: samples = [n*'ab' for n in [0, 1, 3, 4]] + ['baba']

for regexp in ['(ab)*', '(ab)+', '(ab){3}', '(ab){3,4}']:
    # on ajoute \A \Z pour matcher toute la chaîne
    line_regexp = r"\A{}\Z".format(regexp)
    for sample in samples:
        match = re.match(line_regexp, sample)
        print(f"{sample:>8s} / {line_regexp:14s} → {nice(match)}")

/ \A(ab)*\Z      → Match!
ab / \A(ab)*\Z    → Match!
ababab / \A(ab)*\Z → Match!
abababab / \A(ab)*\Z → Match!
baba / \A(ab)*\Z  → no
/ \A(ab+)\Z      → no
ab / \A(ab+)\Z    → Match!
ababab / \A(ab+)\Z → Match!
abababab / \A(ab+)\Z → Match!
baba / \A(ab+)\Z  → no
/ \A(ab){3}\Z    → no
ab / \A(ab){3}\Z  → no
ababab / \A(ab){3}\Z → Match!
abababab / \A(ab){3}\Z → no
baba / \A(ab){3}\Z → no
/ \A(ab){3,4}\Z  → no
ab / \A(ab){3,4}\Z → no
ababab / \A(ab){3,4}\Z → Match!
abababab / \A(ab){3,4}\Z → Match!
baba / \A(ab){3,4}\Z → no
```

### Groupes et contraintes

Nous avons déjà vu un exemple de groupe nommé (voir `needle` plus haut), les opérateurs que l'on peut citer dans cette catégorie sont :

- (...) les parenthèses définissent un groupe anonyme,
- (?P<name>...) définit un groupe nommé,
- (?:...) permet de mettre des parenthèses mais sans créer un groupe, pour optimiser l'exécution puisqu'on n'a pas besoin de conserver les liens vers la chaîne d'entrée,
- (?P=name) qui ne matche que si l'on retrouve à cet endroit de l'entrée la même sous-chaîne que celle trouvée pour le groupe name en amont,
- enfin (?=...), (?!\...) et (?<=...) permettent des contraintes encore plus élaborées, nous vous laissons le soin d'expérimenter avec elles si vous êtes intéressés; sachez toutefois que l'utilisation de telles constructions peut en théorie rendre l'interprétation de votre expression régulière beaucoup moins efficace.

### Greedy vs non-greedy

Lorsqu'on stipule une répétition un nombre indéfini de fois, il se peut qu'il existe **plusieurs** façons de filtrer l'entrée avec l'expression régulière. Que ce soit avec `*`, ou `+`, ou `?`, l'algorithme va toujours essayer de trouver la **séquence la plus longue**, c'est pourquoi on qualifie l'approche de *greedy* - quelque chose comme glouton en français.

```
In [48]: # un fragment d'HTML
line='<h1>Title</h1>'

# si on cherche un texte quelconque entre crochets
# c'est-à-dire l'expression régulière "<.*>"
re_greedy = '<.*>'

# on obtient ceci
# on rappelle que group(0) montre la partie du fragment
# HTML qui matche l'expression régulière
match = re.match(re_greedy, line)
match.group(0)
```

```
Out[48]: '<h1>Title</h1>'
```

Ça n'est pas forcément ce qu'on voulait faire, aussi on peut spécifier l'approche inverse, c'est-à-dire de trouver la **plus-petite** chaîne qui matche, dans une approche dite *non-greedy*, avec les opérateurs suivants :

- `*? : *` mais *non-greedy*,
- `+? : +` mais *non-greedy*,
- `?? : ?` mais *non-greedy*,

```
In [49]: # ici on va remplacer * par *? pour rendre l'opérateur * non-greedy
re_non_greedy = re_greedy = '<.*?>'

# mais on continue à chercher un texte entre <> naturellement
# si bien que cette fois, on obtient
match = re.match(re_non_greedy, line)
match.group(0)
```

```
Out[49]: '<h1>'
```

### S'agissant du traitement des fins de ligne

Il peut être utile, pour conclure cette présentation, de préciser un peu le comportement de la librairie vis-à-vis des fins de ligne.

Historiquement, les expressions régulières telles qu'on les trouve dans les librairies C, donc dans `sed`, `grep` et autres utilitaires Unix, sont associées au modèle mental où on filtre les entrées ligne par ligne.

Le module `re` en garde des traces, puisque

```
In [50]: # un exemple de traitement des 'newline'
sample = """une entrée
sur
plusieurs
lignes
"""
```

```
In [51]: match = re.compile("(.*").match(sample)
match.groups()
```

```
Out[51]: ('une entrée',)
```

Vous voyez donc que l'attrape-tout ' .' en fait n'attrape pas le caractère de fin de ligne `\n`, puisque si c'était le cas et compte tenu du côté *greedy* de l'algorithme on devrait voir ici tout le contenu de `sample`. Il existe un *flag* `re.DOTALL` qui permet de faire de `.` un vrai attrape-tout qui capture aussi les *newline*

```
In [52]: match = re.compile("(.*", flags=re.DOTALL).match(sample)
match.groups()
```

```
Out[52]: ('une entrée\nsur\nplusieurs\nlignes\n',)
```

Cela dit, le caractère *newline* est par ailleurs considéré comme un caractère comme un autre, on peut le mentionner **dans une regexp** comme les autres. Voici quelques exemples pour illustrer tout ceci

```
In [53]: # sans mettre le flag unicode \w ne matche que l'ASCII
match = re.compile("([\w]*)").match(sample)
match.groups()
```

```
Out[53]: ('une entrée',)
```

```
In [54]: # sans mettre le flag unicode \w ne matche que l'ASCII
match = re.compile("([\w]*)", flags=re.U).match(sample)
match.groups()
```

```
Out[54]: ('une entrée',)
```

```
In [55]: # si on ajoute \n à la liste des caractères attendus
# on obtient bien tout le contenu initial

# attention ici il ne FAUT PAS utiliser un raw string,
# car on veut vraiment écrire un newline dans la regexp

match = re.compile("([\w \n]*)", flags=re.UNICODE).match(sample)
match.groups()
```

```
Out[55]: ('une entrée\nsur\nplusieurs\nlignes\n',)
```

## Conclusion

La mise au point d'expressions régulières est certes un peu exigeante, et demande pas mal de pratique, mais permet d'écrire en quelques lignes des fonctionnalités très puissantes, c'est un investissement très rentable :)

Je vous signale enfin l'existence de **sites web** qui évaluent une expression régulière **de manière interactive** et qui peuvent rendre la mise au point moins fastidieuse.

Je vous signale notamment <https://pythex.org/>, et il en existe beaucoup d'autres.

## Pour en savoir plus

Pour ceux qui ont quelques rudiments de la théorie des langages, vous savez qu'on distingue en général

- l'**analyse lexicale**, qui découpe le texte en morceaux (qu'on appelle des *tokens*),
- et l'**analyse syntaxique** qui décrit pour simplifier à l'extrême l'ordre dans lequel on peut trouver les tokens.

Avec les expressions régulières, on adresse le niveau de l'analyse lexicale. Pour l'analyse syntaxique, qui est franchement au delà des objectifs de ce cours, il existe de nombreuses alternatives, parmi lesquelles :

- [pyparsing](#)
- [PLY \(Python Lex-Yacc\)](#)
- [ANTLR](#) qui est un outil écrit en Java mais qui peut générer des parsers en python,
- ...



## 2.6 Expressions régulières

Nous vous proposons dans ce notebook quelques exercices sur les expressions régulières. Faisons quelques remarques avant de commencer :

- nous nous concentrons sur l'écriture de l'expression régulière en elle-même, et pas sur l'utilisation de la bibliothèque ;
- en particulier, tous les exercices font appel à `re.match` entre votre *regex* et une liste de chaînes d'entrée qui servent de jeux de test.

**Liens utiles** Pour travailler sur ces exercices, il pourra être profitable d'avoir sous la main :

- la [documentation officielle](#) ;
- et [cet outil interactif sur https://pythex.org/](https://pythex.org/) qui permet d'avoir un retour presque immédiat, et donc d'accélérer la mise au point.

### 2.6.1 Exercice - niveau intermédiaire (1)

#### Identificateurs Python

```
In [ ]: # évaluez cette cellule pour charger l'exercice
        from regex_pythonid import exo_pythonid
```

On vous demande d'écrire une expression régulière qui décrit les noms de variable en Python. Pour cet exercice on se concentre sur les caractères ASCII. On exclut donc les noms de variables qui pourraient contenir des caractères exotiques comme les caractères accentués ou autres lettres grecques.

Il s'agit donc de reconnaître toutes les chaînes qui commencent par une lettre ou un `_`, suivi de lettres, chiffres ou `_`.

```
In [ ]: # quelques exemples de résultat attendus
        exo_pythonid.example()
```

```
In [ ]: # à vous de jouer: écrivez ici
        # sous forme de chaîne votre expression régulière
```

```
        regex_pythonid = r"<votre_regex>"
```

```
In [ ]: # évaluez cette cellule pour valider votre code
        exo_pythonid.correction(regex_pythonid)
```

### 2.6.2 Exercice - niveau intermédiaire (2)

#### Lignes avec nom et prénom

```
In [ ]: # pour charger l'exercice
        from corrections.regex_agenda import exo_agenda
```

On veut reconnaître dans un fichier toutes les lignes qui contiennent un nom et un prénom.

```
In [ ]: exo_agenda.example()
```

Plus précisément, on cherche les chaînes qui :

- commencent par une suite - possiblement vide - de caractères alphanumériques (vous pouvez utiliser `\w`) ou tiret haut (-) qui constitue le prénom;
- contiennent ensuite comme séparateur le caractère 'deux-points' ;
- contiennent ensuite une suite - cette fois jamais vide - de caractères alphanumériques, qui constitue le nom;
- et enfin contiennent un deuxième caractère : mais optionnellement seulement.

On vous demande de construire une expression régulière qui définit les deux groupes nom et prénom, et qui rejette les lignes qui ne satisfont pas ces critères.

```
In [ ]: # entrez votre regexp ici
        # il faudra la faire terminer par \Z
        # regardez ce qui se passe si vous ne le faites pas

        regexp_agenda = r"<votre regexp>\Z"

In [ ]: # évaluez cette cellule pour valider votre code
        exo_agenda.correction(regexp_agenda)
```

### 2.6.3 Exercice - niveau intermédiaire (3)

#### Numéros de téléphone

```
In [ ]: # pour charger l'exercice
        from corrections.regexp_phone import exo_phone
```

Cette fois on veut reconnaître des numéros de téléphone français, qui peuvent être :

- soit au format contenant 10 chiffres dont le premier est un 0 ;
- soit un format international commençant par +33 suivie de 9 chiffres.

Dans tous les cas on veut trouver dans le groupe 'number' les 9 chiffres vraiment significatifs, comme ceci :

```
In [ ]: exo_phone.example()

In [ ]: # votre regexp
        # à nouveau il faut terminer la regexp par \Z
        regexp_phone = r"<votre regexp>\Z"

In [ ]: # évaluez cette cellule pour valider votre code
        exo_phone.correction(regexp_phone)
```

### 2.6.4 Exercice - niveau avancé

Vu comment sont conçus les exercices, vous ne pouvez pas passer à `re.compile` un drapeau comme `re.IGNORECASE` ou autre ; sachez cependant que vous pouvez **embarquer ces drapeaux dans la regexp** elle-même ; par exemple pour rendre la regexp insensible à la casse de caractères, au lieu d'appeler `re.compile` avec le flag `re.I`, vous pouvez utiliser `(?i)` comme ceci :

```
In [ ]: import re

In [ ]: # on peut embarquer les flags comme IGNORECASE
        # directement dans la regexp
        # c'est équivalent de faire ceci

        re_obj = re.compile("abc", flags=re.IGNORECASE)
        re_obj.match("ABC").group(0)
```

```
In [ ]: # ou cela

re.match("(?i)abc", "ABC").group(0)

In [ ]: # les flags comme (?i) doivent apparaître
# en premier dans la regexp
re.match("abc(?i)", "ABC").group(0)
```

Pour plus de précisions sur ce trait, que nous avons laissé de côté dans le complément pour ne pas trop l'alourdir, voyez [la documentation sur les expressions régulières](#) et cherchez la première occurrence de `ilmsux`.

### Décortiquer une URL

On vous demande d'écrire une expression régulière qui permette d'analyser des URLs. Voici les conventions que nous avons adoptées pour l'exercice :

- la chaîne contient les parties suivantes :
  - `<protocol>://<location>/<path>`;
- l'URL commence par le nom d'un protocole qui doit être parmi `http`, `https`, `ftp`, `ssh`;
- le nom du protocole peut contenir de manière indifférente des minuscules ou des majuscules;
- ensuite doit venir la séquence `://`;
- ensuite on va trouver une chaîne `<location>` qui contient :
  - potentiellement un nom d'utilisateur, et s'il est présent, potentiellement un mot de passe;
  - obligatoirement un nom de `hostname`;
  - potentiellement un numéro de port;
- lorsque les 4 parties sont présentes dans `<location>`, cela se présente comme ceci :
  - `<location> = <user>:<password>@<hostname>:<port>`;
- si l'on note entre crochets les parties optionnelles, cela donne :
  - `<location> = [<user>[:<password>]@]<hostname>[:<port>]`;
- le champ `<user>` ne peut contenir que des caractères alphanumériques ; si le `@` est présent le champ `<user>` ne peut pas être vide;
- le champ `<password>` peut contenir tout sauf un `:` et de même, si le `:` est présent le champ `<password>` ne peut pas être vide;
- le champ `<hostname>` peut contenir une suite non-vide de caractères alphanumériques, underscores, ou `.` ;
- le champ `<port>` ne contient que des chiffres, et il est non vide si le `:` est spécifié;
- le champ `<path>` peut être vide.

Enfin, vous devez définir les groupes `proto`, `user`, `password`, `hostname`, `port` et `path` qui sont utilisés pour vérifier votre résultat. Dans la case `Résultat attendu`, vous trouverez soit `None` si la regexp ne filtre pas l'intégralité de l'entrée, ou bien une liste ordonnée de tuples qui donnent la valeur de ces groupes ; vous n'avez rien à faire pour construire ces tuples, c'est l'exercice qui s'en occupe.

```
In [ ]: # pour charger l'exercice
from corrections.regex_url import exo_url
```

```
In [ ]: # exemples du résultat attendu
        exo_url.example()
```

```
In [ ]: # n'hésitez pas à construire votre regexp petit à petit
```

```
        regexp_url = "<votre_regexp>"
```

```
In [ ]: exo_url.correction(regexp_url)
```

## 2.7 Les slices en Python

### 2.7.1 Complément - niveau basique

Ce support de cours reprend les notions de *slicing* vues dans la vidéo.

Nous allons illustrer les slices sur la chaîne suivante, rappelez-vous toutefois que ce mécanisme fonctionne avec toutes les séquences que l'on verra plus tard, comme les listes ou les tuples.

```
In [1]: chaine = "abcdefghijklmnopqrstuvwxy"
        print(chaine)
```

```
abcdefghijklmnopqrstuvwxy
```

#### Slice sans pas

On a vu en cours qu'une slice permet de désigner toute une plage d'éléments d'une séquence. Ainsi on peut écrire :

```
In [2]: chaine[2:6]
```

```
Out[2]: 'cdef'
```

#### Conventions de début et fin

Les débutants ont parfois du mal avec les bornes. Il faut se souvenir que :

- les indices **commencent** comme toujours à **zéro** ;
- le premier indice début est **inclus** ;
- le second indice fin est **exclu** ;
- on obtient en tout fin-début items dans le résultat.

Ainsi, ci-dessus, le résultat contient  $6 - 2 = 4$  éléments.

Pour vous aider à vous souvenir des conventions de début et de fin, souvenez-vous qu'on veut pouvoir facilement juxtaposer deux slices qui ont une borne commune.

C'est-à-dire qu'avec :

	0	1	2	3	4	5	6	7	8	9
	abcdefghijklmnop									
[0:3]	[	x	x	x	[					
[3:7]					[	x	x	x	x	[
[0:7]	[	x	x	x	x	x	x	x	[	

début et fin

```
In [3]: # chaine[a:b] + chaine[b:c] == chaine[a:c]
        chaine[0:3] + chaine[3:7] == chaine[0:7]
```

```
Out[3]: True
```

### Bornes omises

On peut omettre une borne :

```
In [4]: # si on omet la première borne, cela signifie que
        # la slice commence au début de l'objet
        chaine[:6]
```

```
Out[4]: 'abcdef'
```

```
In [5]: # et bien entendu c'est la même chose si on omet la deuxième borne
        chaine[24:]
```

```
Out[5]: 'yz'
```

```
In [6]: # ou même omettre les deux bornes, auquel cas on
        # fait une copie de l'objet - on y reviendra plus tard
        chaine[:]
```

```
Out[6]: 'abcdefghijklmnopqrstuvwxyz'
```

### Indices négatifs

On peut utiliser des indices négatifs pour compter à partir de la fin :

```
In [7]: chaine[3:-3]
```

```
Out[7]: 'defghijklmnopqrstuvw'
```

```
In [8]: chaine[-3:]
```

```
Out[8]: 'xyz'
```

### Slice avec pas

Il est également possible de préciser un *pas*, de façon à ne choisir par exemple, dans la plage donnée, qu'un élément sur deux :

```
In [9]: # le pas est précisé après un deuxième deux-points (:)
        # ici on va choisir un caractère sur deux dans la plage [3:-3]
        chaine[3:-3:2]
```

```
Out[9]: 'dfhjlnprtv'
```

Comme on le devine, le troisième élément de la slice, ici 2, détermine le pas. On ne retient donc, dans la chaîne defghi... que d, puis f, et ainsi de suite.

On peut préciser du coup la borne de fin (ici -3) avec un peu de liberté, puisqu'ici on obtiendrait un résultat identique avec -4.

```
In [10]: chaine[3:-4:2]
```

```
Out[10]: 'dfhjlnprtv'
```

**Pas négatif**

Il est même possible de spécifier un pas négatif. Dans ce cas, de manière un peu contre-intuitive, il faut préciser un début (le premier indice de la slice) qui soit *plus à droite* que la fin (le second indice).

Pour prendre un exemple, comme l'élément d'indice -3, c'est-à-dire *x*, est plus à droite que l'élément d'indice 3, c'est-à-dire *d*, évidemment si on ne précisait pas le pas (qui revient à choisir un pas égal à 1), on obtiendrait une liste vide :

```
In [11]: chaine[-3:3]
```

```
Out[11]: ''
```

Si maintenant on précise un pas négatif, on obtient cette fois :

```
In [12]: chaine[-3:3:-2]
```

```
Out[12]: 'xvtrpnljhf'
```

**Conclusion**

À nouveau, souvenez-vous que tous ces mécanismes fonctionnent avec de nombreux autres types que les chaînes de caractères. En voici deux exemples qui anticipent tous les deux sur la suite, mais qui devraient illustrer les vastes possibilités qui sont offertes avec les slices.

**Listes** Par exemple sur les listes :

```
In [13]: liste = [0, 2, 4, 8, 16, 32, 64, 128]
         liste
```

```
Out[13]: [0, 2, 4, 8, 16, 32, 64, 128]
```

```
In [14]: liste[-1:1:-2]
```

```
Out[14]: [128, 32, 8]
```

Et même ceci, qui peut être déroutant. Nous reviendrons dessus.

```
In [15]: liste[2:4] = [100, 200, 300, 400, 500]
         liste
```

```
Out[15]: [0, 2, 100, 200, 300, 400, 500, 16, 32, 64, 128]
```

**2.7.2 Complément - niveau avancé**

**numpy** La bibliothèque numpy permet de manipuler des tableaux ou des matrices. En anticipant (beaucoup) sur son usage que nous reverrons bien entendu en détail, voici un aperçu de ce que l'on peut faire avec des slices sur des objets numpy :

```
In [16]: # ces deux premières cellules sont à admettre
         # on construit un tableau ligne
         import numpy as np

         un_cinq = np.array([1, 2, 3, 4, 5])
         un_cinq
```

```
Out[16]: array([1, 2, 3, 4, 5])
```

```
In [17]: # ces deux premières cellules sont à admettre
         # on le combine avec lui-même - et en utilisant une slice un peu magique
         # pour former un tableau carré 5x5
```

```
array = 10 * un_cinq[:, np.newaxis] + un_cinq
array
```

```
Out[17]: array([[11, 12, 13, 14, 15],
               [21, 22, 23, 24, 25],
               [31, 32, 33, 34, 35],
               [41, 42, 43, 44, 45],
               [51, 52, 53, 54, 55]])
```

Sur ce tableau de taille 5x5, nous pouvons aussi faire du slicing et extraire le sous-tableau 3x3 au centre :

```
In [18]: centre = array[1:4, 1:4]
         centre
```

```
Out[18]: array([[22, 23, 24],
               [32, 33, 34],
               [42, 43, 44]])
```

On peut bien sûr également utiliser un pas :

```
In [19]: coins = array[:, :4, ::4]
         coins
```

```
Out[19]: array([[11, 15],
               [51, 55]])
```

Ou bien retourner complètement dans une direction :

```
In [20]: tete_en_bas = array[:, :-1, :]
         tete_en_bas
```

```
Out[20]: array([[51, 52, 53, 54, 55],
               [41, 42, 43, 44, 45],
               [31, 32, 33, 34, 35],
               [21, 22, 23, 24, 25],
               [11, 12, 13, 14, 15]])
```



## 2.8 Méthodes spécifiques aux listes

### 2.8.1 Complément - niveau basique

Voici quelques unes des méthodes disponibles sur le type `list`.

#### Trouver l'information

Pour commencer, rappelons comment retrouver la liste des méthodes définies sur le type `list`:

```
In [1]: help(list)
```

Help on class list in module builtins:

```
class list(object)
| list() -> new empty list
| list(iterable) -> new list initialized from iterable's items
|
| Methods defined here:
|
| __add__(self, value, /)
|     Return self+value.
|
| __contains__(self, key, /)
|     Return key in self.
|
| __delitem__(self, key, /)
|     Delete self[key].
|
| __eq__(self, value, /)
|     Return self==value.
|
| __ge__(self, value, /)
|     Return self>=value.
|
| __getattr__(self, name, /)
|     Return getattr(self, name).
|
| __getitem__(...)
|     x.__getitem__(y) <==> x[y]
|
| __gt__(self, value, /)
|     Return self>value.
|
| __iadd__(self, value, /)
|     Implement self+=value.
|
| __imul__(self, value, /)
|     Implement self*=value.
|
| __init__(self, /, *args, **kwargs)
|     Initialize self. See help(type(self)) for accurate signature.
```

```

|
|  __iter__(self, /)
|      Implement iter(self).
|
|  __le__(self, value, /)
|      Return self<=value.
|
|  __len__(self, /)
|      Return len(self).
|
|  __lt__(self, value, /)
|      Return self<value.
|
|  __mul__(self, value, /)
|      Return self*value.n
|
|  __ne__(self, value, /)
|      Return self!=value.
|
|  __new__(*args, **kwargs) from builtins.type
|      Create and return a new object. See help(type) for accurate signature.
|
|  __repr__(self, /)
|      Return repr(self).
|
|  __reversed__(...)
|      L.__reversed__() -- return a reverse iterator over the list
|
|  __rmul__(self, value, /)
|      Return self*value.
|
|  __setitem__(self, key, value, /)
|      Set self[key] to value.
|
|  __sizeof__(...)
|      L.__sizeof__() -- size of L in memory, in bytes
|
|  append(...)
|      L.append(object) -> None -- append object to end
|
|  clear(...)
|      L.clear() -> None -- remove all items from L
|
|  copy(...)
|      L.copy() -> list -- a shallow copy of L
|
|  count(...)
|      L.count(value) -> integer -- return number of occurrences of value
|
|  extend(...)
|      L.extend(iterable) -> None -- extend list by appending elements from the iterable

```

```

|
|  index(...)
|      L.index(value, [start, [stop]]) -> integer -- return first index of value.
|      Raises ValueError if the value is not present.
|
|  insert(...)
|      L.insert(index, object) -- insert object before index
|
|  pop(...)
|      L.pop([index]) -> item -- remove and return item at index (default last).
|      Raises IndexError if list is empty or index is out of range.
|
|  remove(...)
|      L.remove(value) -> None -- remove first occurrence of value.
|      Raises ValueError if the value is not present.
|
|  reverse(...)
|      L.reverse() -- reverse *IN PLACE*
|
|  sort(...)
|      L.sort(key=None, reverse=False) -> None -- stable sort *IN PLACE*
|
|  -----
|  Data and other attributes defined here:
|
|  __hash__ = None

```

Ignorez les méthodes dont le nom commence et termine par `__` (nous parlerons de ceci en semaine 6), vous trouvez alors les méthodes utiles listées entre `append` et `sort`.

Certaines de ces méthodes ont été vues dans la vidéo sur les séquences, c'est le cas notamment de `count` et `index`.

Nous allons à présent décrire les autres, partiellement et brièvement. Un autre complément décrit la méthode `sort`. Reportez-vous au lien donné en fin de notebook pour obtenir une information plus complète.

Donnons-nous pour commencer une liste témoin :

```

In [2]: liste = [0, 1, 2, 3]
        print('liste', liste)

```

```
liste [0, 1, 2, 3]
```

#### Avertissements :

- soyez bien attentifs au nombre de fois où vous exécutez les cellules de ce notebook;
- par exemple une liste renversée deux fois peut donner l'impression que `reverse` ne marche pas;
- n'hésitez pas à utiliser le menu *Cell -> Run All* pour réexécuter en une seule fois le notebook entier.

## append

La méthode `append` permet d'ajouter **un élément** à la fin d'une liste :

```
In [3]: liste.append('ap')
        print('liste', liste)
```

```
liste [0, 1, 2, 3, 'ap']
```

## extend

La méthode `extend` réalise la même opération, mais avec **tous les éléments** de la liste qu'on lui passe en argument :

```
In [4]: liste2 = ['ex1', 'ex2']
        liste.extend(liste2)
        print('liste', liste)
```

```
liste [0, 1, 2, 3, 'ap', 'ex1', 'ex2']
```

## append vs +

Ces deux méthodes `append` et `extend` sont donc assez voisines ; avant de voir d'autres méthodes de `list`, prenons un peu le temps de comparer leur comportement avec l'addition `+` de liste. L'élément clé ici, on l'a déjà vu dans la vidéo, est que la liste est un objet **mutable**. `append` et `extend` **modifient** la liste sur laquelle elles travaillent, alors que l'addition **crée un nouvel objet**.

```
In [5]: # pour créer une liste avec les n premiers entiers, on utilise
        # la fonction built-in range(), que l'on convertit en liste
        # on aura l'occasion d'y revenir
        a1 = list(range(3))
        print(a1)
```

```
[0, 1, 2]
```

```
In [6]: a2 = list(range(10, 13))
        print(a2)
```

```
[10, 11, 12]
```

```
In [7]: # le fait d'utiliser + crée une nouvelle liste
        a3 = a1 + a2
```

```
In [8]: # si bien que maintenant on a trois objets différents
        print('a1', a1)
        print('a2', a2)
        print('a3', a3)
```

```
a1 [0, 1, 2]
a2 [10, 11, 12]
a3 [0, 1, 2, 10, 11, 12]
```

Comme on le voit, après une addition, les deux termes de l'addition sont inchangés. Pour bien comprendre, voyons exactement le même scénario sous `pythontutor` :

```
In [ ]: %load_ext ipythontutor
```

**Note** : une fois que vous avez évalué la cellule avec `%%ipythontutor`, vous devez cliquer sur le bouton Forward pour voir pas à pas le comportement du programme.

```
In [ ]: %%ipythontutor height=230 ratio=0.7
        a1 = list(range(3))
        a2 = list(range(10, 13))
        a3 = a1 + a2
```

Alors que si on avait utilisé `extend`, on aurait obtenu ceci :

```
In [ ]: %%ipythontutor height=200 ratio=0.75
        e1 = list(range(3))
        e2 = list(range(10, 13))
        e3 = e1.extend(e2)
```

Ici on tire profit du fait que la liste est un objet mutable : `extend` **modifie** l'objet sur lequel on l'appelle (ici `e1`). Dans ce scénario on ne crée en tout que deux objets, et du coup il est inutile pour `extend` de renvoyer quoi que ce soit, et c'est pourquoi `e3` ici vaut `None`.

C'est pour cette raison que :

- l'addition est disponible sur tous les types séquences - on peut toujours réaliser l'addition puisqu'on crée un nouvel objet pour stocker le résultat de l'addition ;
- mais `append` et `extend` ne sont par exemple **pas disponibles** sur les chaînes de caractères, qui sont **immuables** - si `e1` était une chaîne, on ne pourrait pas la modifier pour lui ajouter des éléments.

## insert

Reprenons notre inventaire des méthodes de `list`, et pour cela rappelons nous le contenu de la variable `liste` :

```
In [9]: liste
```

```
Out[9]: [0, 1, 2, 3, 'ap', 'ex1', 'ex2']
```

La méthode `insert` permet, comme le nom le suggère, d'insérer un élément à une certaine position ; comme toujours les indices commencent à zéro et donc :

```
In [10]: # insérer à l'index 2
          liste.insert(2, '1 bis')
          print('liste', liste)
```

```
liste [0, 1, '1 bis', 2, 3, 'ap', 'ex1', 'ex2']
```

On peut remarquer qu'un résultat analogue peut être obtenu avec une affectation de slice ; par exemple pour insérer au rang 5 (i.e. avant `ap`), on pourrait aussi bien faire :

```
In [11]: liste[5:5] = ['3 bis']
         print('liste', liste)

liste [0, 1, '1 bis', 2, 3, '3 bis', 'ap', 'ex1', 'ex2']
```

### remove

La méthode `remove` détruit la **première occurrence** d'un objet dans la liste :

```
In [12]: liste.remove(3)
         print('liste', liste)

liste [0, 1, '1 bis', 2, '3 bis', 'ap', 'ex1', 'ex2']
```

### pop

La méthode `pop` prend en argument un indice ; elle permet d'extraire l'élément à cet indice. En un seul appel on obtient la valeur de l'élément et on l'enlève de la liste :

```
In [13]: popped = liste.pop(0)
         print('popped', popped, 'liste', liste)

popped 0 liste [1, '1 bis', 2, '3 bis', 'ap', 'ex1', 'ex2']
```

Si l'indice n'est pas précisé, c'est le dernier élément de la liste qui est visé :

```
In [14]: popped = liste.pop()
         print('popped', popped, 'liste', liste)

popped ex2 liste [1, '1 bis', 2, '3 bis', 'ap', 'ex1']
```

### reverse

Enfin `reverse` renverse la liste, le premier élément devient le dernier :

```
In [15]: liste.reverse()
         print('liste', liste)

liste ['ex1', 'ap', '3 bis', 2, '1 bis', 1]
```

On peut remarquer ici que le résultat se rapproche de ce qu'on peut obtenir avec une opération de slicing comme ceci :

```
In [16]: liste2 = liste[::-1]
         print('liste2', liste2)

liste2 [1, '1 bis', 2, '3 bis', 'ap', 'ex1']
```

**À la différence toutefois** qu'avec le slicing c'est une copie de la liste initiale qui est retournée, la liste de départ quant à elle n'est pas modifiée.

**Pour en savoir plus**

<https://docs.python.org/3/tutorial/datastructures.html#more-on-lists>

**Note spécifique aux notebooks**

help avec ? Je vous signale en passant que dans un notebook vous pouvez obtenir de l'aide avec un point d'interrogation ? inséré avant ou après un symbole. Par exemple pour obtenir des précisions sur la méthode `list.pop`, on peut faire soit :

```
In [17]: # fonctionne dans tous les environnements Python
         help(list.pop)
```

Help on method\_descriptor:

```
pop(...)
L.pop([index]) -> item -- remove and return item at index (default last).
Raises IndexError if list is empty or index is out of range.
```

```
In [18]: # spécifique aux notebooks
         # l'affichage obtenu est légèrement différent
         # tapez la touche 'Esc' - ou cliquez la petite croix
         # pour faire disparaître le dialogue qui apparaît en bas
         list.pop?
```

**Complétion avec Tab**

Dans un notebook vous avez aussi la complétion; si vous tapez - dans une cellule de code - le début d'un symbole connu dans l'environnement :

```
In [ ]: # placez votre curseur à la fin de la ligne après 'li'
         # et appuyez sur la touche 'Tab'
         li
```

Vous voyez apparaître un dialogue avec les noms connus qui commencent par `li` ; utilisez les flèches pour choisir, et 'Return' pour sélectionner.

## 2.9 Objets mutables et objets immuables

### 2.9.1 Complément - niveau basique

#### Les chaînes sont des objets immuables

Voici un exemple d'un fragment de code qui illustre le caractère immuable des chaînes de caractères. Nous l'exécutons sous [pythontutor](#), afin de bien illustrer les relations entre variables et objets.

```
In [ ]: # il vous faut charger cette cellule
        # pour pouvoir utiliser les suivantes
        %load_ext ipythontutor
```

**Note** : une fois que vous avez évalué la cellule avec `%%ipythontutor`, vous devez cliquer sur le bouton Forward pour voir pas à pas le comportement du programme.

Le scénario est très simple, on crée deux variables `s1` et `s2` vers le même objet `'abc'`, puis on fait une opération `+=` sur la variable `s1`.

Comme l'objet est une chaîne, il est donc immuable, on ne **peut pas modifier l'objet** directement; pour obtenir l'effet recherché (à savoir que `s1` s'allonge de `'def'`), Python **crée un deuxième objet**, comme on le voit bien sous `pythontutor` :

```
In [ ]: %%ipythontutor heapPrimitives=true
        # deux variables vers le même objet
        s1 = 'abc'
        s2 = s1
        # on essaie de modifier l'objet
        s1 += 'def'
        # pensez à cliquer sur `Forward`
```

#### Les listes sont des objets mutables

Voici ce qu'on obtient par contraste pour le même scénario mais qui cette fois utilise des listes, qui sont des objets mutables :

```
In [ ]: %%ipythontutor heapPrimitives=true ratio=0.8
        # deux variables vers le même objet
        liste1 = ['a', 'b', 'c']
        liste2 = liste1
        # on modifie l'objet
        liste1 += ['d', 'e', 'f']
        # pensez à cliquer sur `Forward`
```

#### Conclusion

Ce comportement n'est pas propre à l'usage de l'opérateur `+=` - que pour cette raison d'ailleurs nous avons tendance à déconseiller.

Les objets mutables et immuables ont par essence un comportement différent, il est très important d'avoir ceci présent à l'esprit.

Nous aurons notamment l'occasion d'approfondir cela dans la séquence consacrée aux références partagées, en semaine 3.



## 2.10 Tris de listes

### 2.10.1 Complément - niveau basique

Python fournit une méthode standard pour trier une liste, qui s'appelle, sans grande surprise, `sort`.

#### La méthode `sort`

Voyons comment se comporte `sort` sur un exemple simple :

```
In [1]: liste = [8, 7, 4, 3, 2, 9, 1, 5, 6]
        print('avant tri', liste)
        liste.sort()
        print('apres tri', liste)
```

```
avant tri [8, 7, 4, 3, 2, 9, 1, 5, 6]
apres tri [1, 2, 3, 4, 5, 6, 7, 8, 9]
```

On retrouve ici, avec l'instruction `liste.sort()` un cas d'appel de méthode (ici `sort`) sur un objet (ici `liste`), comme on l'avait vu dans la vidéo sur la notion d'objet.

La première chose à remarquer est que la liste d'entrée a été modifiée, on dit "en place", ou encore "par effet de bord". Voyons cela sous `pythontutor` :

```
In [ ]: %load_ext ipythontutor

In [ ]: %%ipythontutor height=200 ratio=0.8
        liste = [3, 2, 9, 1]
        liste.sort()
```

On aurait pu imaginer que la liste d'entrée soit restée inchangée, et que la méthode de tri renvoie une copie triée de la liste, ce n'est pas le choix qui a été fait, cela permet d'économiser des allocations mémoire autant que possible et d'accélérer sensiblement le tri.

#### La fonction `sorted`

Si vous avez besoin de faire le tri sur une copie de votre liste, la fonction `sorted` vous permet de le faire :

```
In [ ]: %%ipythontutor height=200 ratio=0.8
        liste1 = [3, 2, 9, 1]
        liste2 = sorted(liste1)
```

#### Tri décroissant

Revenons à la méthode `sort` et aux tris *en place*. Par défaut la liste est triée par ordre croissant, si au contraire vous voulez l'ordre décroissant, faites comme ceci :

```
In [2]: liste = [8, 7, 4, 3, 2, 9, 1, 5, 6]
        print('avant tri', liste)
        liste.sort(reverse=True)
        print('apres tri décroissant', liste)
```

```
avant tri [8, 7, 4, 3, 2, 9, 1, 5, 6]
apres tri décroissant [9, 8, 7, 6, 5, 4, 3, 2, 1]
```

Nous n'avons pas encore vu à quoi correspond cette formule `reverse=True` dans l'appel à la méthode - ceci sera approfondi dans le chapitre sur les appels de fonction - mais dans l'immédiat vous pouvez utiliser cette technique telle quelle.

### Chaînes de caractères

Cette technique fonctionne très bien sur tous les types numériques (enfin, à l'exception des complexes; en guise d'exercice, pourquoi?), ainsi que sur les chaînes de caractères :

```
In [3]: liste = ['spam', 'egg', 'bacon', 'beef']
        liste.sort()
        print('après tri', liste)

après tri ['bacon', 'beef', 'egg', 'spam']
```

Comme on s'y attend, il s'agit cette fois d'un **tri lexicographique**, dérivé de l'ordre sur les caractères. Autrement dit, c'est l'ordre du dictionnaire. Il faut souligner toutefois, pour les personnes n'ayant jamais été exposées à l'informatique, que cet ordre, quoique déterministe, est arbitraire en dehors des lettres de l'alphabet.

Ainsi par exemple :

```
In [4]: # deux caractères minuscules se comparent
        # comme on s'y attend
        'a' < 'z'
```

```
Out[4]: True
```

Bon, mais par contre :

```
In [5]: # si l'un est en minuscule et l'autre en majuscule,
        # ce n'est plus le cas
        'Z' < 'a'
```

```
Out[5]: True
```

Ce qui à son tour explique ceci :

```
In [6]: # la conséquence de 'Z' < 'a', c'est que
        liste = ['abc', 'Zoo']
        liste.sort()
        print(liste)

['Zoo', 'abc']
```

Et lorsque les chaînes contiennent des espaces ou autres ponctuations, le résultat du tri peut paraître surprenant :

```
In [7]: # attention ici notre premiere chaîne commence par une espace
        # et le caractère 'Espace' est plus petit
        # que tous les autres caractères imprimables
        liste = [' zoo', 'ane']
        liste.sort()
        print(liste)

[' zoo', 'ane']
```

### À suivre

Il est possible de définir soi-même le critère à utiliser pour trier une liste, et nous verrons cela bientôt, une fois que nous aurons introduit la notion de fonction.

## 2.11 Indentations en Python

### 2.11.1 Complément - niveau basique

#### Imbrications

Nous l'avons vu dans la vidéo, la pratique la plus courante est d'utiliser systématiquement une indentation de 4 espaces :

```
In [1]: # la convention la plus généralement utilisée
        # consiste à utiliser une indentation de 4 espaces
        if 'g' in 'egg':
            print('OUI')
        else:
            print('NON')
```

OUI

Voyons tout de suite comment on pourrait écrire plusieurs tests imbriqués :

```
In [2]: entree = 'spam'

        # pour imbriquer il suffit d'indenter de 8 espaces
        if 'a' in entree:
            if 'b' in entree:
                cas11 = True
                print('a et b')
            else:
                cas12 = True
                print('a mais pas b')
        else:
            if 'b' in entree:
                cas21 = True
                print('b mais pas a')
            else:
                cas22 = True
                print('ni a ni b')
```

a mais pas b

Dans cette construction assez simple, remarquez bien **les deux points ':'** à chaque début de bloc, c'est-à-dire à chaque fin de ligne `if` ou `else`.

Cette façon d'organiser le code peut paraître très étrange, notamment aux gens habitués à un autre langage de programmation, puisqu'en général les syntaxes des langages sont conçues de manière à être insensibles aux espaces et à la présentation.

Comme vous le constaterez à l'usage cependant, une fois qu'on s'y est habitué cette pratique est très agréable, une fois qu'on a écrit la dernière ligne du code, on n'a pas à réfléchir à refermer le bon nombre d'accolades ou de *end*.

Par ailleurs, comme pour tous les langages, votre éditeur favori connaît cette syntaxe et va vous aider à respecter la règle des 4 caractères. Nous ne pouvons pas publier ici une liste des commandes disponibles par éditeur, nous vous invitons le cas échéant à échanger entre vous sur le forum pour partager les recettes que vous utilisez avec votre éditeur / environnement de programmation favori.

## 2.11.2 Complément - niveau intermédiaire

### Espaces *vs* tabulations

#### Version courte

Il nous faut par contre donner quelques détails sur un problème que l'on rencontre fréquemment sur du code partagé entre plusieurs personnes quand celles-ci utilisent des environnements différents.

Pour faire court, ce problème est **susceptible d'apparaître dès qu'on utilise des tabulations**, plutôt que des espaces, pour implémenter les indentations. Aussi, le message à retenir ici est **de ne jamais utiliser de tabulations dans votre code Python**. Tout bon éditeur Python devrait faire cela par défaut.

#### Version longue

En version longue, il existe un code ASCII pour un caractère qui s'appelle *Tabulation* (alias Control-i, qu'on note aussi ^I); l'interprétation de ce caractère n'étant pas clairement spécifiée, il arrive qu'on se retrouve dans une situation comme la suivante.

Bernard utilise l'éditeur vim; sous cet éditeur il lui est possible de mettre des tabulations dans son code, et de choisir la valeur de ces tabulations. Aussi il va dans les préférences de vim, choisit Tabulation=4, et écrit un programme qu'il voit comme ceci :

```
In [3]: if 'a' in entree:
        if 'b' in entree:
            cas11 = True
            print('a et b')
        else:
            cas12 = True
            print('a mais pas b')
```

a mais pas b

Sauf qu'en fait, il a mis un mélange de tabulations et d'espaces, et en fait le fichier contient (avec ^I pour tabulation) :

```
if 'a' in entree:
^Iif 'b' in entree:
^I^Icas11 = True
    ^Iprint('a et b')
^Ielse:
^I^Icas12 = True
    ^Iprint('a mais pas b')
```

Remarquez le mélange de Tabulations et d’espaces dans les deux lignes avec `print`. Bernard envoie son code à Alice qui utilise `emacs`. Dans son environnement, `emacs` affiche une tabulation comme 8 caractères. Du coup Alice “voit” le code suivant :

```
if 'a' in entree:
    if 'b' in entree:
        cas11 = True
        print('a et b')
    else:
        cas12 = True
        print('a mais pas b')
```

Bref, c’est la confusion la plus totale. Aussi répétons-le, **n’utilisez jamais de tabulations dans votre code Python**.

Ce qui ne veut pas dire qu’il ne faut pas utiliser la touche `Tab` avec votre éditeur - au contraire, c’est une touche très utilisée - mais faites bien la différence entre le fait d’appuyer sur la touche `Tab` et le fait que le fichier sauvé sur disque contient effectivement un caractère tabulation. Votre éditeur favori propose très certainement une option permettant de faire les remplacements idoines pour ne pas écrire de tabulation dans vos fichiers, tout en vous permettant d’indenter votre code avec la touche `Tab`.

Signalons enfin que Python 3 est plus restrictif que Python 2 à cet égard, et interdit de mélanger des espaces et des tabulations sur une même ligne. Ce qui n’enlève rien à notre recommandation.

### 2.11.3 Complément - niveau avancé

Vous pouvez trouver du code qui ne respecte pas la convention des 4 caractères.

#### Version courte

En version courte : **Utilisez toujours des indentations de 4 espaces**.

#### Version longue

En version longue, et pour les curieux : Python **n’impose pas** que les indentations soient de 4 caractères. Aussi vous pouvez rencontrer un code qui ne respecte pas cette convention, et il nous faut, pour être tout à fait précis sur ce que Python accepte ou non, préciser ce qui est réellement requis par Python.

La règle utilisée pour analyser votre code, c’est que toutes les instructions **dans un même bloc** soient présentées avec le même niveau d’indentation. Si deux lignes successives - modulo les blocs imbriqués - ont la même indentation, elles sont dans le même bloc.

Voyons quelques exemples. Tout d’abord le code suivant est **légal**, quoique, redisons-le pour la dernière fois, **pas du tout recommandé** :

```
In [4]: # code accepté mais pas du tout recommandé
        if 'a' in 'pas du tout recommande':
            succes = True
            print('OUI')
```

```
else:
    print('NON')
```

OUI

En effet, les deux blocs (après if et après else) sont des blocs distincts, ils sont libres d'utiliser deux indentations différentes (ici 2 et 6).

Par contre la construction ci-dessous n'est pas légale :

```
In [5]: # ceci n'est pas correct et est rejeté par Python
```

```
if 'a' in entree:
    if 'b' in entree:
        cas11 = True
        print('a et b')
    else:
        cas12 = True
        print('a mais pas b')
```

```
File "<tokenize>", line 6
else:
^
```

```
IndentationError: unindent does not match any outer indentation level
```

En effet les deux lignes if et else font logiquement partie du même bloc, elles **doivent** donc avoir la même indentation. Avec cette présentation le lecteur Python émet une erreur et ne peut pas interpréter le code.

## 2.12 Bonnes pratiques de présentation de code

### 2.12.1 Complément - niveau basique

#### La PEP-008

On trouve [dans la PEP-008 \(en anglais\)](#) les conventions de codage qui s'appliquent à toute la librairie standard, et qui sont certainement un bon point de départ pour vous aider à trouver le style de présentation qui vous convient.

Nous vous recommandons en particulier les sections sur

- [l'indentation](#)
- [les espaces](#)
- [les commentaires](#)

#### Un peu de lecture : le module pprint

Voici par exemple le code du module pprint (comme PrettyPrint) de la librairie standard qui permet d'imprimer des données.

La fonction du module - le pretty printing - est évidemment accessoire ici, mais vous pouvez y voir illustré

- le *docstring* pour le module : les lignes de 11 à 35,
- les indentations, comme nous l'avons déjà mentionné sont à 4 espaces, et sans tabulation,
- l'utilisation des espaces, notamment autour des affectations et opérateurs, des définitions de fonction, des appels de fonctions...
- les lignes qui restent dans une largeur "raisonnable" (79 caractères)
- vous pouvez regarder notamment la façon de couper les lignes pour respecter cette limite en largeur.

```
In [ ]: from modtools import show_module_html
import pprint
show_module_html(pprint)
```

#### Espaces

Comme vous pouvez le voir dans `pprint.py`, les règles principales concernant les espaces sont les suivantes.

- S'agissant des **affectations** et **opérateurs**, on fera

`x = y + z`

Et non pas

`x=y+z`

Ni

`x =y+z`

Ni encore

`x=y+z`

L'idée étant d'aérer de manière homogène pour faciliter la lecture.



- On **déclare une fonction** comme ceci

```
def foo(x, y, z):
```

Et non pas comme ceci (un espace en trop avant la parenthèse ouvrante)

```
def foo (x,y,z):
```

Ni surtout comme ceci (pas d'espace entre les paramètres)

```
def foo(x,y,z):
```

- La même règle s'applique naturellement aux **appels de fonction** :

```
foo(x, y, z)
```

et non pas ~~foo(x,y,z)~~

```
ni def foo(x,y,z):
```

Il est important de noter qu'il s'agit ici de **règles d'usage** et non pas de règles syntaxiques ; tous les exemples barrés ci-dessus sont en fait **syntactiquement corrects**, l'interpréteur les accepterait sans souci ; mais ces règles sont **très largement adoptées**, et obligatoires pour intégrer du code dans la librairie standard.

## Coupages de ligne

Nous allons à présent zoomer dans ce module pour voir quelques exemples de coupure de ligne. Par contraste avec ce qui précède, il s'agit cette fois surtout de **règles syntaxiques**, qui peuvent rendre un code non valide si elles ne sont pas suivies.

### Coupure de ligne sans *backslash* (\)

```
In [ ]: show_module_html(pprint,
                        beg="def pprint",
                        end="def pformat")
```

La fonction `pprint` (ligne ~47) est une commodité (qui crée une instance de `PrettyPrinter`, sur lequel on envoie la méthode `pprint`).

Vous voyez ici qu'il n'est **pas nécessaire** d'insérer un *backslash* (\) à la fin des lignes 50 et 51, car il y a une parenthèse ouvrante qui n'est pas fermée à ce stade.

De manière générale, lorsqu'une parenthèse ouvrante ( - idem avec les crochets [ et accolades { - n'est pas fermée sur la même ligne, l'interpréteur suppose qu'elle sera fermée plus loin et n'impose pas de *backslash*.

Ainsi par exemple on peut écrire sans *backslash* :

```
valeurs = [
    1,
    2,
    3,
    5,
    7,
]
```

Ou encore

```
x = un_nom_de_fonction_tres_tres_long(
    argument1, argument2,
    argument3, argument4,
)
```

À titre de rappel, signalons aussi les chaînes de caractères à base de `"""` ou `'''` qui permettent elles aussi d'utiliser plusieurs lignes consécutives sans *backslash*, comme

```
texte = """ Les sanglots longs
Des violons
De l'automne"""
```

### Coupure de ligne avec *backslash* (\)

Par contre il est des cas où le *backslash* est nécessaire :

```
In [ ]: show_module_html(pprint,
                        beg="components), readable, recursive",
                        end="elif len(object) ",
                        lineno_width=3)
```

Dans ce fragment au contraire, vous voyez en ligne 522 qu'il **a fallu cette fois** insérer un *backslash* \ comme caractère de continuation pour que l'instruction puisse se poursuivre en ligne 523.

### Coupages de lignes - épilogue

Dans tous le cas où une instruction est répartie sur plusieurs lignes, c'est naturellement l'indentation de **la première ligne** qui est significative pour savoir à quel bloc rattacher cette instruction.

Notez bien enfin qu'on peut toujours mettre un *backslash* même lorsque ce n'est pas nécessaire, mais on évite cette pratique en règle générale car les *backslash* nuisent à la lisibilité.

## 2.12.2 Complément - niveau intermédiaire

### Outils liés à PEP008

Il existe plusieurs outils liés à la PEP0008, pour vérifier si votre code est conforme, ou même le modifier pour qu'il le devienne.

Ce qui nous donne un excellent prétexte pour parler un peu de <https://pypi.python.org>, qui est la plateforme qui distribue les logiciels disponibles via l'outil pip3.

Je vous signale notamment :

- <https://pypi.python.org/pypi/pep8/> pour vérifier, et
- <https://pypi.python.org/pypi/autopep8/> pour modifier automatiquement votre code et le rendre conforme.

### Les deux-points ':'

Dans un autre registre entièrement, vous pouvez **vous reporter à ce lien** si vous êtes intéressé par la question de savoir pourquoi on a choisi un délimiteur (le caractère deux-points :) pour terminer les instructions comme `if`, `for` et `def`.

## 2.13 L'instruction pass

### 2.13.1 Complément - niveau basique

Nous avons vu qu'en Python les blocs de code sont définis par leur indentation.

#### Une fonction vide

Cette convention a une limitation lorsqu'on essaie de définir un bloc vide. Voyons par exemple comment on définirait en C une fonction qui ne fait rien :

```
/* une fonction C qui ne fait rien */
void foo() {}
```

Comme en Python on n'a pas d'accolade pour délimiter les blocs de code, il existe une instruction pass, qui ne fait rien. À l'aide de cette instruction on peut à présent définir une fonction vide comme ceci :

```
In [1]: # une fonction Python qui ne fait rien
def foo():
    pass
```

#### Une boucle vide

Pour prendre un second exemple un peu plus pratique, et pour anticiper un peu sur l'instruction while que nous verrons très bientôt, voici un exemple d'une boucle vide, c'est à dire sans corps, qui permet de "dépiler" dans une liste jusqu'à l'obtention d'une certaine valeur :

```
In [2]: liste = list(range(10))
print('avant', liste)
while liste.pop() != 5:
    pass
print('après', liste)
```

```
avant [0, 1, 2, 3, 4, 5, 6, 7, 8, 9]
après [0, 1, 2, 3, 4]
```

On voit qu'ici encore l'instruction pass a toute son utilité.

### 2.13.2 Complément - niveau intermédiaire

#### Un if sans then

```
In [3]: # on utilise dans ces exemples une condition fausse
condition = False
```

Imaginons qu'on parte d'un code hypothétique qui fasse ceci :

```
In [4]: # la version initiale
if condition:
    print("non")
else:
    print("bingo")
```

```
bingo
```

Et que l'on veuille modifier ce code pour simplement supprimer l'impression de non. La syntaxe du langage **ne permet pas** de simplement commenter le premier print :

```
# si on commente le premier print  
# la syntaxe devient incorrecte  
if condition:  
    # print "non"  
else:  
    print "bingo"
```

Évidemment ceci pourrait être récrit autrement en inversant la condition, mais parfois on s'efforce de limiter au maximum l'impact d'une modification sur le code. Dans ce genre de situation on préférera écrire plutôt :

```
In [5]: # on peut s'en sortir en ajoutant une instruction pass  
        if condition:  
            # print "non"  
            pass  
        else:  
            print("bingo")
```

```
bingo
```

### Une classe vide

Enfin, comme on vient de le voir dans la vidéo, on peut aussi utiliser `pass` pour définir une classe vide comme ceci :

```
In [6]: class Foo:  
        pass
```

```
In [7]: foo = Foo()
```

## 2.14 Fonctions avec ou sans valeur de retour

### 2.14.1 Complément - niveau basique

#### Le style procédural

Une procédure est une fonction qui se contente de dérouler des instructions. Voici un exemple d'une telle fonction :

```
In [1]: def affiche_carre(n):  
        print("le carre de", n, "vaut", n*n)
```

qui s'utiliserait comme ceci :

```
In [2]: affiche_carre(12)
```

```
le carre de 12 vaut 144
```

#### Le style fonctionnel

Mais en fait, dans notre cas, il serait beaucoup plus commode de définir une fonction qui **retourne** le carré d'un nombre, afin de pouvoir écrire quelque chose comme :

```
surface = carre(15)
```

quitte à imprimer cette valeur ensuite si nécessaire. Jusqu'ici nous avons fait beaucoup appel à `print`, mais dans la pratique, imprimer n'est pas un but en soi.

#### L'instruction `return`

Voici comment on pourrait écrire une fonction `carre` qui **retourne** (on dit aussi **renvoie**) le carré de son argument :

```
In [3]: def carre(n):  
        return n*n  
  
        if carre(8) <= 100:  
            print('petit appartement')
```

```
petit appartement
```

La sémantique (le mot savant pour “comportement”) de l'instruction `return` est assez simple. La fonction qui est en cours d'exécution **s'achève** immédiatement, et l'objet cité dans l'instruction `return` est retourné à l'appelant, qui peut utiliser cette valeur comme n'importe quelle expression.

#### Le singleton `None`

Le terme même de fonction, si vous vous rappelez vos souvenirs de mathématiques, suggère qu'on calcule un résultat à partir de valeurs d'entrée. Dans la pratique il est assez rare qu'on définisse une fonction qui ne retourne rien.

En fait **toutes** les fonctions retournent quelque chose. Lorsque le programmeur n'a pas prévu d'instruction `return`, Python retourne un objet spécial, baptisé `None`. Voici par exemple ce qu'on obtient si on essaie d'afficher la valeur de retour de notre première fonction, qui, on le rappelle, ne retourne rien :

```
In [4]: # ce premier appel provoque l'impression d'une ligne
        retour = affiche_carre(15)
```

le carre de 15 vaut 225

```
In [5]: # voyons ce qu'a retourné la fonction affiche_carre
        print('retour =', retour)
```

retour = None

L'objet `None` est un singleton prédéfini par Python, un peu comme `True` et `False`. Ce n'est pas par contre une valeur booléenne, nous aurons l'occasion d'en reparler.

### Un exemple un peu plus réaliste

Pour illustrer l'utilisation de `return` sur un exemple plus utile, voyons le code suivant :

```
In [6]: def premier(n):
        """
        Retourne un booléen selon que n est premier ou non
        Retourne None pour les entrées négatives ou nulles
        """
        # retourne None pour les entrées non valides
        if n <= 0:
            return
        # traiter le cas singulier
        elif n == 1:
            return False
        # chercher un diviseur dans [2..n-1]
        # bien sûr on pourrait s'arrêter à la racine carrée de n
        # mais ce n'est pas notre sujet
        else:
            for i in range(2, n):
                if n % i == 0:
                    # on a trouvé un diviseur,
                    # on peut sortir de la fonction
                    return False
            # à ce stade, le nombre est bien premier
            return True
```

Cette fonction teste si un entier est premier ou non; il s'agit naturellement d'une version d'école, il existe d'autres méthodes beaucoup plus adaptées à cette tâche. On peut toutefois vérifier que cette version est fonctionnelle pour de petits entiers comme suit. On rappelle que 1 n'est pas considéré comme un nombre premier :

```
In [7]: for test in [-2, 1, 2, 4, 19, 35]:
        print(f"premier({test:2d}) = {premier(test)}")
```

```

premier(-2) = None
premier( 1) = False
premier( 2) = True
premier( 4) = False
premier(19) = True
premier(35) = False

```

**return sans valeur** Pour les besoins de cette discussion, nous avons choisi de retourner `None` pour les entiers négatifs ou nuls, une manière comme une autre de signaler que la valeur en entrée n'est pas valide.

Ceci n'est pas forcément une bonne pratique, mais elle nous permet ici d'illustrer que dans le cas où on ne mentionne pas de valeur de retour, Python retourne `None`.

**return interrompt la fonction** Comme on peut s'en convaincre en instrumentant le code - ce que vous pouvez faire à titre d'exercice en ajoutant des fonctions `print` - dans le cas d'un nombre qui n'est pas premier la boucle `for` ne va pas jusqu'à son terme.

On aurait pu d'ailleurs tirer profit de cette propriété pour écrire la fonction de manière légèrement différente comme ceci :

```

In [8]: def premier_sans_else(n):
        """
        Retourne un booléen selon que n est premier ou non
        Retourne None pour les entrées négatives ou nulles
        """
        # retourne None pour les entrées non valides
        if n <= 0:
            return
        # traiter le cas singulier
        if n == 1:
            return False
        # par rapport à la première version, on a supprimé
        # la clause else: qui est inutile
        for i in range(2, n):
            if n % i == 0:
                # on a trouve un diviseur
                return False
        # a ce stade c'est que le nombre est bien premier
        return True

```

C'est une question de style et de goût. En tout cas, les deux versions sont tout à fait équivalentes, comme on le voit ici :

```

In [9]: for test in [-2, 2, 4, 19, 35]:
        print(f"pour n = {test:2d} : premier → {premier(test)}\n"
              f"      premier_sans_else → {premier_sans_else(test)}\n")

```

```

pour n = -2 : premier → None
              premier_sans_else → None

```

```
pour n = 2 : premier → True
             premier_sans_else → True

pour n = 4 : premier → False
             premier_sans_else → False

pour n = 19 : premier → True
             premier_sans_else → True

pour n = 35 : premier → False
             premier_sans_else → False
```

**Digression sur les chaînes** Vous remarquerez dans cette dernière cellule, si vous regardez bien le paramètre de `print`, qu'on peut accoler deux chaînes (ici deux *f-strings*) sans même les ajouter ; un petit détail pour éviter d'alourdir le code :

```
In [10]: # quand deux chaînes apparaissent immédiatement
          # l'une après l'autre sans opérateur, elles sont concaténées
          "abc" "def"

Out[10]: 'abcdef'
```



## 2.15 Formatage des chaînes de caractères

### 2.15.1 Exercice - niveau basique

```
In [ ]: # charger l'exercice
        from exo_label import exo_label
```

Vous devez écrire une fonction qui prend deux arguments :

- une chaîne de caractères qui désigne le prénom d'un élève;
- un entier qui indique la note obtenue.

Elle devra retourner une chaîne de caractères selon que la note est

- $0 \leq \text{note} < 10$
- $10 \leq \text{note} < 16$
- $16 \leq \text{note} \leq 20$

comme on le voit sur les exemples :

```
In [ ]: exo_label.example()
```

```
In [ ]: # à vous de jouer
        def label(prenom, note):
            "votre code"
```

```
In [ ]: # pour corriger
        exo_label.correction(label)
```

## 2.16 Séquences

### 2.16.1 Exercice - niveau basique

#### Slicing

Commençons par créer une chaîne de caractères. Ne vous inquiétez pas si vous ne comprenez pas encore le code d'initialisation utilisé ci-dessous.

Pour les plus curieux, l'instruction `import` permet de charger dans votre programme une boîte à outils que l'on appelle un module. Python vient avec de nombreux modules qui forment la bibliothèque standard. Le plus difficile avec les modules de la bibliothèque standard est de savoir qu'ils existent. En effet, il y en a un grand nombre et bien souvent il existe un module pour faire ce que vous souhaitez.

Ici en particulier nous utilisons le module `string`.

```
In [1]: import string
        chaine = string.ascii_lowercase
        print(chaine)
```

```
abcdefghijklmnopqrstuvwxyz
```

Pour chacune des sous-chaînes ci-dessous, écrire une expression de slicing sur `chaine` qui renvoie la sous-chaîne. La cellule de code doit retourner `True`.

Par exemple, pour obtenir "def" :

```
In [2]: chaine[3:6] == "def"
```

```
Out[2]: True
```

1) Écrivez une slice pour obtenir "vwx" (n'hésitez pas à utiliser les indices négatifs) :

```
In [ ]: chaine[ <votre_code> ] == "vwx"
```

2) Une slice pour obtenir "wxyz" (avec une seule constante) :

```
In [ ]: chaine[ <votre_code> ] == "wxyz"
```

3) Une slice pour obtenir "dfhjlnprtvxz" (avec deux constantes) :

```
In [ ]: chaine[ <votre_code> ] == "dfhjlnprtvxz"
```

4) Une slice pour obtenir "xurolifc" (avec deux constantes) :

```
In [ ]: chaine[ <votre_code> ] == "xurolifc"
```

### 2.16.2 Exercice - niveau intermédiaire

#### Longueur

```
In [ ]: # il vous faut évaluer cette cellule magique
        # pour charger l'exercice qui suit
        # et autoévaluer votre réponse
        from corrections.exo_inconnue import exo_inconnue
```

On vous donne une chaîne `composite` dont on sait qu'elle a été calculée à partir de deux chaînes `inconnue` et `connue` comme ceci :

```
composite = connue + inconnue + connue
```

On vous donne également la chaîne `connue`. Imaginez par exemple que vous avez (ce ne sont pas les vraies valeurs) :

```
connue = '0bf1'
composite = '0bf1a9730e150bf1'
```

alors, dans ce cas :

```
inconnue = 'a9730e15'
```

L'exercice consiste à écrire une fonction qui retourne la valeur de `inconnue` à partir de celles de `composite` et `connue`. Vous pouvez admettre que `connue` n'est pas vide, c'est-à-dire qu'elle contient au moins un caractère.

Vous pouvez utiliser du *slicing*, et la fonction `len()`, qui retourne la longueur d'une chaîne :

```
In [3]: len('abcd')
```

```
Out[3]: 4
```

```
In [ ]: # à vous de jouer
def inconnue(composite, connue):
    "votre code"
```

Une fois votre code évalué, vous pouvez évaluer la cellule suivante pour vérifier votre résultat.

```
In [ ]: # correction
exo_inconnue.correction(inconnue)
```

Lorsque vous évaluez cette cellule, la correction vous montre :

- dans la première colonne l'appel qui est fait à votre fonction ;
- dans la seconde colonne la valeur attendue pour `inconnue` ;
- dans la troisième colonne ce que votre code a réellement calculé.

Si toutes les lignes sont **en vert** c'est que vous avez réussi cet exercice.

Vous pouvez essayer autant de fois que vous voulez, mais il vous faut alors à chaque itération :

- évaluer votre cellule-réponse (là où vous définissez la fonction `inconnue`) ;
- et ensuite évaluer la cellule correction pour la mettre à jour.

## 2.17 Listes

### 2.17.1 Exercice - niveau basique

```
In [ ]: from corrections.exo_laccess import exo_laccess
```

Vous devez écrire une fonction `laccess` qui prend en argument une liste, et qui retourne :

- `None` si la liste est vide ;
- sinon le dernier élément de la liste si elle est de taille paire ;
- et sinon l'élément du milieu.

```
In [ ]: exo_laccess.example()
```

```
In [ ]: # écrivez votre code ici
def laccess(liste):
    return "votre code"
```

```
In [ ]: # pour le corriger
exo_laccess.correction(laccess)
```

Une fois que votre code fonctionne, vous pouvez regarder si par hasard il marcherait aussi avec des chaînes :

```
In [ ]: from corrections.exo_laccess import exo_laccess_strings
```

```
In [ ]: exo_laccess_strings.correction(laccess)
```

## 2.18 Instruction if et fonction def

### 2.18.1 Exercice - niveau basique

#### Fonction de divisibilité

```
In [ ]: # chargement de l'exercice
        from corrections.exo_divisible import exo_divisible
```

L'exercice consiste à écrire une fonction baptisée `divisible` qui retourne une valeur booléenne, qui indique si un des deux arguments est divisible par l'autre.

Vous pouvez supposer les entrées `a` et `b` entiers et non nuls, mais pas forcément positifs.

```
In [ ]: def divisible(a, b):
        "<votre_code>"
```

Vous pouvez à présent tester votre code en évaluant ceci, qui écrira un message d'erreur si un des jeux de test ne donne pas le résultat attendu.

```
In [ ]: # tester votre code
        exo_divisible.correction(divisible)
```

### 2.18.2 Exercice - niveau basique

#### Fonction définie par morceaux

```
In [ ]: # chargement de l'exercice
        from corrections.exo_morceaux import exo_morceaux
```

On veut définir en Python une fonction qui est définie par morceaux :

$$f : x \rightarrow \begin{cases} -x - 5 & \text{si } x \leq -5 \\ 0 & \text{si } x \in [-5, 5] \\ \frac{1}{5}x - 1 & \text{si } x \geq 5 \end{cases}$$

```
In [ ]: # donc par exemple
        exo_morceaux.example()
```

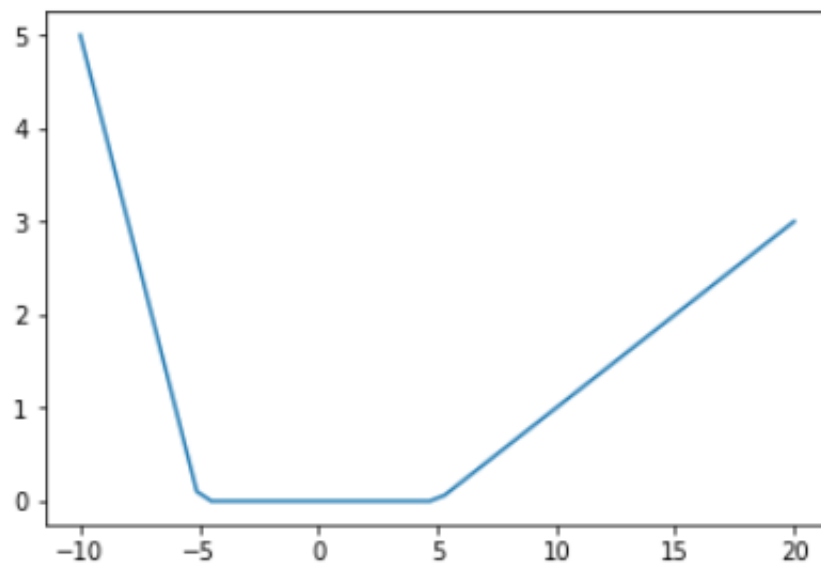
```
In [ ]: # à vous de jouer

        def morceaux(x):
            return 0 # "votre code"
```

```
In [ ]: # pour corriger votre code
        exo_morceaux.correction(morceaux)
```

**Représentation graphique** L'exercice est maintenant terminé, mais nous allons voir ensemble maintenant comment vous pourriez visualiser votre fonction.

Voici ce qui est attendu comme courbe pour `morceaux` (image fixe) :



En partant de votre code, vous pouvez produire votre propre courbe en utilisant numpy et matplotlib comme ceci :

```
In [ ]: # on importe les bibliothèques
import numpy as np
import matplotlib.pyplot as plt

In [ ]: # un échantillon des X entre -10 et 20
X = np.linspace(-10, 20)

# et les Y correspondants
Y = np.vectorize(morceaux)(X)

In [ ]: # on n'a plus qu'à dessiner
plt.plot(X, Y)
plt.show()
```

## 2.19 Comptage dans les chaines

### 2.19.1 Exercice - niveau basique

Nous remercions Benoit Izac pour cette contribution aux exercices.

### 2.19.2 La commande UNIX `wc(1)`

---

Sur les systèmes de type UNIX, la commande `wc` permet de compter le nombre de lignes, de mots et d'octets (ou de caractères) présents sur l'entrée standard ou contenus dans un fichier.

L'exercice consiste à écrire une fonction nommée `wc` qui prendra en argument une chaîne de caractères et retournera une liste contenant trois éléments :

1. le nombre de lignes (plus précisément le nombre de retours à la ligne);
2. le nombre de mots (un mot étant séparé par des espaces);
3. le nombre de caractères (on utilisera uniquement le jeu de caractères ASCII).

```
In [ ]: # chargement de l'exercice
        from corrections.exo_wc import exo_wc
```

```
In [ ]: # exemple
        exo_wc.example()
```

**Indice** : nous avons vu rapidement la boucle `for`, sachez toutefois qu'on peut tout à fait résoudre l'exercice en utilisant uniquement la bibliothèque standard.

**Remarque** : usuellement, ce genre de fonctions retournerait plutôt un tuple qu'une liste, mais comme nous ne voyons les tuples que la semaine prochaine..

À vous de jouer :

```
In [ ]: # la fonction à implémenter
        def wc(string):
            # remplacer pass par votre code
            pass
```

```
In [ ]: # correction
        exo_wc.correction(wc)
```

## 2.20 Compréhensions (1)

### 2.20.1 Exercice - niveau basique

#### Liste des valeurs d'une fonction

```
In [ ]: # Pour charger l'exercice
        from corrections.exo_liste_p import exo_liste_P
```

On se donne une fonction polynomiale :

$$P(x) = 2x^2 - 3x - 2$$

On vous demande d'écrire une fonction `liste_P` qui prend en argument une liste de nombres réels  $x$  et qui retourne la liste des valeurs  $P(x)$ .

```
In [ ]: # voici un exemple de ce qui est attendu
        exo_liste_P.example()
```

Écrivez votre code dans la cellule suivante (On vous suggère d'écrire une fonction  $P$  qui implémente le polynôme mais ça n'est pas strictement indispensable, seul le résultat de `liste_P` compte) :

```
In [ ]: def P(x):
        "<votre code>"

        def liste_P(liste_x):
            "votre code"
```

Et vous pouvez le vérifier en évaluant cette cellule :

```
In [ ]: # pour vérifier votre code
        exo_liste_P.correction(liste_P)
```

---

### 2.20.2 Récréation

Si vous avez correctement implémenté la fonction `liste_P` telle que demandé dans le premier exercice, vous pouvez visualiser le polynôme  $P$  en utilisant `matplotlib` avec le code suivant :

```
In [ ]: # on importe les bibliothèques
        import numpy as np
        import matplotlib.pyplot as plt

In [ ]: # un échantillon des X entre -10 et 10
        X = np.linspace(-10, 10)

        # et les Y correspondants
        Y = liste_P(X)

In [ ]: # on n'a plus qu'à dessiner
        plt.plot(X, Y)
        plt.show()
```



## 2.21 Compréhensions (2)

### 2.21.1 Exercice - niveau intermédiaire

#### Mise au carré

```
In [ ]: # chargement de l'exercice
        from corrections.exo_carre import exo_carre
```

On vous demande à présent d'écrire une fonction dans le même esprit que ci-dessus. Cette fois, chaque ligne contient, séparés par des points-virgules, une liste d'entiers, et on veut obtenir une nouvelle chaîne avec les carrés de ces entiers, séparés par des deux-points.

À nouveau les lignes peuvent être remplies de manière approximative, avec des espaces, des tabulations, ou même des points-virgules en trop, que ce soit au début, à la fin, ou au milieu d'une ligne.

```
In [ ]: # exemples
        exo_carre.example()
```

```
In [ ]: # écrivez votre code ici
        def carre(ligne):
            "<votre_code>"
```

```
In [ ]: # pour corriger
        exo_carre.correction(carre)
```



# Chapitre 3

Renforcement des notions de base,  
références partagées

## 3.1 Les fichiers

### 3.1.1 Complément - niveau basique

Voici quelques utilisations habituelles du type fichier en Python.

#### Avec un context manager

Nous avons vu dans la vidéo les mécanismes de base sur les fichiers. Nous avons vu notamment qu'il est important de bien fermer un fichier après usage. On a vu aussi qu'il est recommandé de **toujours** utiliser l'instruction `with` et de contrôler son encodage. Il est donc recommandé de faire :

```
In [1]: # avec un 'with' on garantit la fermeture du fichier
        with open("foo.txt", "w", encoding='utf-8') as sortie:
            for i in range(2):
                sortie.write(f"{i}\n")
```

#### Les modes d'ouverture

Les modes d'ouverture les plus utilisés sont :

- 'r' (la chaîne contenant l'unique caractère r) pour ouvrir un fichier en lecture seulement;
- 'w' en écriture seulement; le contenu précédent du fichier, s'il existait, est perdu;
- 'a' en écriture seulement; mais pour ajouter du contenu en fin de fichier.

Voici par exemple comment on pourrait ajouter deux lignes de texte dans le fichier `foo.txt` qui contient, à ce stade du notebook, deux entiers :

```
In [2]: # on ouvre le fichier en mode 'a' comme append (= ajouter)
        with open("foo.txt", "a", encoding='utf-8') as sortie:
            for i in range(100, 102):
                sortie.write(f"{i}\n")
```

```
In [3]: # maintenant on regarde ce que contient le fichier
        with open("foo.txt", encoding='utf-8') as entree: # remarquez que sans 'mode', on ouv
            for line in entree:
                # line contient déjà un retour à la ligne
                print(line, end='')
```

```
0
1
100
101
```

Il existe de nombreuses variantes au mode d'ouverture, pour par exemple :

- ouvrir le fichier en lecture *et* en écriture (mode +);
- ouvrir le fichier en mode binaire (mode b).

Ces variantes sont décrites dans [la section sur la fonction built-in open](#) dans la documentation Python.

### 3.1.2 Complément - niveau intermédiaire

#### Un fichier est un itérateur

Nous reparlerons des notions d'itérable et d'itérateur dans les semaines suivantes. Pour l'instant, on peut dire qu'un fichier - qui donc **est itérable** puisqu'on peut le lire par une boucle `for` - est aussi **son propre itérateur**. Cela implique que l'on ne peut le parcourir qu'une fois dans une boucle `for`. Pour le reparcourir, il faut le fermer et l'ouvrir de nouveau.

```
In [4]: # un fichier est son propre itérateur

In [5]: with open("foo.txt", encoding='utf-8') as entree:
        print(entree.__iter__() is entree)

True
```

Par conséquent, écrire deux boucles `for` imbriquées sur **le même objet fichier** ne **fonctionnerait pas** comme on pourrait s'y attendre.

```
In [6]: # Si l'on essaie d'écrire deux boucles imbriquées
        # sur le même objet fichier, le résultat est inattendu
        with open("foo.txt", encoding='utf-8') as entree:
            for l1 in entree:
                # on enlève les fins de ligne
                l1 = l1.strip()
                for l2 in entree:
                    # on enlève les fins de ligne
                    l2 = l2.strip()
                    print(l1, "x", l2)

0 x 1
0 x 100
0 x 101
```

### 3.1.3 Complément - niveau avancé

#### Autres méthodes

Vous pouvez également accéder à des fonctions de beaucoup plus bas niveau, notamment celle fournies directement par le système d'exploitation ; nous allons en décrire deux parmi les plus utiles.

#### Digression - `repr()`

Comme nous allons utiliser maintenant des outils d'assez bas niveau pour lire du texte, pour examiner ce texte nous allons utiliser la fonction `repr()`, et voici pourquoi :

```
In [7]: # construisons à la main une chaîne qui contient deux lignes
        lines = "abc" + "\n" + "def" + "\n"

In [8]: # si on l'imprime on voit bien les retours à la ligne
        # d'ailleurs on sait qu'il n'est pas utile
        # d'ajouter un retour à la ligne à la fin
        print(lines, end="")
```

```
abc
def
```

```
In [9]: # vérifions que repr() nous permet de bien
        # voir le contenu de cette chaîne
        print(repr(lines))
```

```
'abc\ndef\n'
```

### Lire un contenu - bas niveau

Revenons aux fichiers; la méthode `read()` permet de lire dans le fichier un buffer d'une certaine taille :

```
In [10]: # read() retourne TOUT le contenu
        # ne pas utiliser avec de très gros fichiers bien sûr

        # une autre façon de montrer tout le contenu du fichier
        with open("foo.txt", encoding='utf-8') as entree:
            full_contents = entree.read()
            print(f"Contenu complet\n{full_contents}", end="")
```

```
Contenu complet
```

```
0
1
100
101
```

```
In [11]: # lire dans le fichier deux blocs de quatre caractères
        with open("foo.txt", encoding='utf-8') as entree:
            for bloc in range(2):
                print(f"Bloc {bloc} >>{repr(entree.read(4))}<<")
```

```
Bloc 0 >>'0\n1\n'<<
```

```
Bloc 1 >>'100\n'<<
```

On voit donc que chaque bloc contient bien quatre caractères en comptant les sauts de ligne :

bloc #	contenu
0	un 0, un <i>newline</i> , un 1, un <i>newline</i>
1	un 1, deux 0, un <i>newline</i>

### La méthode `flush`

Les entrées-sorties sur fichier sont bien souvent *bufferisées* par le système d'exploitation. Cela signifie qu'un appel à `write` ne provoque pas forcément une écriture immédiate, car pour des

raisons de performance on attend d'avoir suffisamment de matière avant d'écrire sur le disque.

Il y a des cas où ce comportement peut s'avérer gênant, et où on a besoin d'écrire immédiatement (et donc de vider le *buffer*), et c'est le propos de la méthode `flush`.

### Fichiers textuels et fichiers binaires

De la même façon que le langage propose les deux types `str` et `bytes`, il est possible d'ouvrir un fichier en mode *textuel* ou en mode *binaire*.

Les fichiers que nous avons vus jusqu'ici étaient ouverts en mode *textuel* (c'est le défaut), et c'est pourquoi nous avons interagi avec eux avec des objets de type `str` :

```
In [12]: # un fichier ouvert en mode textuel nous donne des str
         with open('foo.txt', encoding='utf-8') as input:
             for line in input:
                 print("on a lu un objet de type", type(line))
```

```
on a lu un objet de type <class 'str'>
on a lu un objet de type <class 'str'>
on a lu un objet de type <class 'str'>
on a lu un objet de type <class 'str'>
```

Lorsque ce n'est pas le comportement souhaité, on peut :

- ouvrir le fichier en mode *binaire* - pour cela on ajoute le caractère `b` au mode d'ouverture ;
- et on peut alors interagir avec le fichier avec des objets de type `bytes`

Pour illustrer ce trait, nous allons : 0. créer un fichier en mode texte, et y insérer du texte en UTF-8 ; 0. relire le fichier en mode binaire, et retrouver le codage des différents caractères.

```
In [13]: # phase 1 : on écrit un fichier avec du texte en UTF-8
         # on ouvre donc le fichier en mode texte
         # en toute rigueur il faut préciser l'encodage,
         # si on ne le fait pas il sera déterminé
         # à partir de vos réglages système
         with open('strbytes', 'w', encoding='utf-8') as output:
             output.write("déjà l'été\n")

In [14]: # phase 2: on ouvre le fichier en mode binaire
         with open('strbytes', 'rb') as rawinput:
             # on lit tout le contenu
             octets = rawinput.read()
             # qui est de type bytes
             print("on a lu un objet de type", type(octets))
             # si on regarde chaque octet un par un
             for i, octet in enumerate(octets):
                 print(f"{i} → {repr(chr(octet))} [{hex(octet)}]")
```

```
on a lu un objet de type <class 'bytes'>
0 → 'd' [0x64]
1 → 'Ã' [0xc3]
```

```

2 → '©' [0xa9]
3 → 'j' [0x6a]
4 → 'Ã' [0xc3]
5 → '\xa0' [0xa0]
6 → ' ' [0x20]
7 → 'l' [0x6c]
8 → '"' [0x27]
9 → 'Ã' [0xc3]
10 → '©' [0xa9]
11 → 't' [0x74]
12 → 'Ã' [0xc3]
13 → '©' [0xa9]
14 → '\r' [0xd]
15 → '\n' [0xa]

```

Vous retrouvez ainsi le fait que l’unique caractère Unicode é a été encodé par UTF-8 sous la forme de deux octets de code hexadécimal 0xc3 et 0xa9.

Vous pouvez également consulter ce site qui visualise l’encodage UTF-8, avec notre séquence d’entrée :

<https://mothereff.in/utf-8#d%C3%A9%C3%A0%20l%27%C3%A9%C3%A9%0A>

```

In [15]: # on peut comparer le nombre d'octets et le nombre de caractères
         with open('strbytes', encoding='utf-8') as textfile:
             print(f"en mode texte, {len(textfile.read())} caractères")
         with open('strbytes', 'rb') as binfile:
             print(f"en mode binaire, {len(binfile.read())} octets")

```

```

en mode texte, 11 caractères
en mode binaire, 16 octets

```

Ce qui correspond au fait que nos quatre caractères non-ASCII (3 x é et 1 x à) sont tous encodés par UTF-8 comme deux octets, comme vous pouvez vous en assurer [ici pour é](#) et [là pour à](#).

### Pour en savoir plus

Pour une description exhaustive vous pouvez vous reporter :

- au [glossaire sur la notion de object file](#),
- et aussi et surtout [au module io](#) qui décrit plus en détail les fonctionnalités disponibles.



## 3.2 Fichiers et utilitaires

### 3.2.1 Complément - niveau basique

Outre les objets fichiers créés avec la fonction `open`, comme on l’a vu dans la vidéo, et qui servent à lire et écrire à un endroit précis, une application a besoin d’un minimum d’utilitaires pour **parcourir l’arborescence de répertoires et fichiers**, c’est notre propos dans ce complément.

#### Le module `os.path` (obsolète)

Avant la version python-3.4, la librairie standard offrait une conjonction d’outils pour ce type de fonctionnalités :

- le module `os.path`, pour faire des calculs sur les chemins et noms de fichiers [doc](#),
- le module `os` pour certaines fonctions complémentaires comme renommer ou détruire un fichier [doc](#),
- et enfin le module `glob` pour la recherche de fichiers, par exemple pour trouver tous les fichiers en `*.txt` [doc](#).

Cet ensemble un peu disparate a été remplacé par une **librairie unique** `pathlib`, qui fournit toutes ces fonctionnalités sous un interface unique et moderne, que nous **recommandons** évidemment d’utiliser pour **du nouveau code**.

Avant d’aborder `pathlib`, voici un très bref aperçu de ces trois anciens modules, pour le cas - assez probable - où vous les rencontreriez dans du code existant; tous les noms qui suivent correspondent à des **fonctions** - par opposition à `pathlib` qui, comme nous allons le voir, offre une interface orientée objet :

- `os.path.join` ajoute `‘/’` ou `‘\’` entre deux morceaux de chemin, selon l’OS
- `os.path.basename` trouve le nom de fichier dans un chemin
- `os.path.dirname` trouve le nom du directory dans un chemin
- `os.path.abspath` calcule un chemin absolu, c’est-à-dire à partir de la racine du filesystem
- `os.path.exists` pour savoir si un chemin existe ou pas (fichier ou répertoire)
- `os.path.isfile` (et `isdir`) pour savoir si un chemin est un fichier (et un répertoire)
- `os.path.getsize` pour obtenir la taille du fichier
- `os.path.getatime` et aussi `getmtime` et `getctime` pour obtenir les dates de création/modification d’un fichier
- `os.remove` (ou son ancien nom `os.unlink`), qui permet de supprimer un fichier
- `os.rmdir` pour supprimer un répertoire (mais qui doit être vide)
- `os.removedirs` pour supprimer tout un répertoire avec son contenu, récursivement si nécessaire
- `os.rename` pour renommer un fichier
- `glob.glob` comme dans par exemple `glob.glob("*.txt")`

#### Le module `pathlib`

C’est la méthode recommandée aujourd’hui pour travailler sur les fichiers et répertoires.

#### Orienté Objet

Comme on l’a mentionné `pathlib` offre une interface orientée objet; mais qu’est-ce

que ça veut dire au juste ?

Ceci nous donne un prétexte pour une première application pratique des notions de module (que nous avons introduits en fin de semaine 2) et de classe (que nous allons voir en fin de semaine).

De même que le langage nous propose les types *builtin* `int` et `str`, le module `pathlib` nous expose **un type** (on dira plutôt **une classe**) qui s'appelle `Path`, que nous allons importer comme ceci :

```
In [1]: from pathlib import Path
```

Nous allons faire tourner un petit scénario qui va créer un fichier :

```
In [2]: # le nom de notre fichier jouet
        nom = 'fichier-temoin'
```

Pour commencer, nous allons vérifier si le fichier en question existe.

Pour ça nous créons un **objet** qui est une **instance** de la classe `Path`, comme ceci :

```
In [3]: # on crée un objet de la classe Path, associé au nom de fichier
        path = Path(nom)
```

Vous remarquez que c'est consistant avec par exemple :

```
In [4]: # transformer un float en int
        i = int(3.5)
```

en ce sens que le type (`int` ou `Path`) se comporte comme une usine pour créer des objets du type en question.

Quoi qu'il en soit, cet objet `path` offre un certain nombre de méthodes ; pour les voir puisque nous sommes dans un notebook, je vous invite dans la cellule suivante à utiliser l'aide en ligne en appuyant sur la touche 'Tabulation' après avoir ajouté un `.` comme si vous alliez envoyer une méthode à cet objet

```
path.[taper la touche TAB]
```

et le notebook vous montrera la liste des méthodes disponibles.

```
In [5]: # ajouter un . et utilisez la touche <Tabulation>
        path.chmod
```

```
Out[5]: <bound method Path.chmod of WindowsPath('fichier-temoin')>
```

Ainsi par exemple on peut savoir si le fichier existe avec la méthode `exists()`

```
In [6]: # au départ le fichier n'existe pas
        path.exists()
```

```
Out[6]: False
```

```
In [7]: # si j'écris dedans je le crée
        with open(nom, 'w', encoding='utf-8') as output:
            output.write('0123456789\n')
```

```
In [8]: # et maintenant il existe
        path.exists()
```

```
Out[8]: True
```

### Métadonnées

Voici quelques exemples qui montrent comment accéder aux métadonnées de ce fichier :

```
In [9]: # cette méthode retourne (en un seul appel système) les métadonnées agrégées
        path.stat()
```

```
Out[9]: os.stat_result(st_mode=33206, st_ino=2251799814156457, st_dev=571766744, st_nlink=1,
```

Pour ceux que ça intéresse, l'objet retourné par cette méthode stat est un `namedtuple`, que l'on va voir très bientôt.

On accède aux différentes informations comme ceci :

```
In [10]: # la taille du fichier en octets est de 11
         # car il faut compter un caractère "newline" en fin de ligne
         path.stat().st_size
```

```
Out[10]: 12
```

```
In [11]: # la date de dernière modification, sous forme d'un entier
         # c'est le nombre de secondes depuis le 1er Janvier 1970
         mtime = path.stat().st_mtime
         mtime
```

```
Out[11]: 1537703765.728249
```

```
In [12]: # que je peux rendre lisible comme ceci
         # en anticipant sur le module datetime
         from datetime import datetime
         mtime_datetime = datetime.fromtimestamp(mtime)
         mtime_datetime
```

```
Out[12]: datetime.datetime(2018, 9, 23, 13, 56, 5, 728249)
```

```
In [13]: # ou encore, si je formate pour n'obtenir que
         # l'heure et la minute
         f"{mtime_datetime:%H:%M}"
```

```
Out[13]: '13:56'
```

### Détruire un fichier

```
In [14]: # je peux maintenant détruire le fichier
         path.unlink()
```

```
In [15]: # ou encore mieux, si je veux détruire
         # seulement dans le cas où il existe je peux aussi faire
         try:
             path.unlink()
         except FileNotFoundError:
             print("no need to remove")
```

```
no need to remove
```

```
In [16]: # et maintenant il n'existe plus
         path.exists()
```

```
Out[16]: False
```

```
In [17]: # je peux aussi retrouver le nom du fichier comme ceci
         # attention ce n'est pas une méthode mais un attribut
         # c'est pourquoi il n'y a pas de parenthèses
         path.name
```

```
Out[17]: 'fichier-temoin'
```

### Recherche de fichiers

Maintenant je voudrais connaître la liste des fichiers de nom \*.json dans le directory data.

La méthode la plus naturelle consiste à créer une instance de Path associée au directory lui-même :

```
In [18]: dirpath = Path('./data/')
```

Sur cet objet la méthode glob nous retourne un itérable qui contient ce qu'on veut :

```
In [19]: # tous les fichiers *.json dans le répertoire data/
         for json in dirpath.glob("*.json"):
             print(json)
```

### Documentation complète

Voyez [la documentation complète ici](#)

## 3.2.2 Complément - niveau avancé

Pour ceux qui sont déjà familiers avec les classes, j'en profite pour vous faire remarquer le type de notre objet path

```
In [20]: type(path)
```

```
Out[20]: pathlib.WindowsPath
```

qui n'est pas Path, mais en fait une sous-classe de Path qui est - sur la plateforme du MOOC au moins, qui fonctionne sous linux - un objet de type PosixPath, qui est une sous-classe de Path, comme vous pouvez le voir :

```
In [21]: from pathlib import PosixPath
         issubclass(PosixPath, Path)
```

```
Out[21]: True
```

Ce qui fait que mécaniquement, path est bien une instance de Path

```
In [22]: isinstance(path, Path)
```

```
Out[22]: True
```

ce qui est heureux puisqu'on avait utilisé Path() pour construire l'objet path au départ :)

## 3.3 Formats de fichiers : JSON et autres

### 3.3.1 Compléments - niveau basique

Voici quelques mots sur des outils Python fournis dans la bibliothèque standard, et qui permettent de lire ou écrire des données dans des fichiers.

#### Le problème

Les données dans un programme Python sont stockées en mémoire (la RAM), sous une forme propice aux calculs. Par exemple un petit entier est fréquemment stocké en binaire dans un mot de 64 bits, qui est prêt à être soumis au processeur pour faire une opération arithmétique.

Ce format ne se prête pas forcément toujours à être transposé tel quel lorsqu'on doit écrire des données sur un support plus pérenne, comme un disque dur, ou encore sur un réseau pour transmission distante - ces deux supports étant à ce point de vue très voisins.

Ainsi par exemple il pourra être plus commode d'écrire notre entier sur disque, ou de le transmettre à un programme distant, sous une forme décimale qui sera plus lisible, sachant que par ailleurs toutes les machines ne codent pas un entier de la même façon.

Il convient donc de faire de la traduction dans les deux sens entre représentations d'une part en mémoire, et d'autre part sur disque ou sur réseau (à nouveau, on utilise en général les mêmes formats pour ces deux usages).

#### Le format JSON

Le format sans aucun doute le plus populaire à l'heure actuelle est [le format JSON](#) pour *JavaScript Object Notation*.

Sans trop nous attarder nous dirons que JSON est un encodage - en anglais [marshalling](#) - qui se prête bien à la plupart des types de base que l'on trouve dans les langages modernes comme Python, Ruby ou JavaScript.

La bibliothèque standard de Python contient [le module json](#) que nous illustrons très rapidement ici :

```
In [1]: import json
```

```
# En partant d'une donnée construite à partir de types de base
data = [
    # des types qui ne posent pas de problème
    [1, 2, 'a', [3.23, 4.32], {'eric': 32, 'jean': 43}],
    # un tuple
    (1, 2, 3),
]

# sauver ceci dans un fichier
with open("s1.json", "w", encoding='utf-8') as json_output:
    json.dump(data, json_output)
```

```
# et relire le résultat
with open("s1.json", encoding='utf-8') as json_input:
    data2 = json.load(json_input)
```

### Limitations de json

Certains types de base ne sont pas supportés par le format JSON (car ils ne sont pas natifs en JavaScript), c'est le cas notamment pour :

- tuple, qui se fait encoder comme une liste;
- complex, set et frozenset, que l'on ne peut pas encoder du tout (sans étendre la bibliothèque).

C'est ce qui explique ce qui suit :

```
In [2]: # le premier élément de data est intact,
        # comme si on avait fait une *deep copy* en fait
        print("première partie de data", data[0] == data2[0])
```

première partie de data True

```
In [3]: # par contre le `tuple` se fait encoder comme une `list`
        print("deuxième partie", "entrée", type(data[1]), "sortie", type(data2[1]))
```

deuxième partie entrée <class 'tuple'> sortie <class 'list'>

Malgré ces petites limitations, ce format est de plus en plus populaire, notamment parce qu'on peut l'utiliser pour communiquer avec des applications Web écrites en JavaScript, et aussi parce qu'il est très léger, et supporté par de nombreux langages.

## 3.3.2 Compléments - niveau intermédiaire

### Le format csv

Le format csv pour *Comma Separated Values*, originaire du monde des tableurs, peut rendre service à l'occasion, il est proposé [dans le module csv](#).

### Le format pickle

Le format pickle remplit une fonctionnalité très voisine de JSON, mais est spécifique à Python. C'est pourquoi, malgré des limites un peu moins sévères, son usage tend à rester plutôt marginal pour l'échange de données, on lui préfère en général le format JSON.

Par contre, pour la sauvegarde locale d'objets Python (pour, par exemple, faire des points de reprises d'un programme), il est très utile. Il est implémenté [dans le module pickle](#).

### Le format XML

Vous avez aussi très probablement entendu parler de XML, qui est un format assez populaire également.

Cela dit, la puissance, et donc le coût, de XML et JSON ne sont pas du tout comparables, XML étant beaucoup plus flexible mais au prix d'une complexité de mise en œuvre très supérieure.

Il existe plusieurs souches différentes de bibliothèques prenant en charge le format XML, [qui sont introduites ici](#).

**Pour en savoir plus**

Voyez la page sur [les formats de fichiers](#) dans la documentation Python.

## 3.4 Fichiers systèmes

### 3.4.1 Complément - niveau avancé

Dans ce complément, nous allons voir comment un programme Python interagit avec ce qu'il est convenu d'appeler le système d'entrées-sorties standard du système d'exploitation.

#### Introduction

Dans un ordinateur, le système d'exploitation (Windows, Linux, macOS, etc.) comprend un noyau (*kernel*) qui est un logiciel qui a l'exclusivité pour interagir physiquement avec le matériel (processeur(s), mémoire, disque(s), périphériques, etc.); il offre aux programmes utilisateur (*userspace*) des abstractions pour interagir avec ce matériel.

La notion de fichier, telle qu'on l'a vue dans la vidéo, correspond à une de ces abstractions ; elle repose principalement sur les quatre opérations élémentaires suivantes :

- open;
- close;
- read;
- write.

Parmi les autres conventions d'interaction entre le système (pour être précis : le *shell*) et une application, il y a les notions de :

- entrée standard (*standard input*, en abrégé *stdin*);
- sortie standard (*standard output*, en abrégé *stdout*);
- erreur standard (*standard error*, en abrégé *stderr*).

Ceci est principalement pertinent dans le contexte d'un terminal. L'idée c'est que l'on a envie de pouvoir *rediriger les entrées-sorties* d'un programme sans avoir à le modifier. De la sorte, on peut également *chaîner* des traitements *à l'aide de pipes*, sans avoir besoin de sauver les résultats intermédiaires sur disque.

Ainsi par exemple lorsque l'on écrit :

```
$ monprogramme < fichier_entree > fichier_sortie
```

Les deux fichiers en question sont ouverts par le *shell*, et passés à *monprogramme* - que celui-ci soit écrit en C, en Python ou en Java - sous la forme des fichiers *stdin* et *stdout* respectivement, et donc **déjà ouverts**.

#### Le module `sys`

L'interpréteur Python vous expose ces trois fichiers sous la forme d'attributs du module `sys` :

```
In [1]: import sys
        for channel in (sys.stdin, sys.stdout, sys.stderr):
            print(channel)

<_io.TextIOWrapper name='<stdin>' mode='r' encoding='cp1252'>
<ipykernel.iostream.OutStream object at 0x041774F0>
<ipykernel.iostream.OutStream object at 0x04177CF0>
```



Dans le contexte du notebook vous pouvez constater que les deux flux de sortie sont implémentés comme des classes spécifiques à IPython. Si vous exécutez ce code localement dans votre ordinateur vous allez sans doute obtenir quelque chose comme :

```
<_io.TextIOWrapper name='<stdin>' mode='r' encoding='UTF-8'>
<_io.TextIOWrapper name='<stdout>' mode='w' encoding='UTF-8'>
<_io.TextIOWrapper name='<stderr>' mode='w' encoding='UTF-8'>
```

On n'a pas extrêmement souvent besoin d'utiliser ces variables en règle générale, mais elles peuvent s'avérer utiles dans des contextes spécifiques.

Par exemple, l'instruction `print` écrit dans `sys.stdout` (c'est-à-dire la sortie standard). Et comme `sys.stdout` est une variable (plus exactement `stdout` est un attribut dans le module référencé par la variable `sys`) et qu'elle référence un objet fichier, on peut lui faire référencer un autre objet fichier et ainsi rediriger depuis notre programme tous les sorties, qui sinon iraient sur le terminal, vers un fichier de notre choix :

```
In [2]: # ici je fais exprès de ne pas utiliser un `with`
        # car très souvent les deux redirections apparaissent
        # dans des fonctions différentes
import sys
        # on ouvre le fichier destination
autre_stdout = open('ma_sortie.txt', 'w', encoding='utf-8')
        # on garde un lien vers le fichier sortie standard
        # pour le réinstaller plus tard si besoin.
tmp = sys.stdout
print('sur le terminal')

        # première redirection
sys.stdout = autre_stdout
print('dans le fichier')

        # on remet comme c'était au début
sys.stdout = tmp
        # et alors pour être propre on n'oublie pas de fermer
autre_stdout.close()
print('de nouveau sur le terminal')
```

```
sur le terminal
de nouveau sur le terminal
```

```
In [3]: # et en effet, dans le fichier on a bien
        with open("ma_sortie.txt", encoding='utf-8') as check:
            print(check.read())
```

```
dans le fichier
```

## 3.5 La construction de tuples

### 3.5.1 Complément - niveau intermédiaire

#### Les tuples et la virgule terminale

Comme on l’a vu dans la vidéo, on peut construire un tuple à deux éléments - un couple - de quatre façons :

```
In [1]: # sans parenthèse ni virgule terminale
couple1 = 1, 2
# avec parenthèses
couple2 = (1, 2)
# avec virgule terminale
couple3 = 1, 2,
# avec parenthèses et virgule
couple4 = (1, 2,)

In [2]: # toutes ces formes sont équivalentes ; par exemple
couple1 == couple4
```

Out [2]: True

Comme on le voit :

- en réalité la **parenthèse est parfois superflue** ; mais il se trouve qu’elle est **largement utilisée** pour améliorer la lisibilité des programmes, sauf dans le cas du *tuple unpacking* ; nous verrons aussi plus bas qu’elle est **parfois nécessaire** selon l’endroit où le tuple apparaît dans le programme ;
- la **dernière virgule est optionnelle** aussi, c’est le cas pour les tuples à au moins 2 éléments - nous verrons plus bas le cas des tuples à un seul élément.

#### Conseil pour la présentation sur plusieurs lignes

En général d’ailleurs, la forme avec parenthèses et virgule terminale est plus pratique. Considérez par exemple l’initialisation suivante ; on veut créer un tuple qui contient des listes (naturellement un tuple peut contenir n’importe quel objet Python), et comme c’est assez long on préfère mettre un élément du tuple par ligne :

```
In [3]: mon_tuple = ([1, 2, 3],
                     [4, 5, 6],
                     [7, 8, 9],
                     )
```

L’avantage lorsqu’on choisit cette forme (avec parenthèses, et avec virgule terminale), c’est d’abord qu’il n’est pas nécessaire de mettre un backslash à la fin de chaque ligne ; parce que l’on est à l’intérieur d’une zone parenthésée, l’interpréteur Python “sait” que l’instruction n’est pas terminée et va se continuer sur la ligne suivante.

Deuxièmement, si on doit ultérieurement ajouter ou enlever un élément dans le tuple, il suffira d’enlever ou d’ajouter toute une ligne, sans avoir à s’occuper des virgules ; si on avait choisi de ne pas faire figurer la virgule terminale, alors pour ajouter un élément dans le tuple après le dernier, il ne faut pas oublier d’ajouter une virgule à la ligne précédente. Cette simplicité se répercute au niveau du gestionnaire de code source, où les différences dans le

code sont plus faciles à visualiser.

Signalons enfin que ceci n'est pas propre aux tuples. La virgule terminale est également optionnelle pour les listes, ainsi d'ailleurs que pour tous les types Python où cela fait du sens, comme les dictionnaires et les ensembles que nous verrons bientôt. Et dans tous les cas où on opte pour une présentation multi-lignes, il est conseillé de faire figurer une virgule terminale.

### Tuples à un élément

Pour revenir à présent sur le cas des tuples à un seul élément, c'est un cas particulier, parmi les quatre syntaxes que l'on a vues ci-dessus, on obtiendrait dans ce cas :

```
In [4]: # ATTENTION : ces deux premières formes ne construisent pas un tuple !
        simple1 = 1
        simple2 = (1)
        # celles-ci par contre construisent bien un tuple
        simple3 = 1,
        simple4 = (1,)
```

- Il est bien évident que la première forme ne crée pas de tuple ;
- et en fait la seconde non plus, car Python lit ceci comme une expression parenthésée, avec seulement un entier.

Et en fait ces deux premières formes créent un entier simple :

```
In [5]: type(simple2)
```

```
Out[5]: int
```

Les deux autres formes créent par contre toutes les deux un tuple à un élément comme on cherchait à le faire :

```
In [6]: type(simple3)
```

```
Out[6]: tuple
```

```
In [7]: simple3 == simple4
```

```
Out[7]: True
```

Pour conclure, disons donc qu'il est conseillé de **toujours mentionner une virgule terminale** lorsqu'on construit des tuples.

### Parenthèse parfois obligatoire

Dans certains cas vous vous apercevrez que la parenthèse est obligatoire. Par exemple on peut écrire :

```
In [8]: x = (1,)
        (1,) == x
```

```
Out[8]: True
```

Mais si on essaie d'écrire le même test sans les parenthèses :

```
In [9]: # ceci provoque une SyntaxError
        1, == x

        File "<ipython-input-9-219921aa55c4>", line 2
        1, == x
            ^
SyntaxError: invalid syntax
```

Python lève une erreur de syntaxe; encore une bonne raison pour utiliser les parenthèses.

### Addition de tuples

Bien que le type tuple soit immuable, il est tout à fait légal d'additionner deux tuples, et l'addition va produire un **nouveau** tuple :

```
In [10]: tuple1 = (1, 2,)
         tuple2 = (3, 4,)
         print('addition', tuple1 + tuple2)

addition (1, 2, 3, 4)
```

Ainsi on peut également utiliser l'opérateur += avec un tuple qui va créer, comme précédemment, un nouvel objet tuple :

```
In [11]: tuple1 = (1, 2,)
         tuple1 += (3, 4,)
         print('apres ajout', tuple1)

apres ajout (1, 2, 3, 4)
```

### Construire des tuples élaborés

Malgré la possibilité de procéder par additions successives, la construction d'un tuple peut s'avérer fastidieuse.

Une astuce utile consiste à penser aux fonctions de conversion, pour construire un tuple à partir de - par exemple - une liste. Ainsi on peut faire par exemple ceci :

```
In [12]: # on fabrique une liste pas à pas
         liste = list(range(10))
         liste[9] = 'Inconnu'
         del liste [2:5]
         liste

Out[12]: [0, 1, 5, 6, 7, 8, 'Inconnu']

In [13]: # on convertit le résultat en tuple
         mon_tuple = tuple(liste)
         mon_tuple

Out[13]: (0, 1, 5, 6, 7, 8, 'Inconnu')
```

**Digression sur les noms de fonctions prédéfinies**

**Remarque :** Vous avez peut-être observé que nous avons choisi de ne pas appeler notre tuple simplement tuple. C’est une bonne pratique en général d’éviter les noms de fonctions prédéfinies par Python.

Ces variables en effet sont des variables “comme les autres”. Imaginez qu’on ait en fait deux tuples à construire comme ci-dessus, voici ce qu’on obtiendrait si on n’avait pas pris cette précaution :

```
In [14]: liste = range(10)
         # ATTENTION : ceci redéfinit le symbole tuple
         tuple = tuple(liste)
         tuple
```

```
Out[14]: (0, 1, 2, 3, 4, 5, 6, 7, 8, 9)
```

```
In [15]: # si bien que maintenant on ne peut plus faire ceci
         # car à ce point, tuple ne désigne plus le type tuple
         # mais l'objet qu'on vient de créer
         autre_liste = range(100)
         autre_tuple = tuple(autre_liste)
```

```
-----
TypeError                                     Traceback (most recent call last)
```

```
<ipython-input-15-15ef2d1e6804> in <module>()
      3 # mais l'objet qu'on vient de créer
      4 autre_liste = range(100)
----> 5 autre_tuple = tuple(autre_liste)
```

```
TypeError: 'tuple' object is not callable
```

Il y a une erreur parce que nous avons remplacé (ligne 2) la valeur de la variable `tuple`, qui au départ référençait le **type** tuple (ou si on préfère la fonction de conversion), par un **objet** tuple. Ainsi en ligne 5, lorsqu’on appelle à nouveau `tuple`, on essaie d’exécuter un objet qui n’est pas ‘appelable’ (*not callable* en anglais).

D’un autre côté, l’erreur est relativement facile à trouver dans ce cas. En cherchant toutes les occurrences de `tuple` dans notre propre code on voit assez vite le problème. De plus, je vous rappelle que votre éditeur de texte **doit** faire de la coloration syntaxique, et que toutes les fonctions built-in (dont `tuple` et `list` font partie) sont colorées spécifiquement (par exemple, en violet sous IDLE). En pratique, avec un bon éditeur de texte et un peu d’expérience, cette erreur est très rare.

## 3.6 Sequence unpacking

### 3.6.1 Complément - niveau basique

**Remarque préliminaire :** nous avons vainement cherché une traduction raisonnable pour ce trait du langage, connue en anglais sous le nom de *sequence unpacking* ou encore parfois *tuple unpacking*, aussi pour éviter de créer de la confusion nous avons finalement décidé de conserver le terme anglais à l'identique.

#### Déjà rencontré

L'affectation dans Python peut concerner plusieurs variables à la fois. En fait nous en avons déjà vu un exemple en Semaine 1, avec la fonction `fibonacci` dans laquelle il y avait ce fragment :

```
for i in range(2, n + 1):
    f2, f1 = f1, f1 + f2
```

Nous allons dans ce complément décortiquer les mécanismes derrière cette phrase qui a probablement excité votre curiosité. :)

#### Un exemple simple

Commençons par un exemple simple à base de tuple. Imaginons que l'on dispose d'un tuple couple dont on sait qu'il a deux éléments :

```
In [1]: couple = (100, 'spam')
```

On souhaite à présent extraire les deux valeurs, et les affecter à deux variables distinctes. Une solution naïve consiste bien sûr à faire simplement :

```
In [2]: gauche = couple[0]
        droite = couple[1]
        print('gauche', gauche, 'droite', droite)
```

```
gauche 100 droite spam
```

Cela fonctionne naturellement très bien, mais n'est pas très pythonique - comme on dit ;) Vous devez toujours garder en tête qu'il est rare en Python de manipuler des indices. Dès que vous voyez des indices dans votre code, vous devez vous demander si votre code est pythonique.

On préférera la formulation équivalente suivante :

```
In [3]: (gauche, droite) = couple
        print('gauche', gauche, 'droite', droite)
```

```
gauche 100 droite spam
```

La logique ici consiste à dire : affecter les deux variables de sorte que le tuple (gauche, droite) soit égal à couple. On voit ici la supériorité de cette notion d'unpacking sur la manipulation d'indices : vous avez maintenant des variables qui expriment la nature de l'objet manipulé, votre code devient expressif, c'est-à-dire auto-documenté.

Remarquons que les parenthèses ici sont optionnelles - comme lorsque l'on construit un tuple - et on peut tout aussi bien écrire, et c'est le cas d'usage le plus fréquent d'omission des parenthèses pour le tuple :

```
In [4]: gauche, droite = couple
        print('gauche', gauche, 'droite', droite)
```

```
gauche 100 droite spam
```

### Autres types

Cette technique fonctionne aussi bien avec d'autres types. Par exemple, on peut utiliser :

- une syntaxe de liste à gauche du = ;
- une liste comme expression à droite du =.

```
In [5]: # comme ceci
        liste = [1, 2, 3]
        [gauche, milieu, droit] = liste
        print('gauche', gauche, 'milieu', milieu, 'droit', droit)
```

```
gauche 1 milieu 2 droit 3
```

Et on n'est même pas obligés d'avoir le même type à gauche et à droite du signe =, comme ici :

```
In [6]: # membre droit: une liste
        liste = [1, 2, 3]
        # membre gauche : un tuple
        gauche, milieu, droit = liste
        print('gauche', gauche, 'milieu', milieu, 'droit', droit)
```

```
gauche 1 milieu 2 droit 3
```

En réalité, les seules contraintes fixées par Python sont que :

- le terme à droite du signe = soit un *itérable* (tuple, liste, string, etc.);
- le terme à gauche soit écrit comme un tuple ou une liste - notons tout de même que l'utilisation d'une liste à gauche est rare et peu pythonique;
- les deux termes aient la même longueur - en tout cas avec les concepts que l'on a vus jusqu'ici, mais voir aussi plus bas l'utilisation de \*arg avec le *extended unpacking*.

La plupart du temps le terme de gauche est écrit comme un tuple. C'est pour cette raison que les deux termes *tuple unpacking* et *sequence unpacking* sont en vigueur.

### La façon *pythonique* d'échanger deux variables

Une caractéristique intéressante de l'affectation par *sequence unpacking* est qu'elle est sûre ; on n'a pas à se préoccuper d'un éventuel ordre d'évaluation, les valeurs à **droite** de l'affectation sont **toutes** évaluées en premier, et ainsi on peut par exemple échanger deux variables comme ceci :

```
In [7]: a = 1
        b = 2
        a, b = b, a
        print('a', a, 'b', b)

a 2 b 1
```

### Extended unpacking

Le *extended unpacking* a été introduit en Python 3 ; commençons par en voir un exemple :

```
In [8]: reference = [1, 2, 3, 4, 5]
        a, *b, c = reference
        print(f"a={a} b={b} c={c}")

a=1 b=[2, 3, 4] c=5
```

Comme vous le voyez, le mécanisme ici est une extension de *sequence unpacking* ; Python vous autorise à mentionner **une seule fois**, parmi les variables qui apparaissent à gauche de l'affectation, une variable **précédée de \***, ici *\*b*.

Cette variable est interprétée comme une **liste de longueur quelconque** des éléments de *reference*. On aurait donc aussi bien pu écrire :

```
In [9]: reference = range(20)
        a, *b, c = reference
        print(f"a={a} b={b} c={c}")

a=0 b=[1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18] c=19
```

Ce trait peut s'avérer pratique, lorsque par exemple on s'intéresse seulement aux premiers éléments d'une structure :

```
In [10]: # si on sait que data contient prenom, nom, et un nombre inconnu d'autres informations
data = [ 'Jean', 'Dupont', '061234567', '12', 'rue du chemin vert', '57000', 'METZ',
        # on peut utiliser la variable _ qui véhicule l'idée que l'on ne s'y intéresse pas
        prenom, nom, *_ = data
        print(f"prenom={prenom} nom={nom}")

prenom=Jean nom=Dupont
```

### 3.6.2 Complément - niveau intermédiaire

On a vu les principaux cas d'utilisation de la *sequence unpacking*, voyons à présent quelques subtilités.



### Plusieurs occurrences d'une même variable

On peut utiliser **plusieurs fois** la même variable dans la partie gauche de l'affectation :

```
In [11]: # ceci en toute rigueur est légal
        # mais en pratique on évite de le faire
        entree = [1, 2, 3]
        a, a, a = entree
        print(f"a = {a}")

a = 3
```

**Attention** toutefois, comme on le voit ici, Python **n'impose pas** que les différentes occurrences de `a` correspondent à **des valeurs identiques** (en langage savant, on dirait que cela ne permet pas de faire de l'unification). De manière beaucoup plus pragmatique, l'interpréteur se contente de faire comme s'il faisait l'affectation plusieurs fois de gauche à droite, c'est-à-dire comme s'il faisait :

```
In [ ]: a = 1; a = 2; a = 3
```

Cette technique n'est utilisée en pratique que pour les parties de la structure dont on n'a que faire dans le contexte. Dans ces cas-là, il arrive qu'on utilise le nom de variable `_`, dont on rappelle qu'il est légal, ou tout autre nom comme `ignored` pour manifester le fait que cette partie de la structure ne sera pas utilisée, par exemple :

```
In [12]: entree = [1, 2, 3]

        _, milieu, _ = entree
        print('milieu', milieu)

        ignored, ignored, right = entree
        print('right', right)

milieu 2
right 3
```

### En profondeur

Le *sequence unpacking* ne se limite pas au premier niveau dans les structures, on peut extraire des données plus profondément imbriquées dans la structure de départ; par exemple avec en entrée la liste :

```
In [13]: structure = ['abc', [(1, 2), ([3], 4)], 5]
```

Si on souhaite extraire la valeur qui se trouve à l'emplacement du 3, on peut écrire :

```
In [14]: (a, (b, ((trois,), c)), d) = structure
        print('trois', trois)

trois 3
```

Ou encore, sans doute un peu plus lisible :

```
In [15]: (a, (b, ([trois], c)), d) = structure
         print('trois', trois)

trois 3
```

Naturellement on aurait aussi bien pu écrire ici quelque chose comme :

```
In [16]: trois = structure[1][1][0][0]
         print('trois', trois)

trois 3
```

Affaire de goût évidemment. Mais n’oublions pas une des phrases du Zen de Python *Flat is better than nested*, ce qui veut dire que ce n’est pas parce que vous pouvez faire des structures imbriquées complexes que vous devez le faire. Bien souvent, cela rend la lecture et la maintenance du code complexe, j’espère que l’exemple précédent vous en a convaincu.

### Extended unpacking et profondeur

On peut naturellement ajouter de l’*extended unpacking* à n’importe quel étage d’un *unpacking* imbriqué :

```
In [17]: # un exemple très alambiqué avec plusieurs variables *extended
         tree = [1, 2, [(3, 33, 'three', 'thirty-three')], ([4, 44, ('forty', 'forty-four')],
         *_), ((_, *x3, _),), (*_, x4) = tree
         print(f"x3={x3}, x4={x4}")

x3=[33, 'three'], x4=('forty', 'forty-four')
```

Dans ce cas, la limitation d’avoir une seule variable de la forme *\*extended* s’applique toujours, naturellement, mais à chaque niveau dans l’imbrication, comme on le voit sur cet exemple.

### 3.6.3 Pour en savoir plus

— [Le PEP \(en anglais\) qui introduit le \*extended unpacking\*.](#)

## 3.7 Plusieurs variables dans une boucle for

### 3.7.1 Complément - niveau basique

Nous avons vu précédemment (séquence ‘Les tuples’, complément ‘Sequence unpacking’) la possibilité d’affecter plusieurs variables à partir d’un seul objet, comme ceci :

```
In [1]: item = (1, 2)
        a, b = item
        print(f"a={a} b={b}")
```

```
a=1 b=2
```

D’une façon analogue, il est possible de faire une boucle for qui itère sur **une seule** liste mais qui *agit* sur **plusieurs variables**, comme ceci :

```
In [2]: entrees = [(1, 2), (3, 4), (5, 6)]
        for a, b in entrees:
            print(f"a={a} b={b}")
```

```
a=1 b=2
```

```
a=3 b=4
```

```
a=5 b=6
```

À chaque itération, on trouve dans *entree* un tuple (d’abord (1, 2), puis à l’itération suivante (3, 4), etc.); à ce stade les variables *a* et *b* vont être affectées à, respectivement, le premier et le deuxième élément du tuple, exactement comme dans le *sequence unpacking*. Cette mécanique est massivement utilisée en Python.

### 3.7.2 Complément - niveau intermédiaire

#### La fonction zip

Voici un exemple très simple qui utilise la technique que l’on vient de voir.

Imaginons qu’on dispose de deux listes de longueurs égales, dont on sait que les entrées correspondent une à une, comme par exemple :

```
In [3]: villes = ["Paris", "Nice", "Lyon"]
        populations = [2*10**6, 4*10**5, 10**6]
```

Afin d’écrire facilement un code qui “associe” les deux listes entre elles, Python fournit une fonction *built-in* baptisée *zip*; voyons ce qu’elle peut nous apporter sur cet exemple :

```
In [4]: list(zip(villes, populations))
```

```
Out[4]: [('Paris', 2000000), ('Nice', 400000), ('Lyon', 1000000)]
```

On le voit, on obtient en retour une liste composée de tuples. On peut à présent écrire une boucle for comme ceci :

```
In [5]: for ville, population in zip(villes, populations):
        print(population, "habitants à", ville)
```

```
2000000 habitants à Paris
400000 habitants à Nice
1000000 habitants à Lyon
```

Qui est, nous semble-t-il, beaucoup plus lisible que ce que l'on serait amené à écrire avec des langages plus traditionnels.

Tout ceci se généralise naturellement à plus de deux variables :

```
In [6]: for i, j, k in zip(range(3), range(100, 103), range(200, 203)):
        print(f"i={i} j={j} k={k}")
```

```
i=0 j=100 k=200
i=1 j=101 k=201
i=2 j=102 k=202
```

**Remarque** : lorsqu'on passe à zip des listes de tailles différentes, le résultat est tronqué, c'est l'entrée de **plus petite taille** qui détermine la fin du parcours.

```
In [7]: # on n'itère que deux fois
        # car le premier argument de zip est de taille 2
        for units, tens in zip([1, 2], [10, 20, 30, 40]):
            print(units, tens)
```

```
1 10
2 20
```

### La fonction enumerate

Une autre fonction très utile permet d'itérer sur une liste avec l'indice dans la liste, il s'agit de `enumerate` :

```
In [8]: for i, ville in enumerate(villes):
        print(i, ville)
```

```
0 Paris
1 Nice
2 Lyon
```

Cette forme est **plus simple** et **plus lisible** que les formes suivantes qui sont équivalentes, mais qui ne sont pas pythoniques :

```
In [9]: for i in range(len(villes)):
        print(i, villes[i])
```

```
0 Paris
1 Nice
2 Lyon
```

```
In [10]: for i, ville in zip(range(len(villes)), villes):  
         print(i, ville)
```

```
0 Paris  
1 Nice  
2 Lyon
```

## 3.8 Fichiers

### 3.8.1 Exercice - niveau basique

#### Calcul du nombre de lignes, de mots et de caractères

```
In [ ]: # chargement de l'exercice
        from corrections.exo_comptage import exo_comptage
```

On se propose d'écrire une *moulinette* qui annote un fichier avec des nombres de lignes, de mots et de caractères.

Le but de l'exercice est d'écrire une fonction `comptage` :

- qui prenne en argument un nom de fichier d'entrée (on suppose qu'il existe) et un nom de fichier de sortie (on suppose qu'on a le droit de l'écrire) ;
- le fichier d'entrée est supposé encodé en UTF-8 ;
- le fichier d'entrée est laissé intact ;
- pour chaque ligne en entrée, le fichier de sortie comporte une ligne qui donne le numéro de ligne, le nombre de mots (**séparés par des espaces**), le nombre de caractères (y compris la fin de ligne), et la ligne d'origine.

```
In [ ]: # un exemple de ce qui est attendu
        exo_comptage.example()
```

```
In [ ]: # votre code
        def comptage(in_filename, out_filename):
            "votre code"
```

**N'oubliez pas de vérifier** que vous ajoutez bien les **fins de ligne**, car la vérification automatique est pointilleuse (elle utilise l'opérateur `==`), et rejettera votre code si vous ne produisez pas une sortie rigoureusement similaire à ce qui est attendu.

```
In [ ]: # pour vérifier votre code
        # voyez aussi un peu plus bas, une cellule d'aide au debugging

        exo_comptage.correction(comptage)
```

La méthode `debug` applique votre fonction au premier fichier d'entrée, et affiche le résultat comme dans l'exemple ci-dessus :

```
In [ ]: # debugging
        exo_comptage.debug(comptage)
```

#### Accès aux fichiers d'exemples

Vous pouvez télécharger les fichiers d'exemples :

- [Romeo and Juliet](#)
- [Lorem Ipsum](#)
- ["Une charogne" en UTF-8](#)

Pour les courageux, je vous donne également [“Une charogne” en ISO-8859-15](#), qui contient le même texte que “Une charogne”, mais encodé en Latin-9, connu aussi sous le nom ISO-8859-15.

Ce dernier fichier n’est pas à prendre en compte dans la version basique de l’exercice, mais vous pourrez vous rendre compte par vous-mêmes, au cas où cela ne serait pas clair encore pour vous, qu’il n’est pas facile d’écrire une fonction `comptage` qui devine l’encodage, c’est-à-dire qui fonctionne correctement avec des entrées indifféremment en Unicode ou Latin, sans que cet encodage soit passé en paramètre à `comptage`.

C’est d’ailleurs le propos de [la bibliothèque `chardet`](#) qui s’efforce de déterminer l’encodage de fichiers d’entrée, sur la base de modèles statistiques.

## 3.9 Sequence unpacking

### 3.9.1 Exercice - niveau basique

```
In [ ]: # chargeons l'exercice
        from corrections.exo_surgery import exo_surgery
```

Cet exercice consiste à écrire une fonction `surgery`, qui prend en argument une liste, et qui retourne la **même** liste **modifiée** comme suit :

- si la liste est de taille 0 ou 1, elle n'est pas modifiée;
- si la liste est de taille paire, on intervertit les deux premiers éléments de la liste;
- si elle est de taille impaire, on intervertit les deux derniers éléments.

```
In [ ]: # voici quelques exemples de ce qui est attendu
        exo_surgery.example()
```

```
In [ ]: # écrivez votre code
        def surgery(liste):
            "<votre_code>"
```

```
In [ ]: # pour le vérifier, évaluez cette cellule
        exo_surgery.correction(surgery)
```



## 3.10 Dictionnaires

### 3.10.1 Complément - niveau basique

Ce document résume les opérations courantes disponibles sur le type `dict`. On rappelle que le type `dict` est un type **mutable**.

#### Création en extension

On l'a vu, la méthode la plus directe pour créer un dictionnaire est en extension comme ceci :

```
In [1]: annuaire = {'marc': 35, 'alice': 30, 'eric': 38}
        print(annuaire)

{'marc': 35, 'alice': 30, 'eric': 38}
```

#### Création - la fonction `dict`

Comme pour les fonctions `int` ou `list`, la fonction `dict` est une fonction de construction de dictionnaire - on dit un constructeur. On a vu aussi dans la vidéo qu'on peut utiliser ce constructeur à base d'une liste de tuples (clé, valeur)

```
In [2]: # le paramètre de la fonction dict est
        # une liste de couples (clé, valeur)
        annuaire = dict([('marc', 35), ('alice', 30), ('eric', 38)])
        print(annuaire)

{'marc': 35, 'alice': 30, 'eric': 38}
```

Remarquons qu'on peut aussi utiliser cette autre forme d'appel à `dict` pour un résultat équivalent :

```
In [3]: annuaire = dict(marc=35, alice=30, eric=38)
        print(annuaire)

{'marc': 35, 'alice': 30, 'eric': 38}
```

Remarquez ci-dessus l'absence de quotes autour des clés comme `marc`. Il s'agit d'un cas particulier de passage d'arguments que nous expliciterons plus longuement en fin de semaine 4.

#### Accès atomique

Pour accéder à la valeur associée à une clé, on utilise la notation à base de crochets `[]` :

```
In [4]: print('la valeur pour marc est', annuaire['marc'])

la valeur pour marc est 35
```

Cette forme d'accès ne fonctionne que si la clé est effectivement présente dans le dictionnaire. Dans le cas contraire, une exception `KeyError` est levée. Aussi si vous n'êtes pas sûr que la clé soit présente, vous pouvez utiliser la méthode `get` qui accepte une valeur par défaut :

```
In [5]: print('valeur pour marc', annuaire.get('marc', 0))
        print('valeur pour inconnu', annuaire.get('inconnu', 0))
```

```
valeur pour marc 35
valeur pour inconnu 0
```

Le dictionnaire est un type **mutable**, et donc on peut **modifier la valeur** associée à une clé :

```
In [6]: annuaire['eric'] = 39
        print(annuaire)

{'marc': 35, 'alice': 30, 'eric': 39}
```

Ou encore, exactement de la même façon, **ajouter une entrée** :

```
In [7]: annuaire['bob'] = 42
        print(annuaire)

{'marc': 35, 'alice': 30, 'eric': 39, 'bob': 42}
```

Enfin pour **détruire une entrée**, on peut utiliser l'instruction `del` comme ceci :

```
In [8]: # pour supprimer la clé 'marc' et donc sa valeur aussi
        del annuaire['marc']
        print(annuaire)

{'alice': 30, 'eric': 39, 'bob': 42}
```

Pour savoir si une clé est présente ou non, il est conseillé d'utiliser l'opérateur d'appartenance `in` comme ceci :

```
In [9]: # forme recommandée
        print('john' in annuaire)

False
```

### Parcourir toutes les entrées

La méthode la plus fréquente pour parcourir tout un dictionnaire est à base de la méthode `items`; voici par exemple comment on pourrait afficher le contenu :

```
In [10]: for nom, age in annuaire.items():
         print(f"{nom}, age {age}")
```

```
alice, age 30  
eric, age 39  
bob, age 42
```

On remarque d'abord que les entrées sont listées dans le désordre, plus précisément, il n'y a pas de notion d'ordre dans un dictionnaire; ceci est dû à l'action de la fonction de hachage, que nous avons vue dans la vidéo précédente.

On peut obtenir séparément la liste des clés et des valeurs avec :

```
In [11]: for clé in annuaire.keys():  
         print(clé)
```

```
alice  
eric  
bob
```

```
In [12]: for valeur in annuaire.values():  
         print(valeur)
```

```
30  
39  
42
```

### La fonction len

On peut comme d'habitude obtenir la taille d'un dictionnaire avec la fonction len :

```
In [13]: print(f"{len(annuaire)} entrées dans annuaire")
```

```
3 entrées dans annuaire
```

### Pour en savoir plus sur le type dict

Pour une liste exhaustive reportez-vous à la page de la documentation Python ici :

<https://docs.python.org/3/library/stdtypes.html#mapping-types-dict>

### 3.10.2 Complément - niveau intermédiaire

#### La méthode `update`

On peut également modifier un dictionnaire avec le contenu d'un autre dictionnaire avec la méthode `update` :

```
In [14]: print(f"avant: {list(annuaire.items())}")
```

```
avant: [('alice', 30), ('eric', 39), ('bob', 42)]
```

```
In [15]: annuaire.update({'jean':25, 'eric':70})
         list(annuaire.items())
```

```
Out[15]: [('alice', 30), ('eric', 70), ('bob', 42), ('jean', 25)]
```

#### `collections.OrderedDict` : dictionnaire et ordre d'insertion

**Attention** : un dictionnaire est **non ordonné** ! Il ne se souvient pas de l'ordre dans lequel les éléments ont été insérés. C'était particulièrement visible dans les versions de Python jusque 3.5 :

```
In [ ]: %%python2
```

```
# cette cellule utilise python-2.7 pour illustrer le fait
# que les dictionnaires ne sont pas ordonnés

d = {'c' : 3, 'b' : 1, 'a' : 2}
for k, v in d.items():
    print k, v
```

En réalité, et depuis la version 3.6 de Python, il se trouve qu'**incidemment** l'implémentation CPython (la plus répandue donc) a été modifiée, et maintenant on peut avoir l'**impression** que les dictionnaires sont ordonnés :

```
In [17]: d = {'c' : 3, 'b' : 1, 'a' : 2}
         for k, v in d.items():
             print(k, v)
```

```
c 3
b 1
a 2
```

Il faut insister sur le fait qu'il s'agit d'un **détail d'implémentation**, et que vous ne devez pas écrire du code qui suppose que les dictionnaires sont ordonnés.

Si vous avez besoin de dictionnaires qui sont **garantis** ordonnés, voyez dans le [module `collections`](#) la classe `OrderedDict`, qui est une personnalisation (une sous-classe) du type `dict`, qui cette fois possède cette bonne propriété :

```
In [18]: from collections import OrderedDict
         d = OrderedDict()
         for i in ['a', 7, 3, 'x']:
             d[i] = i
         for k, v in d.items():
             print('OrderedDict', k, v)

OrderedDict a a
OrderedDict 7 7
OrderedDict 3 3
OrderedDict x x
```

#### `collections.defaultdict` : initialisation automatique

Imaginons que vous devez gérer un dictionnaire dont les valeurs sont des listes, et que votre programme ajoute des valeurs au fur et à mesure dans ces listes.

Avec un dictionnaire de base, cela peut vous amener à écrire un code qui ressemble à ceci :

```
In [19]: # imaginons qu'on a lu dans un fichier des couples (x, y)
         tuples = [
             (1, 2),
             (2, 1),
             (1, 3),
             (2, 4),
         ]

In [20]: # et on veut construire un dictionnaire
         # x -> [liste de tous les y connectés à x]
         resultat = {}

         for x, y in tuples:
             if x not in resultat:
                 resultat[x] = []
             resultat[x].append(y)

         for key, value in resultat.items():
             print(key, value)

1 [2, 3]
2 [1, 4]
```

Cela fonctionne, mais n'est pas très élégant. Pour simplifier ce type de traitement, vous pouvez utiliser `defaultdict`, une sous-classe de `dict` dans le module `collections` :

```
In [21]: from collections import defaultdict

         # on indique que les valeurs doivent être créées à la volée
         # en utilisant la fonction list
         resultat = defaultdict(list)
```

```

# du coup plus besoin de vérifier la présence de la clé
for x, y in tuples:
    resultat[x].append(y)

for key, value in resultat.items():
    print(key, value)

```

1 [2, 3]  
2 [1, 4]

Cela fonctionne aussi avec le type `int`, lorsque vous voulez par exemple compter des occurrences :

```

In [22]: compteurs = defaultdict(int)

phrase = "une phrase dans laquelle on veut compter les caractères"

for c in phrase:
    compteurs[c] += 1

sorted(compteurs.items())

```

Out [22]: [(' ', 8),  
('a', 5),  
('c', 3),  
('d', 1),  
('e', 8),  
('h', 1),  
('l', 4),  
('m', 1),  
('n', 3),  
('o', 2),  
('p', 2),  
('q', 1),  
('r', 4),  
('s', 4),  
('t', 3),  
('u', 3),  
('v', 1),  
('è', 1)]

Signalons enfin une fonctionnalité un peu analogue, quoiqu'un peut moins élégante à mon humble avis, mais qui est présente avec les dictionnaires `dict` standard. Il s'agit de [la méthode `setdefault`](#) qui permet, en un seul appel, de retourner la valeur associée à une clé et de créer cette clé au besoin, c'est-à-dire si elle n'est pas encore présente :

```

In [23]: # avant
annuaire

Out [23]: {'alice': 30, 'bob': 42, 'eric': 70, 'jean': 25}

In [24]: # ceci sera sans effet car eric est déjà présent
annuaire.setdefault('eric', 50)

```

```
Out[24]: 70
```

```
In [25]: # par contre ceci va insérer une entrée dans le dictionnaire
annuaire.setdefault('inconnu', 50)
```

```
Out[25]: 50
```

```
In [26]: # comme on le voit
annuaire
```

```
Out[26]: {'alice': 30, 'bob': 42, 'eric': 70, 'inconnu': 50, 'jean': 25}
```

Notez bien que `setdefault` peut éventuellement créer une entrée mais ne **modifie jamais** la valeur associée à une clé déjà présente dans le dictionnaire, comme le nom le suggère d'ailleurs.

### 3.10.3 Complément - niveau avancé

Pour bien appréhender les dictionnaires, il nous faut souligner certaines particularités, à propos de la valeur de retour des méthodes comme `items()`, `keys()` et `values()`.

#### Ce sont des objets itérables

Les méthodes `items()`, `keys()` et `values()` ne retournent pas des listes (comme c'était le cas en Python 2), mais des **objets itérables** :

```
In [27]: d = {'a' : 1, 'b' : 2}
keys = d.keys()
keys
```

```
Out[27]: dict_keys(['a', 'b'])
```

Comme ce sont des itérables, on peut naturellement faire un `for` avec, on l'a vu :

```
In [28]: for key in keys:
print(key)
```

```
a
b
```

Et un test d'appartenance avec `in` :

```
In [29]: print('a' in keys)
```

```
True
```

```
In [30]: print('x' in keys)
```

```
False
```

**Mais ce ne sont pas des listes**

```
In [31]: isinstance(keys, list)
```

```
Out[31]: False
```

Ce qui signifie qu'on n'a **pas alloué de mémoire** pour stocker toutes les clés, mais seulement un objet qui ne prend pas de place, ni de temps à construire :

```
In [32]: # construisons un dictionnaire
         # pour anticiper un peu sur la compréhension de dictionnaire
```

```
big_dict = {k : k**2 for k in range(1_000_000)}
```

```
In [33]: %%timeit -n 10000
         # créer un objet vue est très rapide
big_keys = big_dict.keys()
```

193 ns  $\pm$  51 ns per loop (mean  $\pm$  std. dev. of 7 runs, 10000 loops each)

```
In [34]: # on répète ici car timeit travaille dans un espace qui lui est propre
         # et donc on n'a pas défini big_keys pour notre interpréteur
big_keys = big_dict.keys()
```

```
In [35]: %%timeit -n 20
         # si on devait vraiment construire la liste ce serait beaucoup plus long
big_lkeys = list(big_keys)
```

20.1 ms  $\pm$  391  $\mu$ s per loop (mean  $\pm$  std. dev. of 7 runs, 20 loops each)

**En fait ce sont des vues** Une autre propriété un peu inattendue de ces objets, c'est que **ce sont des vues**; ce qu'on veut dire par là (pour ceux qui connaissent, cela fait référence à la notion de vue dans les bases de données) c'est que la vue *voit* les changements fait sur l'objet dictionnaire *même après sa création* :

```
In [36]: d = {'a' : 1, 'b' : 2}
         keys = d.keys()
```

```
In [37]: # sans surprise, il y a deux clés dans keys
         for k in keys:
             print(k)
```

```
a
b
```

```
In [38]: # mais si maintenant j'ajoute un objet au dictionnaire
         d['c'] = 3
         # alors on va 'voir' cette nouvelle clé à partir
         # de l'objet keys qui pourtant est inchangé
         for k in keys:
             print(k)
```



a  
b  
c

Reportez vous à [la section sur les vues de dictionnaires](#) pour plus de détails.

**Python 2** Ceci est naturellement en fort contraste avec tout ce qui se passait en Python 2, où l'on avait des méthodes distinctes, par exemple `keys()`, `iterkeys()` et `viewkeys()`, selon le type d'objets que l'on souhaitait construire.

## 3.11 Clés immuables

### 3.11.1 Complément - niveau intermédiaire

Nous avons vu comment manipuler un dictionnaire, il nous reste à voir un peu plus en détail les contraintes qui sont mises par le langage sur ce qui peut servir de clé dans un dictionnaire. On parle dans ce complément spécifiquement des clefs construites à partir des types `built-in`. Le cas de vos propres classes utilisées comme clefs de dictionnaires n'est pas abordé dans ce complément.

#### Une clé doit être immuable

Si vous vous souvenez de la vidéo sur les tables de hash, la mécanique interne du dictionnaire repose sur le calcul, à partir de chaque clé, d'une fonction de hachage.

C'est-à-dire que, pour simplifier, on localise la présence d'une clé en calculant d'abord  $f(cl) = hash$

puis on poursuit la recherche en utilisant `hash` comme indice dans le tableau contenant les couples (clé, valeur).

On le rappelle, c'est cette astuce qui permet de réaliser les opérations sur les dictionnaires en temps constant - c'est-à-dire indépendamment du nombre d'éléments.

Cependant, pour que ce mécanisme fonctionne, il est indispensable que **la valeur de la clé reste inchangée** pendant la durée de vie du dictionnaire. Sinon, bien entendu, on pourrait avoir le scénario suivant :

- on range un tuple (`clef`, `valeur`) à un premier indice  $f(clef) = hash_1$ ;
- on modifie la valeur de `clef` qui devient `clef'`;
- on recherche notre valeur à l'indice  $f(clef') = hash_2 \neq hash_1$ .

et donc, avec ces hypothèses, on n'a plus la garantie de bon fonctionnement de la logique.

#### Une clé doit être globalement immuable

Nous avons depuis le début du cours longuement insisté sur le caractère mutable ou immuable des différents types prédéfinis de Python. Vous devez donc à présent avoir au moins en partie ce tableau en tête :

Type	Mutable?
<code>int, float</code>	immuable
<code>complex, bool</code>	immuable
<code>str</code>	immuable
<code>list</code>	mutable
<code>dict</code>	mutable
<code>set</code>	mutable
<code>frozenset</code>	immuable

Le point important ici, est qu'il **ne suffit pas**, pour une clé, d'être **de type immuable**.

On peut le voir sur un exemple très simple ; donnons-nous donc un dictionnaire :

```
In [1]: d = {}
```

Et commençons avec un objet de type immuable, un tuple d'entiers :

```
In [2]: bonne_cle = (1, 2)
```

Cet objet est non seulement **de type immuable**, mais tous ses composants et sous-composants sont **immuables**, on peut donc l'utiliser comme clé dans le dictionnaire :

```
In [3]: d[bonne_cle] = "pas de probleme ici"
        print(d)
{(1, 2): 'pas de probleme ici'}
```

Si à présent on essaie d'utiliser comme clé un tuple qui contient une liste :

```
In [4]: mauvaise_cle = (1, [1, 2])
```

Il se trouve que cette clé, **bien que de type immuable**, peut être **indirectement modifiée** puisque :

```
In [5]: mauvaise_cle[1].append(3)
        print(mauvaise_cle)
(1, [1, 2, 3])
```

Et c'est pourquoi on ne peut pas utiliser cet objet comme clé dans le dictionnaire :

```
In [6]: # provoque une exception
        d[mauvaise_cle] = 'on ne peut pas faire ceci'
```

```
-----
TypeError                                Traceback (most recent call last)

<ipython-input-6-08a3625a4eed> in <module>()
    1 # provoque une exception
----> 2 d[mauvaise_cle] = 'on ne peut pas faire ceci'

TypeError: unhashable type: 'list'
```

Pour conclure, il faut retenir qu'un objet n'est éligible pour être utilisé comme clé que s'il est **composé de types immuables de haut en bas** de la structure de données.

La raison d'être principale du type `tuple`, que nous avons vu la semaine passée, et du type `frozenset`, que nous verrons très prochainement, est précisément de construire de tels objets globalement immuables.

## Épilogue

Tout ceci est valable pour les types *built-in*. Nous verrons que pour les types définis par l'utilisateur - les classes donc - que nous effleurons à la fin de cette semaine et que nous étudions plus en profondeur en semaine 6, c'est un autre mécanisme qui est utilisé pour calculer la clé de hachage d'une instance de classe.

## 3.12 Gérer des enregistrements

### 3.12.1 Complément - niveau intermédiaire

#### Implémenter un enregistrement comme un dictionnaire

Il nous faut faire le lien entre dictionnaire Python et la notion d'enregistrement, c'est-à-dire une donnée composite qui contient plusieurs champs. (À cette notion correspond, selon les langages, ce qu'on appelle un `struct` ou un `record`.)

Imaginons qu'on veuille manipuler un ensemble de données concernant des personnes ; chaque personne est supposée avoir un nom, un âge et une adresse mail.

Il est possible, et assez fréquent, d'utiliser le dictionnaire comme support pour modéliser ces données comme ceci :

```
In [1]: personnes = [
        {'nom': 'Pierre', 'age': 25, 'email': 'pierre@example.com'},
        {'nom': 'Paul', 'age': 18, 'email': 'paul@example.com'},
        {'nom': 'Jacques', 'age': 52, 'email': 'jacques@example.com'},
    ]
```

Bon, très bien, nous avons nos données, il est facile de les utiliser.

Par exemple, pour l'anniversaire de Pierre on fera :

```
In [2]: personnes[0]['age'] += 1
```

Ce qui nous donne :

```
In [3]: for personne in personnes:
        print(10*"=")
        for info, valeur in personne.items():
            print(f"{info} -> {valeur}")
```

```
=====
nom -> Pierre
age -> 26
email -> pierre@example.com
=====
nom -> Paul
age -> 18
email -> paul@example.com
=====
nom -> Jacques
age -> 52
email -> jacques@example.com
```

#### Un dictionnaire pour indexer les enregistrements

Cela dit, il est bien clair que cette façon de faire n'est pas très pratique ; pour marquer l'anniversaire de Pierre on ne sait bien entendu pas que son enregistrement est le premier dans la liste. C'est pourquoi il est plus adapté, pour modéliser ces informations, d'utiliser non

pas une liste, mais à nouveau... un dictionnaire.

Si on imagine qu'on a commencé par lire ces données séquentiellement dans un fichier, et qu'on a calculé l'objet personnes comme la liste qu'on a vue ci-dessus, alors il est possible de construire un index de ces dictionnaires, (un dictionnaire de dictionnaires, donc).

C'est-à-dire, en anticipant un peu sur la construction de dictionnaires par compréhension :

```
In [4]: # on crée un index permettant de retrouver rapidement
        # une personne dans la liste
        index_par_nom = {personne['nom']: personne for personne in personnes}
        index_par_nom

Out[4]: {'Jacques': {'age': 52, 'email': 'jacques@example.com', 'nom': 'Jacques'},
        'Paul': {'age': 18, 'email': 'paul@example.com', 'nom': 'Paul'},
        'Pierre': {'age': 26, 'email': 'pierre@example.com', 'nom': 'Pierre'}}
```

```
In [5]: # du coup pour accéder à l'enregistrement pour Pierre
        index_par_nom['Pierre']

Out[5]: {'age': 26, 'email': 'pierre@example.com', 'nom': 'Pierre'}
```

Attardons-nous un tout petit peu; nous avons construit un dictionnaire par compréhension, en créant autant d'entrées que de personnes. Nous aborderons en détail la notion de compréhension de sets et de dictionnaires en semaine 5, donc si cette notation vous paraît étrange pour le moment, pas d'inquiétude.

Le résultat est donc un dictionnaire qu'on peut afficher comme ceci :

```
In [6]: for nom, record in index_par_nom.items():
        print(f"Nom : {nom} -> enregistrement : {record}")

Nom : Pierre -> enregistrement : {'nom': 'Pierre', 'age': 26, 'email': 'pierre@example.com'}
Nom : Paul -> enregistrement : {'nom': 'Paul', 'age': 18, 'email': 'paul@example.com'}
Nom : Jacques -> enregistrement : {'nom': 'Jacques', 'age': 52, 'email': 'jacques@example.com'}
```

Dans cet exemple, le premier niveau de dictionnaire permet de trouver rapidement un objet à partir d'un nom; dans le second niveau au contraire on utilise le dictionnaire pour implémenter un enregistrement, à la façon d'un struct en C.

### Techniques similaires

Notons enfin qu'il existe aussi, en Python, un autre mécanisme qui peut être utilisé pour gérer ce genre d'objets composites, ce sont les classes que nous verrons en semaine 6, et qui permettent de définir de nouveaux types plutôt que, comme nous l'avons fait ici, d'utiliser un type prédéfini. Dans ce sens, l'utilisation d'une classe permet davantage de souplesse, au prix de davantage d'effort.

### 3.12.2 Complément - niveau avancé

#### La même idée, mais avec une classe `Personne`

Je vais donner ici une implémentation du code ci-dessus, qui utilise une classe pour modéliser les personnes. Naturellement je n'entre pas dans les détails, que l'on verra en semaine 6, mais j'espère vous donner un aperçu des classes dans un usage réaliste, et vous montrer les avantages de cette approche.

Pour commencer je définis la classe `Personne`, qui va me servir à modéliser chaque personne :

```
In [7]: class Personne:

        # le constructeur - vous ignorez le paramètre self,
        # on pourra construire une personne à partir de
        # 3 paramètres
        def __init__(self, nom, age, email):
            self.nom = nom
            self.age = age
            self.email = email

        # je définis cette méthode pour avoir
        # quelque chose de lisible quand je print()
        def __repr__(self):
            return f"{self.nom} ({self.age} ans) sur {self.email}"
```

Pour construire ma liste de personnes, je fais alors :

```
In [8]: personnes2 = [
        Personne('Pierre', 25, 'pierre@example.com'),
        Personne('Paul', 18, 'paul@example.com'),
        Personne('Jacques', 52, 'jacques@example.com'),
    ]
```

Si je regarde un élément de la liste j'obtiens :

```
In [9]: personnes2[0]

Out[9]: Pierre (25 ans) sur pierre@example.com
```

Je peux indexer tout ceci comme tout à l'heure, si j'ai besoin d'un accès rapide :

```
In [10]: # je dois utiliser cette fois personne.nom et non plus personne['nom']
        index2 = {personne.nom : personne for personne in personnes2}
```

Le principe ici est exactement identique à ce qu'on a fait avec le dictionnaire de dictionnaires, mais on a construit un dictionnaire d'instances.

Et de cette façon :

```
In [11]: print(index2['Pierre'])

Pierre (25 ans) sur pierre@example.com
```

Rendez-vous en semaine 6 pour approfondir la notion de classes et d'instances.

### 3.13 Dictionnaires et listes

#### 3.13.1 Exercice - niveau basique

```
In [ ]: from corrections.exo_graph_dict import exo_graph_dict
```

On veut implémenter un petit modèle de graphes. Comme on a les données dans des fichiers, on veut analyser des fichiers d'entrée qui ressemblent à ceci :

```
In [ ]: !cat data/graph1.txt
```

qui signifierait :

- un graphe à 3 sommets  $s1$ ,  $s2$  et  $s3$ ;
- et 4 arêtes
  - une entre  $s1$  et  $s2$  de longueur 10;
  - une entre  $s2$  et  $s3$  de longueur 12;
  - etc...

On vous demande d'écrire une fonction qui lit un tel fichier texte, et construit (et retourne) un dictionnaire Python qui représente ce graphe.

Dans cet exercice on choisit :

- de modéliser le graphe comme un dictionnaire indexé sur les (noms de) sommets;
- et chaque valeur est une liste de tuples de la forme (*suivant*, *longueur*), dans l'ordre d'apparition dans le fichier d'entrée.

```
In [ ]: # voici ce qu'on obtiendrait par exemple avec les données ci-dessus
        exo_graph_dict.example()
```

#### Notes

- Vous remarquerez que l'exemple ci-dessus retourne un dictionnaire standard ; une solution qui utiliserait `defaultdict` est acceptable également;
- Notez bien également que dans le résultat, la longueur d'un arc est attendue comme un `int`.

```
In [ ]: # n'oubliez pas d'importer si nécessaire
```

```
    # à vous de jouer
    def graph_dict(filename):
        "votre code"
```

```
In [ ]: exo_graph_dict.correction(graph_dict)
```

## 3.14 Fusionner des données

### 3.14.1 Exercices

Cet exercice vient en deux versions, une de niveau basique et une de niveau intermédiaire.

La version basique est une application de la technique d’indexation que l’on a vue dans le complément “Gérer des enregistrements”. On peut très bien faire les deux versions dans l’ordre, une fois qu’on a fait la version basique on est en principe un peu plus avancé pour aborder la version intermédiaire.

#### Contexte

Nous allons commencer à utiliser des données un peu plus réalistes. Il s’agit de données obtenues auprès de [MarineTraffic](#) - et légèrement simplifiées pour les besoins de l’exercice. Ce site expose les coordonnées géographiques de bateaux observées en mer au travers d’un réseau de collecte de type *crowdsourcing*.

De manière à optimiser le volume de données à transférer, l’API de MarineTraffic offre deux modes pour obtenir les données :

- **mode étendu** : chaque mesure (bateau x position x temps) est accompagnée de tous les détails du bateau (id, nom, pays de rattachement, etc.);
- **mode abrégé** : chaque mesure est uniquement attachée à l’id du bateau.

En effet, chaque bateau possède un identifiant unique qui est un entier, que l’on note id.

#### Chargement des données

Commençons par charger les données de l’exercice :

```
In [ ]: from corrections.exo_marine_dict import extended, abbreviated
```

#### Format des données

Le format de ces données est relativement simple, il s’agit dans les deux cas d’une liste d’entrées - une par bateau.

Chaque entrée à son tour est une liste qui contient :

```
mode étendu: [id, latitude, longitude, date_heure, nom_bateau, code_pays, ...]
mode abrégé: [id, latitude, longitude, date_heure]
```

sachant que les entrées après le code pays dans le format étendu ne nous intéressent pas pour cet exercice.

```
In [ ]: # une entrée étendue est une liste qui ressemble à ceci
        sample_extended_entry = extended[3]
        print(sample_extended_entry)

In [ ]: # une entrée abrégée ressemble à ceci
        sample_abbreviated_entry = abbreviated[0]
        print(sample_abbreviated_entry)
```

On précise également que les deux listes `extended` et `abbreviated` :



- possèdent exactement **le même nombre** d'entrées ;
- et correspondent **aux mêmes bateaux** ;
- mais naturellement **à des moments différents** ;
- et **pas forcément dans le même ordre**.

### Exercice - niveau basique

```
In [ ]: # chargement de l'exercice
        from corrections.exo_marine_dict import exo_index
```

**But de l'exercice** On vous demande d'écrire une fonction `index` qui calcule, à partir de la liste des données étendues, un dictionnaire qui est :

- indexé par l'id de chaque bateau ;
- et qui a pour valeur la liste qui décrit le bateau correspondant.

De manière plus imagée, si :

```
extended = [ bateau1, bateau2, ... ]
```

Et si :

```
bateau1 = [ id1, latitude, ... ]
```

On doit obtenir comme résultat de `index` un dictionnaire :

```
{
    id1 -> [ id_bateau1, latitude, ... ],
    id2 ...
}
```

Bref, on veut pouvoir retrouver les différents éléments de la liste `extended` par accès direct, en ne faisant qu'une seule recherche dans l'index.

```
In [ ]: # le résultat attendu
        result_index = exo_index.resultat(extended)

        # on en profite pour illustrer le module pprint
        from pprint import pprint

        # à quoi ressemble le résultat pour un bateau au hasard
        for key, value in result_index.items():
            print("=== clé")
            pprint(key)
            print("=== valeur")
            pprint(value)
            break
```

Remarquez ci-dessus l'utilisation d'un utilitaire parfois pratique : le [module pprint](#) pour [pretty-printer](#).

### Votre code

```
In [ ]: def index(extended):
        "<votre_code>"
```

**Validation**

```
In [ ]: exo_index.correction(index, abbreviated)
```

Vous remarquerez d'ailleurs que la seule chose que l'on utilise dans cet exercice, c'est que l'id des bateaux arrive en première position (dans la liste qui matérialise le bateau), aussi votre code doit marcher à l'identique avec les bateaux étendus :

```
In [ ]: exo_index.correction(index, extended)
```

**Exercice - niveau intermédiaire**

```
In [ ]: # chargement de l'exercice
        from corrections.exo_marine_dict import exo_merge
```

**But de l'exercice** On vous demande d'écrire une fonction merge qui fasse une consolidation des données, de façon à obtenir en sortie un dictionnaire :

```
id -> [nom_bateau, code_pays, position_etendu, position_abrege]
```

dans lequel les deux objets position sont tous les deux des tuples de la forme :

```
(latitude, longitude, date_heure)
```

Voici par exemple un couple clé-valeur dans le résultat attendu :

```
In [ ]: # le résultat attendu
        result_merge = exo_merge.resultat(extended, abbreviated)

        # à quoi ressemble le résultat pour un bateau au hasard
        from pprint import pprint
        for key_value in result_merge.items():
            pprint(key_value)
            break
```

**Votre code**

```
In [ ]: def merge(extended, abbreviated):
        "votre code"
```

**Validation**

```
In [ ]: exo_merge.correction(merge, extended, abbreviated)
```

### Les fichiers de données complets

Signalons enfin pour ceux qui sont intéressés que les données chargées dans cet exercice sont disponibles au format JSON - qui est précisément celui exposé par marinetraffic.

Nous avons beaucoup simplifié les données d'entrée pour vous permettre une mise au point plus facile. Si vous voulez vous amuser à charger des données un peu plus significatives, sachez que :

- vous avez accès aux fichiers de données plus complets :
  - `data/marine-e1-ext.json`
  - `data/marine-e1-abb.json`
- pour charger ces fichiers, qui sont donc au [format JSON](#), la connaissance intime de ce format n'est pas nécessaire, on peut tout simplement utiliser le [module json](#). Voici le code utilisé dans l'exercice pour charger ces JSON en mémoire ; il utilise des notions que nous verrons dans les semaines à venir :

```
In [ ]: # load data from files
import json

with open("data/marine-e1-ext.json", encoding="utf-8") as feed:
    extended_full = json.load(feed)

with open("data/marine-e1-abb.json", encoding="utf-8") as feed:
    abbreviated_full = json.load(feed)
```

Une fois que vous avez un code qui fonctionne vous pouvez le lancer sur ces données plus copieuses en faisant :

```
In [ ]: exo_merge.correction(merge, extended_full, abbreviated_full)
```

## 3.15 Ensembles

### 3.15.1 Complément - niveau basique

Ce document résume les opérations courantes disponibles sur le type `set`. On rappelle que le type `set` est un type **mutable**.

#### Création en extension

On crée un ensemble avec les accolades, comme les dictionnaires, mais sans utiliser le caractère `:`, et cela donne par exemple :

```
In [1]: heteroclite = {'marc', 12, 'pierre', (1, 2, 3), 'pierre'}
        print(heteroclite)

{'pierre', 12, 'marc', (1, 2, 3)}
```

#### Création - la fonction `set`

Il devrait être clair à ce stade que, le nom du type étant `set`, la fonction `set` est un constructeur d'ensemble. On aurait donc aussi bien pu faire :

```
In [2]: heteroclite2 = set(['marc', 12, 'pierre', (1, 2, 3), 'pierre'])
        print(heteroclite2)

{'pierre', 12, 'marc', (1, 2, 3)}
```

#### Créer un ensemble vide

Il faut remarquer que l'on ne peut pas créer un ensemble vide en extension. En effet :

```
In [3]: type({})

Out[3]: dict
```

Ceci est lié à des raisons historiques, les ensembles n'ayant fait leur apparition que tardivement dans le langage en tant que citoyen de première classe.

Pour créer un ensemble vide, la pratique la plus courante est celle-ci :

```
In [4]: ensemble_vide = set()
        print(type(ensemble_vide))

<class 'set'>
```

Ou également, moins élégant mais que l'on trouve parfois dans du vieux code :

```
In [5]: autre_ensemble_vide = set([])
        print(type(autre_ensemble_vide))

<class 'set'>
```

### Un élément dans un ensemble doit être globalement immuable

On a vu précédemment que les clés dans un dictionnaire doivent être globalement immuables. Pour exactement les mêmes raisons, les éléments d'un ensemble doivent aussi être globalement immuables :

```
# on ne peut pas insérer un tuple qui contient une liste
>>> ensemble = {(1, 2, [3, 4])}
Traceback (most recent call last):
  File "<stdin>", line 1, in <module>
TypeError: unhashable type: 'list'
```

Le type set étant lui-même mutable, on ne peut pas créer un ensemble d'ensembles :

```
>>> ensemble = {{1, 2}}
Traceback (most recent call last):
  File "<stdin>", line 1, in <module>
TypeError: unhashable type: 'set'
```

Et c'est une des raisons d'être du type frozenset.

### Création - la fonction frozenset

Un frozenset est un ensemble qu'on ne peut pas modifier, et qui donc peut servir de clé dans un dictionnaire, ou être inclus dans un autre ensemble (mutable ou pas).

Il n'existe pas de raccourci syntaxique comme les {} pour créer un ensemble immuable, qui doit être créé avec la fonction frozenset. Toutes les opérations documentées dans ce notebook, et qui n'ont pas besoin de modifier l'ensemble, sont disponibles sur un frozenset.

Parmi les fonctions exclues sur un frozenset, on peut citer : update, pop, clear, remove ou discard.

### Opérations simples

```
In [6]: # pour rappel
        heteroclite

Out[6]: {(1, 2, 3), 12, 'marc', 'pierre'}
```

### Test d'appartenance

```
In [7]: (1, 2, 3) in heteroclite

Out[7]: True
```

### Cardinal

```
In [8]: len(heteroclite)

Out[8]: 4
```

**Manipulations**

```
In [9]: ensemble = {1, 2, 1}
        ensemble
```

```
Out[9]: {1, 2}
```

```
In [10]: # pour nettoyer
         ensemble.clear()
         ensemble
```

```
Out[10]: set()
```

```
In [11]: # ajouter un element
         ensemble.add(1)
         ensemble
```

```
Out[11]: {1}
```

```
In [12]: # ajouter tous les elements d'un autre *ensemble*
         ensemble.update({2, (1, 2, 3), (1, 3, 5)})
         ensemble
```

```
Out[12]: {(1, 2, 3), (1, 3, 5), 1, 2}
```

```
In [13]: # enlever un element avec discard
         ensemble.discard((1, 3, 5))
         ensemble
```

```
Out[13]: {(1, 2, 3), 1, 2}
```

```
In [14]: # discard fonctionne même si l'élément n'est pas présent
         ensemble.discard('foo')
         ensemble
```

```
Out[14]: {(1, 2, 3), 1, 2}
```

```
In [15]: # enlever un élément avec remove
         ensemble.remove((1, 2, 3))
         ensemble
```

```
Out[15]: {1, 2}
```

```
In [16]: # contrairement à discard, l'élément doit être présent,
         # sinon il y a une exception
         try:
             ensemble.remove('foo')
         except KeyError as e:
             print("remove a levé l'exception", e)
```

```
remove a levé l'exception 'foo'
```

La capture d'exception avec `try` et `except` sert à capturer une erreur d'exécution du programme (que l'on appelle exception) pour continuer le programme. Le but de cet exemple est simplement de montrer (d'une manière plus élégante que de voir simplement le programme planter avec une exception non capturée) que l'expression `ensemble.remove('foo')` génère une exception. Si ce concept vous paraît obscur, pas d'inquiétude, nous l'aborderons cette semaine et nous y reviendrons en détail en semaine 6.

```
In [17]: # pop() ressemble à la méthode éponyme sur les listes
         # sauf qu'il n'y a pas d'ordre dans un ensemble
         while ensemble:
             element = ensemble.pop()
             print("element", element)
         print("et bien sûr maintenant l'ensemble est vide", ensemble)

element 1
element 2
et bien sûr maintenant l'ensemble est vide set()
```

### Opérations classiques sur les ensembles

Donnons-nous deux ensembles simples :

```
In [18]: A2 = set([0, 2, 4, 6])
         print('A2', A2)
         A3 = set([0, 6, 3])
         print('A3', A3)
```

```
A2 {0, 2, 4, 6}
A3 {0, 3, 6}
```

N'oubliez pas que les ensembles, comme les dictionnaires, ne sont **pas ordonnés**.

#### Remarques :

- les notations des opérateurs sur les ensembles rappellent les opérateurs “bit-à-bit” sur les entiers;
- ces opérateurs sont également disponibles sous la forme de méthodes.

#### Union

```
In [19]: A2 | A3
```

```
Out[19]: {0, 2, 3, 4, 6}
```

#### Intersection

```
In [20]: A2 & A3
```

```
Out[20]: {0, 6}
```

**Différence**

```
In [21]: A2 - A3
```

```
Out[21]: {2, 4}
```

```
In [22]: A3 - A2
```

```
Out[22]: {3}
```

**Différence symétrique** On rappelle que  $A \Delta B = (A - B) \cup (B - A)$

```
In [23]: A2 ^ A3
```

```
Out[23]: {2, 3, 4}
```

**Comparaisons**

Ici encore on se donne deux ensembles :

```
In [24]: superset = {0, 1, 2, 3}
         print('superset', superset)
         subset = {1, 3}
         print('subset', subset)
```

```
superset {0, 1, 2, 3}
```

```
subset {1, 3}
```

**Égalité**

```
In [25]: heteroclite == heteroclite2
```

```
Out[25]: True
```

**Inclusion**

```
In [26]: subset <= superset
```

```
Out[26]: True
```

```
In [27]: subset < superset
```

```
Out[27]: True
```

```
In [28]: heteroclite < heteroclite2
```

```
Out[28]: False
```

**Ensembles disjoints**

```
In [29]: heteroclite.isdisjoint(A3)
```

```
Out[29]: True
```

**Pour en savoir plus**

Reportez vous à [la section sur les ensembles](#) dans la documentation Python.



## 3.16 Ensembles

### 3.16.1 Exercice - niveau basique

```
In [ ]: # charger l'exercice
        from corrections.exo_read_set import exo_read_set
```

On se propose d'écrire une fonction `read_set` qui construit un ensemble à partir du contenu d'un fichier. Voici par exemple un fichier d'entrée :

```
In [ ]: !cat data/setref1.txt
```

`read_set` va prendre en argument un nom de fichier (vous pouvez supposer qu'il existe), enlever les espaces éventuelles au début et à la fin de chaque ligne, et construire un ensemble de toutes les lignes ; par exemple :

```
In [ ]: exo_read_set.example()
```

```
In [ ]: # écrivez votre code ici
        def read_set(filename):
            "votre code"
```

```
In [ ]: # vérifiez votre code ici
        exo_read_set.correction(read_set)
```

### 3.16.2 Deuxième partie - niveau basique

```
In [ ]: # la définition de l'exercice
        from corrections.exo_read_set import exo_search_in_set
```

Ceci étant acquis, on veut écrire une deuxième fonction `search_in_set` qui prend en argument deux fichiers :

- `filename_reference` est le nom d'un fichier contenant des mots de référence ;
- `filename` est le nom d'un fichier contenant des mots, dont on veut savoir s'ils sont ou non dans les références.

Pour cela `search_in_set` doit retourner une liste, contenant pour chaque ligne du fichier `filename` un tuple avec :

- la ligne (sans les espaces de début et de fin, ni la fin de ligne) ;
- un booléen qui indique si ce mot est présent dans les références ou pas.

Par exemple :

```
In [ ]: !cat data/setref1.txt
```

```
In [ ]: !cat data/setsample1.txt
```

```
In [ ]: exo_search_in_set.example()
```

```
In [ ]: # à vous
        def search_in_set(filename_reference, filename):
            "votre code"
```

```
In [ ]: # vérifiez
        exo_search_in_set.correction(search_in_set)
```

## 3.17 Exercice sur les ensembles

### 3.17.1 Exercice - niveau intermédiaire

```
In [ ]: # chargement de l'exercice
        from corrections.exo_marine_set import exo_diff
```

#### Les données

Nous reprenons le même genre de données marines en provenance de MarineTraffic que nous avons vues dans l'exercice précédent.

```
In [ ]: from corrections.exo_marine_set import abbreviated, extended
```

#### Rappels sur les formats

étendu: [id, latitude, longitude, date\_heure, nom\_bateau, code\_pays...]  
abrégé: [id, latitude, longitude, date\_heure]

```
In [ ]: print(extended[0])
```

```
In [ ]: print(abbreviated[0])
```

#### But de l'exercice

```
In [ ]: # chargement de l'exercice
        from corrections.exo_marine_set import exo_diff
```

Notez bien une différence importante avec l'exercice précédent : cette fois **il n'y a plus correspondance** entre les bateaux rapportés dans les données étendues et abrégées.

Le but de l'exercice est précisément d'étudier la différence, et pour cela on vous demande d'écrire une fonction

```
diff(extended, abbreviated)
```

qui retourne un tuple à trois éléments :

- l'ensemble (set) des **noms** des bateaux présents dans extended mais pas dans abbreviated;
- l'ensemble des **noms** des bateaux présents dans extended et dans abbreviated;
- l'ensemble des **id** des bateaux présents dans abbreviated mais pas dans extended (par construction, les données ne nous permettent pas d'obtenir les noms de ces bateaux).

```
In [ ]: # le résultat attendu
        result = exo_diff.resultat(extended, abbreviated)

        # combien de bateaux sont concernés
        def show_result(extended, abbreviated, result):
            """
            Affiche divers décomptes sur les arguments
            en entrée et en sortie de diff
            """
            print(10*'- ', "Les entrées")
```

```

print(f"Dans extended: {len(extended)} entrées")
print(f"Dans abbreviated: {len(abbreviated)} entrées")
print(10*'- ', "Le résultat du diff")
extended_only, both, abbreviated_only = result
print(f"Dans extended mais pas dans abbreviated {len(extended_only)}")
print(f"Dans les deux {len(both)}")
print(f"Dans abbreviated mais pas dans extended {len(abbreviated_only)}")

show_result(extended, abbreviated, result)

```

### Votre code

```

In [ ]: def diff(extended, abbreviated):
        "<votre_code>"

```

### Validation

```

In [ ]: exo_diff.correction(diff, extended, abbreviated)

```

### Des fichiers de données plus réalistes

Comme pour l'exercice précédent, les données fournies ici sont très simplistes; vous pouvez, si vous le voulez, essayer votre code avec des données (un peu) plus réalistes en chargeant des fichiers de données plus complets :

- `data/marine-e2-ext.json`
- `data/marine-e2-abb.json`

Ce qui donnerait en Python :

```

In [ ]: # load data from files
import json

with open("data/marine-e2-ext.json", encoding="utf-8") as feed:
    extended_full = json.load(feed)

with open("data/marine-e2-abb.json", encoding="utf-8") as feed:
    abbreviated_full = json.load(feed)

In [ ]: # le résultat de votre fonction sur des données plus vastes
# attention, show_result fait des hypothèses sur le type de votre résultat
# aussi si vous essayez d'exécuter ceci avec comme fonction diff
# la version vide qui est dans le notebook original
# cela peut provoquer une exception
diff_full = diff(extended_full, abbreviated_full)
show_result(extended_full, abbreviated_full, diff_full)

```

Je signale enfin à propos de ces données plus complètes que :

- on a supprimé les entrées correspondant à des bateaux différents mais de même nom; cette situation peut arriver dans la réalité (c'est pourquoi d'ailleurs les bateaux ont un *id*) mais ici ce n'est pas le cas;
- il se peut par contre qu'un même bateau fasse l'objet de plusieurs mesures dans `extended` et/ou dans `abbreviated`.

## 3.18 try ... else ... finally

### 3.18.1 Complément - niveau intermédiaire

L'instruction `try` est généralement assortie d'une ou plusieurs clauses `except`, comme on l'a vu dans la vidéo.

Sachez que l'on peut aussi utiliser - après toutes les clauses `except` :

- une clause `else`, qui va être exécutée si aucune exception n'est attrapée ;
- t/ou une clause `finally` qui sera alors exécutée quoi qu'il arrive.

Voyons cela sur des exemples.

`finally`

C'est sans doute `finally` qui est la plus utile de ces deux clauses, car elle permet de faire un nettoyage **dans tous les cas de figure** - de ce point de vue, cela rappelle un peu les *context managers*.

Et par exemple, comme avec les *context managers*, une fonction peut faire des choses même après un `return`.

```
In [1]: # une fonction qui fait des choses après un return
def return_with_finally(number):
    try:
        return 1/number
    except ZeroDivisionError as e:
        print(f"OOPS, {type(e)}, {e}")
        return("zero-divide")
    finally:
        print("on passe ici même si on a vu un return")
```

```
In [2]: # sans exception
        return_with_finally(1)
```

on passe ici même si on a vu un return

```
Out[2]: 1.0
```

```
In [3]: # avec exception
        return_with_finally(0)
```

```
OOPS, <class 'ZeroDivisionError'>, division by zero
on passe ici même si on a vu un return
```

```
Out[3]: 'zero-divide'
```

else

La logique ici est assez similaire, sauf que le code du `else` n'est exécuté que dans le cas où aucune exception n'est attrapée.

En première approximation, on pourrait penser que c'est équivalent de mettre du code dans la clause `else` ou à la fin de la clause `try`. En fait il y a une différence subtile :

*The use of the `else` clause is better than adding additional code to the `try` clause because it avoids accidentally catching an exception that wasn't raised by the code being protected by the `try... except` statement.*

Dit autrement, si le code dans la clause `else` lève une exception, celle-ci ne **sera pas attrapée** par le `try` courant, et sera donc propagée.

Voici un exemple rapide, en pratique on rencontre assez peu souvent une clause `else` dans un `try` :

```
In [4]: # pour montrer la clause else dans un usage banal
```

```
def function_with_else(number):
    try:
        x = 1/number
    except ZeroDivisionError as e:
        print(f"OOPS, {type(e)}, {e}")
    else:
        print("on passe ici seulement avec un nombre non nul")
    return 'something else'
```

```
In [5]: # sans exception
```

```
function_with_else(1)
```

on passe ici seulement avec un nombre non nul

```
Out[5]: 'something else'
```

```
In [6]: # avec exception
```

```
function_with_else(0)
```

```
OOPS, <class 'ZeroDivisionError'>, division by zero
```

```
Out[6]: 'something else'
```

Remarquez que `else` ne présente pas cette particularité de “traverser” le `return`, que l’on a vue avec `finally` :

```
In [7]: # la clause else ne traverse pas les return
```

```
def return_with_else(number):
    try:
        return 1/number
    except ZeroDivisionError as e:
        print(f"OOPS, {type(e)}, {e}")
        return("zero-divide")
    else:
        print("on ne passe jamais ici à cause des return")
```

```
In [8]: # sans exception  
        return_with_else(1)
```

```
Out[8]: 1.0
```

```
In [9]: # avec exception  
        return_with_else(0)
```

```
OOPS, <class 'ZeroDivisionError'>, division by zero
```

```
Out[9]: 'zero-divide'
```

### Pour en savoir plus

Voyez [le tutorial sur les exceptions](#) dans la documentation officielle.

## 3.19 L'opérateur is

### 3.19.1 Complément - niveau basique

```
In [ ]: %load_ext ipythontutor
```

Les opérateurs `is` et `==`

- nous avons déjà parlé de l'opérateur `==` qui **compare la valeur** de deux objets ;
- python fournit aussi un opérateur `is` qui permet de savoir si deux valeurs correspondent **au même objet** en mémoire.

Nous allons illustrer la différence entre ces deux opérateurs.

#### Scénario 1

```
In [1]: # deux listes identiques
a = [1, 2]
b = [1, 2]

# les deux objets se ressemblent
print('==', a == b)
```

`== True`

```
In [2]: # mais ce ne sont pas les mêmes objets
print('is', a is b)
```

`is False`

#### Scénario 2

```
In [3]: # par contre ici il n'y a qu'une liste
a = [1, 2]

# et les deux variables
# référencent le même objet
b = a

# non seulement les deux expressions se ressemblent
print('==', a == b)
```

`== True`

```
In [4]: # mais elles désignent le même objet
print('is', a is b)
```

`is True`

**La même chose sous pythontutor****Scénario 1**

```
In [ ]: %%ipythontutor curInstr=2
        a = [1, 2]
        b = [1, 2]
```

**Scénario 2**

```
In [ ]: %%ipythontutor curInstr=1
        # équivalent à la forme ci-dessus
        # a = [1, 2]
        # b = a
        a = b = [1, 2]
```

**Utilisez `is` plutôt que `==` lorsque c'est possible**

La pratique usuelle est d'utiliser `is` lorsqu'on compare avec un objet qui est un singleton, comme typiquement `None`.

Par exemple on préférera écrire :

```
In [5]: undef = None

        if undef is None:
            print('indéfini')

indéfini
```

Plutôt que :

```
In [6]: if undef == None:
        print('indéfini')

indéfini
```

Qui se comporte de la même manière (à nouveau, parce qu'on compare avec `None`), mais est légèrement moins lisible, et franchement moins pythonique. :)

Notez aussi et surtout que `is` est **plus efficace** que `==`. En effet `is` peut être évalué en temps constant, puisqu'il s'agit essentiellement de comparer les deux adresses. Alors que pour `==` il peut s'agir de parcourir toute une structure de données possiblement très complexe.

**3.19.2 Complément - niveau intermédiaire****La fonction `id`**

Pour bien comprendre le fonctionnement de `is` nous allons voir la fonction `id` qui retourne un identificateur unique pour chaque objet; un modèle mental acceptable est celui d'adresse mémoire.

```
In [7]: id(True)
```



```
Out[7]: 1924796304
```

Comme vous vous en doutez, l'opérateur `is` peut être décrit formellement à partir de `id` comme ceci :

$$(a \text{ is } b) \iff (\text{id}(a) == \text{id}(b))$$

### Certains types de base sont des singletons

Un singleton est un objet qui n'existe qu'en un seul exemplaire dans la mémoire. Un usage classique des singletons en Python est de minimiser le nombre d'objets immuables en mémoire. Voyons ce que cela nous donne avec des entiers :

```
In [8]: a = 3
        b = 3
        print('a', id(a), 'b', id(b))
```

```
a 1924978512 b 1924978512
```

Tiens, c'est curieux, nous avons ici deux objets, que l'on pourrait penser différents, mais en fait ce sont les mêmes ; `a` et `b` désignent **le même objet** python, et on a :

```
In [9]: a is b
```

```
Out[9]: True
```

Il se trouve que, dans le cas des petits entiers, python réalise une optimisation de l'utilisation de la mémoire. Quel que soit le nombre de variables dont la valeur est 3, un seul objet correspondant à l'entier 3 est alloué et créé, pour éviter d'engorger la mémoire. On dit que l'entier 3 est implémenté comme un singleton ; nous reverrons ceci en exercice.

On trouve cette optimisation avec quelques autres objets python, comme par exemple :

```
In [10]: a = ""
         b = ""
         a is b
```

```
Out[10]: True
```

Ou encore, plus surprenant :

```
In [11]: a = "foo"
         b = "foo"
         a is b
```

```
Out[11]: True
```

**Conclusion** cette optimisation ne touche aucun type mutable (heureusement) ; pour les types immuables, il n'est pas extrêmement important de savoir en détail quels objets sont implémentés de la sorte.

Ce qui est par contre extrêmement important est de comprendre la différence entre `is` et `==`, et de les utiliser à bon escient au risque d'écrire du code fragile.

**Pour en savoir plus**

Aux étudiants de niveau avancé, nous recommandons la lecture de la section “Objects, values and types” dans la documentation Python :

<https://docs.python.org/3/reference/datamodel.html#objects-values-and-types>

qui aborde également la notion de “garbage collection”, que nous n’aurons pas le temps d’approfondir dans ce MOOC.

## 3.20 Listes infinies & références circulaires

### 3.20.1 Complément - niveau intermédiaire

```
In [ ]: %load_ext ipythontutor
```

Nous allons maintenant construire un objet un peu abscons. Cet exemple précis n'a aucune utilité pratique, mais permet de bien comprendre la logique du langage.

Construisons une liste à un seul élément, peu importe quoi :

```
In [1]: infini_1 = [None]
```

À présent nous allons remplacer le premier et seul élément de la liste par... la liste elle-même :

```
In [2]: infini_1[0] = infini_1
        print(infini_1)
```

```
[[...]]
```

Pour essayer de décrire l'objet liste ainsi obtenu, on pourrait dire qu'il s'agit d'une liste de taille 1 et de profondeur infinie, une sorte de fil infini en quelque sorte.

Naturellement, l'objet obtenu est difficile à imprimer de manière convaincante. Pour faire en sorte que cet objet soit tout de même imprimable, et éviter une boucle infinie, python utilise l'ellipse ... pour indiquer ce qu'on appelle une référence circulaire. Si on n'y prenait pas garde en effet, il faudrait écrire [[[[ etc . ]]]] avec une infinité de crochets.

Voici la même séquence exécutée sous <http://pythontutor.com> ; il s'agit d'un site très utile pour comprendre comment python implémente les objets, les références et les partages.

Cliquez sur le bouton Forward pour avancer dans l'exécution de la séquence. À la fin de la séquence vous verrez - ce n'est pas forcément clair - la seule cellule de la liste à se référencer elle-même :

```
In [ ]: %%ipythontutor height=230
        infini_1 = [None]
        infini_1[0] = infini_1
```

Toutes les fonctions de python ne sont pas aussi intelligentes que print. Bien qu'on puisse comparer cette liste avec elle-même :

```
In [3]: infini_1 == infini_1
```

```
Out[3]: True
```

il n'en est pas de même si on la compare avec un objet analogue mais pas identique :

```
In [4]: infini_2 = [0]
        infini_2[0] = infini_2
        print(infini_2)
        infini_1 == infini_2
```

```
[[...]]
```

```
-----

RecursionError                                Traceback (most recent call last)

<ipython-input-4-3c044c24fccb> in <module>()
      2 infini_2[0] = infini_2
      3 print(infini_2)
----> 4 infini_1 == infini_2

RecursionError: maximum recursion depth exceeded in comparison
```

### Généralisation aux références circulaires

On obtient un phénomène équivalent dès lors qu'un élément contenu dans un objet fait référence à l'objet lui-même. Voici par exemple comment on peut construire un dictionnaire qui contient une référence circulaire :

```
In [5]: collection_de_points = [
        {'x': 10, 'y': 20},
        {'x': 30, 'y': 50},
        # imaginez plein de points
    ]

    # on rajoute dans chaque dictionnaire une clé 'points'
    # qui référence la collection complète
    for point in collection_de_points:
        point['points'] = collection_de_points

    # la structure possède maintenant des références circulaires
    print(collection_de_points)

[{'x': 10, 'y': 20, 'points': [...]}, {'x': 30, 'y': 50, 'points': [...]}]
```

On voit à nouveau réapparaître les ellipses, qui indiquent que pour chaque point, le nouveau champ `points` est un objet qui a déjà été imprimé.

Cette technique est cette fois très utile et très utilisée dans la pratique, dès lors qu'on a besoin de naviguer de manière arbitraire dans une structure de données compliquée. Dans cet exemple, pas très réaliste naturellement, on pourrait à présent accéder depuis un point à tous les autres points de la collection dont il fait partie.

À nouveau il peut être intéressant de voir le comportement de cet exemple avec <http://pythontutor.com> pour bien comprendre ce qui se passe, si cela ne vous semble pas clair à première vue :

```
In [ ]: %%ipythontutor curInstr=7
        points = [
            {'x': 10, 'y': 20},
            {'x': 30, 'y': 50},
        ]

        for point in points:
            point['points'] = points
```

## 3.21 Les différentes copies

```
In [ ]: %load_ext ipythontutor
```

### 3.21.1 Complément - niveau basique

#### Deux types de copie

Pour résumer les deux grands types de copie que l'on a vus dans la vidéo :

- La *shallow copy* - de l'anglais *shallow* qui signifie superficiel ;
- La *deep copy* - de *deep* qui signifie profond.

#### Le module `copy`

Pour réaliser une copie, la méthode la plus simple, en ceci qu'elle fonctionne avec tous les types de manière identique, consiste à utiliser [le module standard `copy`](#), et notamment :

- `copy.copy` pour une copie superficielle ;
- `copy.deepcopy` pour une copie en profondeur.

```
In [1]: import copy
        #help(copy.copy)
        #help(copy.deepcopy)
```

#### Un exemple

Nous allons voir le résultat des deux formes de copie sur un même sujet de départ.

#### La copie superficielle / *shallow copy* / `copy.copy`

N'oubliez pas de cliquer le bouton Forward dans la fenêtre pythontutor :

```
In [ ]: %%ipythontutor height=410 curInstr=6
import copy
# On se donne un objet de départ
source = [
    [1, 2, 3], # une liste
    {1, 2, 3}, # un ensemble
    (1, 2, 3), # un tuple
    '123',     # un string
    123,       # un entier
]
# une copie simple renvoie ceci
shallow_copy = copy.copy(source)
```

Vous remarquez que :

- la source et la copie partagent tous leurs (sous-)éléments, et notamment la liste `source[0]` et l'ensemble `source[1]` ;
- ainsi, après cette copie, on peut modifier l'un de ces deux objets (la liste ou l'ensemble), et ainsi modifier la source **et** la copie.

On rappelle aussi que, la source étant une liste, on aurait pu aussi bien faire la copie superficielle avec

```
shallow2 = source[:]
```

**La copie profonde / *deep copy* / `copy.deepcopy`** Sur le même objet de départ, voici ce que fait la copie profonde :

```
In [ ]: %%ipythontutor height=410 curInstr=6
import copy
# On se donne un objet de départ
source = [
    [1, 2, 3], # une liste
    {1, 2, 3}, # un ensemble
    (1, 2, 3), # un tuple
    '123',     # un string
    123,       # un entier
]
# une copie profonde renvoie ceci
deep_copy = copy.deepcopy(source)
```

Ici, il faut remarquer que :

- les deux objets mutables accessibles via `source`, c'est-à-dire **la liste** `source[0]` et **l'ensemble** `source[1]`, ont été tous deux dupliqués;
- **le tuple** correspondant à `source[2]` n'est **pas dupliqué**, mais comme il n'est **pas mutable** on ne peut pas modifier la copie au travers de la source;
- de manière générale, on a la bonne propriété que la source et sa copie ne partagent rien qui soit modifiable;
- et donc on ne peut pas modifier l'un au travers de l'autre.

On retrouve donc à nouveau l'optimisation qui est mise en place dans python pour implémenter les types immuables comme des singletons lorsque c'est possible. Cela a été vu en détail dans le complément consacré à l'opérateur `is`.

### 3.21.2 Complément - niveau intermédiaire

```
In [2]: # on répète car le code précédent a seulement été exposé à pythontutor
import copy
source = [
    [1, 2, 3], # une liste
    {1, 2, 3}, # un ensemble
    (1, 2, 3), # un tuple
    '123',     # un string
    123,       # un entier
]
shallow_copy = copy.copy(source)
deep_copy = copy.deepcopy(source)
```

#### Objets égaux au sens logique

Bien sûr ces trois objets se ressemblent si on fait une comparaison *logique* avec `==` :

```
In [3]: print('source == shallow_copy:', source == shallow_copy)
        print('source == deep_copy:', source == deep_copy)
```

```
source == shallow_copy: True
source == deep_copy: True
```

### Inspectons les objets de premier niveau

Mais par contre si on compare l'**identité** des objets de premier niveau, on voit que source et shallow\_copy partagent leurs objets :

```
In [4]: # voir la cellule ci-dessous si ceci vous parait peu clair
        for i, (source_item, copy_item) in enumerate(zip(source, shallow_copy)):
            compare = source_item is copy_item
            print(f"source[{i}] is shallow_copy[{i}] -> {compare}")

source[0] is shallow_copy[0] -> True
source[1] is shallow_copy[1] -> True
source[2] is shallow_copy[2] -> True
source[3] is shallow_copy[3] -> True
source[4] is shallow_copy[4] -> True
```

```
In [5]: # rappel au sujet de zip et enumerate
        # la cellule ci-dessous est essentiellement équivalente à
        for i in range(len(source)):
            compare = source[i] is shallow_copy[i]
            print(f"source[{i}] is shallow_copy[{i}] -> {compare}")

source[0] is shallow_copy[0] -> True
source[1] is shallow_copy[1] -> True
source[2] is shallow_copy[2] -> True
source[3] is shallow_copy[3] -> True
source[4] is shallow_copy[4] -> True
```

Alors que naturellement ce **n'est pas le cas** avec la copie en profondeur :

```
In [6]: for i, (source_item, deep_item) in enumerate(zip(source, deep_copy)):
        compare = source_item is deep_item
        print(f"source[{i}] is deep_copy[{i}] -> {compare}")

source[0] is deep_copy[0] -> False
source[1] is deep_copy[1] -> False
source[2] is deep_copy[2] -> True
source[3] is deep_copy[3] -> True
source[4] is deep_copy[4] -> True
```

On retrouve ici ce qu'on avait déjà remarqué sous pythontutor, à savoir que les trois derniers objets - immuables - n'ont pas été dupliqués comme on aurait pu s'y attendre.

### On modifie la source

Il doit être clair à présent que, précisément parce que deep\_copy est une copie en profondeur, on peut modifier source sans impacter du tout deep\_copy.

S'agissant de shallow\_copy, par contre, seuls les éléments de premier niveau ont été copiés. Aussi si on fait une modification par exemple à l'**intérieur** de la liste qui est le premier fils de source, cela sera **répercuté** dans shallow\_copy :



```
In [7]: print("avant, source      ", source)
        print("avant, shallow_copy", shallow_copy)
        source[0].append(4)
        print("après, source      ", source)
        print("après, shallow_copy", shallow_copy)

avant, source      [[1, 2, 3], {1, 2, 3}, (1, 2, 3), '123', 123]
avant, shallow_copy [[1, 2, 3], {1, 2, 3}, (1, 2, 3), '123', 123]
après, source      [[1, 2, 3, 4], {1, 2, 3}, (1, 2, 3), '123', 123]
après, shallow_copy [[1, 2, 3, 4], {1, 2, 3}, (1, 2, 3), '123', 123]
```

Si par contre on remplace complètement un élément de premier niveau dans la source, cela ne sera pas répercuté dans la copie superficielle :

```
In [8]: print("avant, source      ", source)
        print("avant, shallow_copy", shallow_copy)
        source[0] = 'remplacement'
        print("après, source      ", source)
        print("après, shallow_copy", shallow_copy)

avant, source      [[1, 2, 3, 4], {1, 2, 3}, (1, 2, 3), '123', 123]
avant, shallow_copy [[1, 2, 3, 4], {1, 2, 3}, (1, 2, 3), '123', 123]
après, source      ['remplacement', {1, 2, 3}, (1, 2, 3), '123', 123]
après, shallow_copy [[1, 2, 3, 4], {1, 2, 3}, (1, 2, 3), '123', 123]
```

### Copie et circularité

Le module `copy` est capable de copier - même en profondeur - des objets contenant des références circulaires.

```
In [9]: l = [None]
        l[0] = l
        l

Out[9]: [[...]]

In [10]: copy.copy(l)

Out[10]: [[[...]]]

In [11]: copy.deepcopy(l)

Out[11]: [[...]]
```

### Pour en savoir plus

On peut se reporter à [la section sur le module `copy`](#) dans la documentation Python.

## 3.22 L'instruction `del`

### 3.22.1 Complément - niveau basique

Voici un récapitulatif sur l'instruction `del` selon le contexte dans lequel elle est utilisée.

#### Sur une variable

On peut annuler la définition d'une variable, avec `del`.

Pour l'illustrer, nous utilisons un bloc `try ... except ...` pour attraper le cas échéant l'exception `NameError`, qui est produite lorsqu'on référence une variable qui n'est pas définie.

```
In [1]: # la variable a n'est pas définie
        try:
            print('a=', a)
        except NameError as e:
            print("a n'est pas définie")
```

a n'est pas définie

```
In [2]: # on la définit
        a = 10

        # aucun souci ici, l'exception n'est pas levée
        try:
            print('a=', a)
        except NameError as e:
            print("a n'est pas définie")
```

a= 10

```
In [3]: # maintenant on peut effacer la variable
        del a

        # c'est comme si on ne l'avait pas définie
        # dans la cellule précédente
        try:
            print('a=', a)
        except NameError as e:
            print("a n'est pas définie")
```

a n'est pas définie

#### Sur une liste

On peut enlever d'une liste les éléments qui correspondent à une *slice* :

```
In [4]: # on se donne une liste
        l = list(range(12))
        print(l)
```

```
[0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11]
```

```
In [5]: # on considère une slice dans cette liste
        print('slice=', l[2:10:3])

        # voyons ce que ça donne si on efface cette slice
        del l[2:10:3]
        print("après del", l)

slice= [2, 5, 8]
après del [0, 1, 3, 4, 6, 7, 9, 10, 11]
```

### Sur un dictionnaire

Avec `del` on peut enlever une clé, et donc la valeur correspondante, d'un dictionnaire :

```
In [6]: # partons d'un dictionnaire simple
        d = dict(foo='bar', spam='eggs', a='b')
        print(d)

{'foo': 'bar', 'spam': 'eggs', 'a': 'b'}
```

```
In [7]: # on peut enlever une clé avec del
        del d['a']
        print(d)

{'foo': 'bar', 'spam': 'eggs'}
```

### On peut passer plusieurs arguments à `del`

```
In [8]: # Voyons où en sont nos données
        print('l', l)
        print('d', d)

l [0, 1, 3, 4, 6, 7, 9, 10, 11]
d {'foo': 'bar', 'spam': 'eggs'}
```

```
In [9]: # on peut invoquer 'del' avec plusieurs expressions
        # séparées par une virgule
        del l[3:], d['spam']

        print('l', l)
        print('d', d)

l [0, 1, 3]
d {'foo': 'bar'}
```

### Pour en savoir plus

La page sur [l'instruction `del`](#) dans la documentation Python.

### 3.23 Affectation simultanée

#### 3.23.1 Complément - niveau basique

Nous avons déjà parlé de l'affectation par *sequence unpacking* (en Semaine 3, séquence “Les tuples”), qui consiste à affecter à plusieurs variables des “morceaux” d’un objet, comme dans :

```
In [1]: x, y = ['spam', 'egg']
```

Dans ce complément nous allons voir une autre forme de l’affectation, qui consiste à affecter **le même objet** à plusieurs variables. Commençons par un exemple simple :

```
In [2]: a = b = 1
        print('a', a, 'b', b)
```

```
a 1 b 1
```

La raison pour laquelle nous abordons cette construction maintenant est qu’elle a une forte relation avec les références partagées ; pour bien le voir, nous allons utiliser une valeur mutable comme valeur à affecter :

```
In [3]: # on affecte a et b au même objet liste vide
        a = b = []
```

Dès lors nous sommes dans le cas typique d’une référence partagée ; une modification de a va se répercuter sur b puisque ces deux variables désignent **le même objet** :

```
In [4]: a.append(1)
        print('a', a, 'b', b)
```

```
a [1] b [1]
```

Ceci est à mettre en contraste avec plusieurs affectations séparées :

```
In [5]: # si on utilise deux affectations différentes
        a = []
        b = []

        # alors on peut changer a sans changer b
        a.append(1)
        print('a', a, 'b', b)
```

```
a [1] b []
```

On voit que dans ce cas chaque affectation crée une liste vide différente, et les deux variables ne partagent plus de donnée.

D’une manière générale, utiliser l’affectation simultanée vers un objet mutable crée mécaniquement des **références partagées**, aussi vérifiez bien dans ce cas que c’est votre intention.

## 3.24 Les instructions += et autres revisit es

### 3.24.1 Compl ment - niveau interm diaire

Nous avons vu en premi re semaine (S quence “Les types num riques”) une premi re introduction   l’instruction += et ses d riv es comme \*=, \*\*=, etc.

#### Ces constructions ont une d finition   g om trie variable

En C quand on utilise += (ou encore ++) on modifie la m moire en place - historiquement, cet op rateur permettait au programmeur d’aider   l’optimisation du code pour utiliser les instructions assembleur idoines.

Ces constructions en Python s’inspirent clairement de C, aussi dans l’esprit ces constructions devraient fonctionner en **modifiant** l’objet r f renc  par la variable.

Mais les types num riques en Python ne sont **pas mutables**, alors que les listes le sont. Du coup le comportement de += est **diff rent** selon qu’on l’utilise sur un nombre ou sur une liste, ou plus g n ralement selon qu’on l’invoque sur un type mutable ou non. Voyons cela sur des exemples tr s simples :

```
In [1]: # Premier exemple avec un entier
```

```
# on commence avec une r f rence partag e
a = b = 3
a is b
```

```
Out[1]: True
```

```
In [2]: # on utilise += sur une des deux variables
```

```
a += 1

# ceci n'a pas modifi  b
# c'est normal, l'entier n'est pas mutable

print(a)
print(b)
print(a is b)
```

```
4
```

```
3
```

```
False
```

```
In [3]: # Deuxi me exemple, cette fois avec une liste
```

```
# la m me r f rence partag e
a = b = []
a is b
```

```
Out[3]: True
```

```
In [4]: # pareil, on fait += sur une des variables
        a += [1]

        # cette fois on a modifié a et b
        # car += a pu modifier la liste en place
        print(a)
        print(b)
        print(a is b)
```

```
[1]
[1]
True
```

Vous voyez donc que la sémantique de += (c'est bien entendu le cas pour toutes les autres formes d'instructions qui combinent l'affectation avec un opérateur) **est différente** suivant que l'objet référencé par le terme de gauche est **mutable ou immuable**.

Pour cette raison, c'est là une opinion personnelle, cette famille d'instructions n'est pas le trait le plus réussi dans le langage, et je ne recommande pas de l'utiliser.

### Précision sur la définition de +=

Nous avons dit en première semaine, et en première approximation, que :

$x += y$

était équivalent à :

$x = x + y$

Au vu de ce qui précède, on voit que ce n'est **pas tout à fait exact**, puisque :

```
In [5]: # si on fait x += y sur une liste
        # on fait un effet de bord sur la liste
        # comme on vient de le voir

        a = []
        print("avant", id(a))
        a += [1]
        print("après", id(a))
```

```
avant 82877176
après 82877176
```

```
In [6]: # alors que si on fait x = x + y sur une liste
        # on crée un nouvel objet liste

        a = []
        print("avant", id(a))
        a = a + [1]
        print("après", id(a))
```

avant 82981752

après 82877496

Vous voyez donc que vis-à-vis des références partagées, ces deux façons de faire mènent à un résultat différent.

## 3.25 Classe

### 3.25.1 Exercice - niveau basique

```
In [ ]: # charger l'exercice
        from corrections.cls_fifo import exo_fifo
```

On veut implémenter une classe pour manipuler une queue d'événements. La logique de cette classe est que :

- on la crée sans argument;
- on peut toujours ajouter un élément avec la méthode `incoming`;
- et tant que la queue contient des éléments on peut appeler la méthode `outgoing`, qui retourne et enlève un élément dans la queue.

Cette classe s'appelle `Fifo` pour *First In, First Out*, c'est-à-dire que les éléments retournés par `outgoing` le sont dans le même ordre où ils ont été ajoutés.

La méthode `outgoing` retourne `None` lorsqu'on l'appelle sur une pile vide.

```
In [ ]: # voici un exemple de scénario
        exo_fifo.example()
```

```
In [ ]: # vous pouvez définir votre classe ici
        class Fifo:
            def __init__(self):
                "votre code"
            def incoming(self, value):
                "votre code"
            def outgoing(self):
                "votre code"
```

```
In [ ]: # et la vérifier ici
        exo_fifo.correction(Fifo)
```