

Introduction

Vous travaillerez sur des fichiers avec Python dans la SAE11. Ce notebook vous présente donc une introduction sur les formes de fichiers texte ou binaire ainsi que l'ouverture, la lecture et l'écriture dans un fichier avec Python.

Format de fichiers texte ou binaire

De manière générale, tous les fichiers peuvent être classés en deux grands formats : texte et binaire.

Les **fichiers binaires** peuvent stocker n’importe quel type de données comme l’image, la vidéo et l’audio dans le même fichier. C'est par exemple le cas des fichier **png**, **pdf**, **ppt**, **mpeg**, La seule condition qu’ils présentent est d’avoir un programme approprié pour lire ce type de données présentes dans le système.

Les fichiers binaires contiennent un en-tête au sommet utilisé pour stocker les informations qui identifient le contenu du fichier. Habituellement, les en-têtes contiennent le type de fichier et d’autres métadonnées comme la taille et la date de dernière modification. Si l'en-tête d’un fichier binaire est endommagé, vous ne pouvez plus accéder aux données significatives du fichier.

Les **fichiers texte** peuvent être considérés comme une version réduite des fichiers binaires, ne pouvant stocker que des données textuelles. C'est par exemple le cas des fichiers **txt**, **rtf**, **HTML**, **xml**, **json** ou **csv**. Tous les fichiers texte suivent la norme ASCII au minimum (ou un norme plus vaste englobant la norme ASCII comme UTF8), pour stocker des données. Les fichiers texte peuvent être visualisés par n’importe quel éditeur de texte.

Ouverture, lecture et fermeture d'un fichier

Les étapes d’ouverture et de fermeture sont toujours présentes en ce qui concerne les fichiers. Il s’agit d’indiquer au système d’exploitation que le programme souhaite accéder à un fichier et interdire à tout autre programme l’accès à ce fichier. Un autre programme qui souhaiterait créer un fichier du même nom ne le pourrait pas tant que l’étape de fermeture n'est pas exécutée. En revanche, il pourrait tout à fait le lire car la lecture ne perturbe pas l’écriture. Pour lire un fichier, il faut donc tout d'abord l'ouvrir avec la fonction **open(fichier, mode)** :

- fichier** : chemin du fichier (relatif ou absolu)
- mode** : le mode est composé de deux lettres, les droits d'accès (rwx) et l'encodage (bf). Par défaut le fichier est ouvert avec le mode 'rt'.
 - r**, ouverture en lecture seule (READ).
 - w**, ouverture en écriture (WRITE), à chaque ouverture le contenu du fichier est écrasé. Si le fichier n'existe pas, Python le crée.
 - a**, ouverture en écriture avec ajout (APPEND) à la fin du fichier. Si le fichier n'existe pas, Python le crée.
 - x**, ouverture en écriture seulement si le fichier n'existe pas.
 - b**, pour une ouverture en mode binaire.
 - t**, pour une ouverture en mode texte.

Une fois le fichier en lecture ouvert, on peut utiliser les méthodes **read()** qui retournent une chaîne de caractères contenant l'intégralité du fichier ou **readlines()** retournant une liste où chaque élément est une ligne du fichier. Chaque ligne est terminée par `"\n"` qui représente un retour à la ligne.

Pour fermer le flux de fichier on utilise la méthode **close()** sur la variable représentant le flux.

```
In [1]: print('Lecture du contenu du fichier IUT_de_Vélizy.txt qui est un fichier au format texte :\n')
fichier = open("IUT_de_Vélizy.txt", 'rt')
texte = fichier.read()
print(texte)
fichier.close()
fichier = open("IUT_de_Vélizy.txt", 'rt')
lignes = fichier.readlines()
print(lignes)
fichier.close()
```

Lecture du contenu du fichier IUT_de_Vélizy.txt qui est un fichier au format texte :

L'Institut universitaire de technologie de Vélizy (IUT) est une composante de l'université de Versailles-Saint-Quentin-en-Yvelines dans les Yvelines.

Il propose des formations professionnelles en deux ans accessibles après le baccalauréat, les DUT (diplôme universitaire de technologie) et des formations en un an accessibles après un bac+2, les licences professionnelles.

Ces formations sont dispensées dans les domaines des technologies industrielles et tertiaires (industrie, commerce, services).

"L'Institut universitaire de technologie de Vélizy (IUT) est une composante de l'université de Versailles-Saint-Quentin-en-Yvelines dans les Yvelines.\n", '\n', 'Il propose des formations professionnelles en deux ans accessibles après le baccalauréat, les DUT (diplôme universitaire de technologie) et des formations en un an accessibles après un bac+2, les licences professionnelles.\n', '\n', 'Ces formations sont dispensées dans les domaines des technologies industrielles et tertiaires (industrie, commerce, services). \n', ' ']

```
In [5]: print('Lecture du début du contenu du fichier IUT_de_Vélizy.pdf qui est un fichier pdf au format binaire :\n')
fichier = open("IUT_de_Vélizy.pdf", 'rb')
texte = fichier.readlines()
print(texte[0:10])
```

Lecture du début du contenu du fichier IUT_de_Vélizy.pdf qui est un fichier pdf au format binaire :

[b'%PDF-1.3\n', b'%\xc4\xe5\xf2\xe5\xeb\xa7\xf3\xa0\xd0\xc4\xc6\n', b'4 0 obj\n', b'<< /Length 5 0 R /Filter /FlateDecode >>\n', b'stream\n', b'x\x01\xadY\xd9r\xdb6\x14}\xe7W \x95\x1c\x8bV\x04\x11\x1b\x97\xae\xae\x934m\xfa\xae4\x19\xbdUy\xf2L\xda\xe9\xdb\x0f\x19\xff\xffL\xcf\xc5\xbd\$\x01\x91\xb2d\xbb\xceD\x14H\xe0\xe0.\xe7.\xa0\xbe\xa9[\xf5Mm\xdf?\x1au\xf7\xa8*ml\xf0je\xe3\xb7\xce\x1b\xd3\xd27\x17\xff\x4d\x3e\x1d\xa6V\xdadU\xd1?\xfab\xac\xaa\x3;\xab\xee\x1e\xa8\xa9b\x9d\xdaT\xba\xaa\x9c\xda\xdd)\xcf\xb3p\x1b\xb6\x6\x1d\xfeTpF\xed\x1e\x4v\xb7\xb3\n', b'\xdf\xbe\xaa\x5\x9b\xefJ\xb5\xfbW}\xdcE)\x8e@\x17w\x0f\xaa\x876\xcd\x01vmt\x15\x8c0\xb0!\x14a\xff\xc5\xe8V\xad\x16\xee\xac\x96\x17\xcb\xb7e\x81\xel\xf2\xf2m\xa9p\xc5\x6\x8\xbc\x8d\xaf\xe2\xe3}\x19/W<\xf9b\xb9\x8e\xc3\x8b+\xb9\x7f\xf9.\xce\xeeGK|)\x00\xba\x89\xb3t\x82\xdf+\xbe\x18\x9d\x82\xb1\x0f01|\xe8\x80P\xac\xaa\x0b\x1f\x91B\xa9\xbe\xa8\xdd\xe7\x13\x06\x81\xad\x9_\xa3\x85\x8d\xe9:\xedC +\x14\xb9\x85\xa1\xcf\x030\x1f\xac1\xba63p\xcb\x11\x0e\x9e\xf6\xce\xb8\xc6\x6\x91\x0c\xa12|]\xb50\x92\xe4)0\x1a\xa2\x89ifX\x020^\xb0\xcb.\x07\xf3\x14\xb7\x91\xa1/\xe6\xeaH\xa8\xae\xc8\xc8j1W\xeb:\xb8\x84PBV\x88\x1_z\x9eDw\xad\x17\xa0\x5e\x03\xel\x4\x02t\x00=0\x00\x9fU\xa9\x9cZ\x81\x07\x3\x9d\xab\xf8\x1d\xce\x1f\x6\xad\x05sA\\\xc0&\xcc\x96\x85/\x8b\x17q\xa1\xad\x0\x99\xa6\x9c\x15((\xda\x0\xae\x9e\x0b/p\x9e\xb0\x84#\xfc\x0hm\x5j?o\x3bw\x9b\xbaT\x1bDZ\xb3/[|\x83\x86Wr\x071\xb6\x01\xf9;p\x94n\xbb}\x89o\xa3\xce\xdb0HJ\x7f?*W\x5\x94j\x9a\x0e\xb5\xa9\x5\x83r\x6i\xdb\xf6\xe3\xfbq\xec\x4\xf7\x05\x6\xfd|\x1a\xab{\xf5\x8f\xfa\x8aX\t\x94\xa4\x06\x8cP\x05z(\xe3{5\x8ce\r\xcf/\x9a\x90`@\x1eK\xf2<\x95\x03\x07^\xb9\xf6 Q\xf5I\x0d\xb7U\x1e\xa2\xe0\x5\xf7\xe4s2\x01\x19\x08\xa6\x02Gh\x04\x92\xd0\x85\xac\x05\x12!C\$\x17`Q\x0d\x4\xdb1\xa7\x1b3\x867\x4r\x8c!N\x9f\x93\x91\x0d\x88%\x8c,9\x85\x97\8\x0d\x14IA[\xeb\x02A04\xfd1\x7t\x9f9\x934\x04\x11\xcePB\x2\x7\xa8Dp\x95\xee<\xa8\x90(\x81\x08:W\x894\x1b\xf9:wW\x00_\x02\x2\xea\x14\x1a\xee\xfa\tQ0n\xa8>\xc0uS\x164\$: \xc30\xe2.\xa8v\x86N\x93\xd4[w\xadn\xbcK\xa6\x6{\xbecr\x95\x1a\xe3\xb43v\x0ey_B\xda>5\x9cr\xfaf\xff\x06\xc5\xb7\xc3\xc7\xff!\xeb\x01\x89\x9a\xa6\x6\x1f0\x81\x9dB\xc3\xfe?7L\x9b>j\$2F\xc3\x13\xa7\xfa\x87i\x80\x14\xaf\xea0\xf2\x9a\x8b\xe6Hw\x06\x05oBs j8\xce5\xe6\\\xbfd4!\xd1Uu=\x05\x86\xea\xcen\xa8E!\x05I\xebH\xc2T\xc3\xe3\xc9i\x12=\xc6\x98FWH\x0\x6\xdb1\x93(0_(\xa8a\x13\xad\x0\x14hP_Q\xe5j\xba)\xe7\x9d\x7\x8u\xec\x05\xa2\x87\x9f\x9d9\xb7\xe6~\x85\x80^\x07\xa4\xef\xa9_\x7f9\xdf\xafc\x90\x18*Gu\x173m\x9e\x0e\x9f\xc3\x93\x14\xcf\xb4\xba\x513\x1d\x7\xebK\xe7\xbc\xfb\xa8I\xabK\xedQD\x89\x5V=\x14\x6\x5\x1ad\x96\x1f1\xbdJ\x6\x165t\x9cM#\xaa\xa8\xb6\xb5\xba\xab\xfb\x05T\x95\xad\xee\x7S\x11\x1e\x86\x4 \x99\x9\xdc\xaf?\x19\x8aRn\x8bS\xf9(\x86\x0d\xe4\0\x94\xdb*\xe4\xf5r<\xa3f\x9b"h\x1c1}Y\xcb\xad\xda\x4I\xa3\x14B\xd3\xb2\xbe\x0c\x1d\x86\x9eeq\xcd\x8bzn\x0f\x8b\x9d[S\x06\x2\x02~\xc4g\x03\x9a~\xb8\xd3\xb5\x003\x14\xcf\x98Y\ca\x002WD\x96K\x02\x00p\x19\xa1\x0cM\xa4\xfe\x4W\x16"m:E.\xbe\xc5\x1d+\x7f\x0e\x0d\x04\x96\x9c9v\x13Qf\$\x181\xe7\xfb806\x88\xd3\x8b\xa1\xa9e\x8c\xe9\xae|G\x0y\x0c0\xa2\x7D\xe3|\xeb\x0f\x8b\xe74\x5\xcc\x0c4\x04\xb3\xa7Q\x17B\x1e?`\x86`\xed\x91T!\xfcK\x8aw\xa4\xb9\xad\xbd\x9a`\xa7\x1b1~<&\x8fR=mu|\w\xa0P\x17\x10\xb9\xe8\x8c\xb1Ou\xe9\x82>\xbe\x5\x89\xa9/EC\x0\xbd>G\xba\x6\xe3D\x8f\xed\x9e\x4\xf3\xa86\xe9I\x02~V\xe6\x1e\x75\xda\x84\xaa\x99B\xc3=K\x8eD\x89\x15\xa6\x9bp\xe6\xb7H\x13\x8eF\xa1pr8+bW\x06B\xa3s\x0c\xa7Lh3z\xaf\x99c\x02\xde\xf6 Y\x0f0\x8e\xacF\x81\x1c1\x1b1\xf2)\xa2g\x9cr\x0d\x8b4<\xf8=>\x16@\xc1\x1b\x8f\x90k\xe4\x15\x82\x92\x07\x12/\xcff\xcc+\xb2XB?=\x8d.X\x1d\x16N\x06\x3qN\x06\x9\x0cq\x13\xc5\x12Mr\xe9\x18s.\xce\x3\xad\x93\xe8)NW"\x14\x12\x14\x0e\xbcq\xa0B\xa4L\xe55\x0ef<\xbc0\x86\x16Ue\x98\xfa\xdc*r\x2/\x9ak\xc4\x86\x17Wn," \x1cY\xe0\xe2\x8d\x5\x7f|\xfe \x85\x84/8\x1KG\x96\x7h\x4\x00\$\x15\x8d6\x1b\xff\xfb\xfe\xcd\x8ix{\xcd\xab\x1f\xa4?\x1807\x9b})h\xe4\x07n\xf3\x0d~\xbcs`u\xa8\xec\xce;\x87s\x1f1A\x2K\x93\xd2\xa9^\x1d\x1b\x90.T\xdaw~\x82\xff5x1\xef\x83\xe0\x1f\x0cn~^\xcb\xa4o'n', b'a\xf0?\x89\x8aT^A\x6X\xbd\x90d=\xd8\x07\x11G7\xe5\x00\x8d<\x80\x01\x88J\xf7\x6\xec\x16\xf2\x02\x8eg2\x11q1\xc5\x12\xe4|\n', b'\x1d\xeb\x0\x01\xb9\x14\x17\x4,\xad\x130Y\x80\x80\xa0\x9b\xbd\x8096\n', b'X\xfa0\x95\xbe\x00\\|(\x92\xe6\x0beS"\x00\xe4\xee\x1\xd3\xf5\$N\\ \xbfe\xa92\xb5ED\x1|ry\xf6\x10\xcd|r@\xa1\xaa(\xde?D\xf2 =\x90\xf3\xe8"\xb0b*\x91QF\xc8{4\xa5\x86PG\x05\xcf\xa5\x135\x90Dh\xa1\xa8(\xfBg\x97r\x020\x0d0\x7'\x8d\x3K\x9e{P\xe8\x0d?\x14\x6\x88\x1c\x11K\xfc!#\x99\xd2\xaf\x18=\xf0\xa58\xf7]-\xb7"\xc7\xe8\x8e\x7d\x9e\x13\xba/is64>\xb9\n']

Ecriture dans un fichier

On peut écrire dans un fichier si celui-ci est ouvert en écriture. Les trois modes possibles sont "w", "x" et "a". À l'instar de read() et readlines(), on utilisera write(chaine) pour écrire une chaîne de caractères et writelines(lignes) avec lignes une liste ou un tuple dont chaque élément est une ligne à écrire.

```
In [4]: fichier = open("IUT_de_Vélizy.txt", 'at')
fichier.write(" \nLe meilleur département c'est incontestablement R&T ")
fichier.close()
fichier = open("IUT_de_Vélizy.txt", 'rt')
texte = fichier.read()
print(texte)
```

L'Institut universitaire de technologie de Vélizy (IUT) est une composante de l'université de Versailles-Saint-Quentin-en-Yvelines dans les Yvelines.

Il propose des formations professionnelles en deux ans accessibles après le baccalauréat, les DUT (diplôme universitaire de technologie) et des formations en un an accessibles après un bac+2, les licences professionnelles.

Ces formations sont dispensées dans les domaines des technologies industrielles et tertiaires (industrie, commerce, services).

Le meilleur département c'est incontestablement R&T

```
In [ ]:
```