

## Séance 8 : ENSEMBLES, DICTIONNAIRES ET FICHIERS TEXTE

L1 – Université Côte d'Azur

Récupérez le fichier `dictionnaire.txt` sur la page du cours (en faisant `enregistrez-sous` et non copier-coller).

<https://upinfo.univ-cotedazur.fr/~obaldellon/L1/bi1/tp8/dictionnaire.txt>

puis placez le dans votre répertoire de travail, il servira pour tout le TP. Ce fichier contient les mots du dictionnaire à raison d'un par ligne.

### Exercice 1 — Ensembles et dictionnaires

Dans cet exercice, `chemin` et `chemin_bilan` sont des chaînes de caractères indiquant le nom (ou le chemin) d'un fichier texte dans le répertoire courant. Typiquement, on pourra choisir "`dictionnaire.txt`" comme exemple pour `chemin`.

- Écrire une fonction `ensemble_de_lettres(chemin)` qui renvoie l'ensemble des lettres contenu dans le fichier donnée en paramètre. On ne comptera pas les ponctuations comme étant des lettres Combien y a-t-il de lettres distinctes dans le fichier `dictionnaire.txt`?
- Écrire une fonction `statistique_lettre(chemin)` qui renvoie le dictionnaire indiquant pour chaque lettre (les clés) son nombre d'occurrences.

```
1 >>> dico = statistique_lettre("toto.txt")
2 >>> dico['e'] # la lettre « e » apparaît 123 fois dans le fichiers toto.txt
3 123
```

- Écrire une fonction `bilan(chemin, chemin_bilan)` qui écrit dans le fichier dont le nom est fourni par la chaîne `chemin_bilan` les deux phrases suivantes où on aura remplacé les « ... » par les bonnes valeurs  
La lettre la plus présente dans le fichier ... est ...  
La seconde lettre la plus présente dans le fichier ... est ...

### Exercice 2 — Utilisation du dictionnaire

À partir de cet exercice, on appellera `dico` la liste (de type `list` et non `dict`) de tous les mots du dictionnaire.

- Écrivez une fonction `extraire_dictionnaire()` qui lit le fichier `dictionnaire.txt` et renvoie une liste dont les éléments sont les mots du dictionnaire. Sauvegarder le résultat de cette fonction dans une variable `dico`.

```
1 def extraire_dictionnaire():
2     s="dictionnaire.txt"
3     f_in = open(s, 'r', encoding = 'utf-8')
4     L=[]
5     for ligne in f_in:
6         mot=ligne[0:-1] # le dernier symbole correspond au symbole '\n'
7         L.append(mot)
8     return L
9
10 dico=extraire_dictionnaire()
```

- Faites afficher les vingt premiers mots de `dico`, un par ligne, en donnant leur indice dans `dico` et leur longueur.

```

1 0 abaissable 10
2 1 abaissante 10
3 2 abaissée 8
4 ...
5 19 abarticulation 14

```

```

1 for i in range(20) :
2     print( i , dico[i] , len(dico[i]) )

```

3. Calculez **sans l'afficher** la liste L des longueurs des mots du dictionnaire. Affichez L[:3], vous devrez trouver [10,10,8]

*Première méthode en bouclant sur les indices.*

```

1 n=len(dico)
2 L=[]
3 for i in range(n) :
4     L.append(len(dico[i]))

```

*Seconde méthode en bouclant directement sur les mots.*

```

1 n=len(dico)
2 L=[]
3 for mot in dico :
4     L.append(len(mot))

```

*Troisième méthode avec une compréhension.*

```

1 [ len(mot) for mot in dico ]

```

4. À l'aide d'une boucle **for**, calculez **longueur\_max(dict)** la longueur du plus long mot de **dict**. Vérifiez en comparant **longueur\_max(dico)** à **max(L)**.

```

1 def longueur_max(dico):
2     l_max=0
3     for i in range(len(dico)):
4         if len(dico[i])>l_max:
5             l_max=len(dico[i])
6     return l_max

```

*Avec une boucle for plus pythonesque.*

```

1 def longueur_max(dico):
2     l_max=0
3     for mot in dico:
4         if len(mot)>l_max:
5             l_max=len(mot)
6     return l_max

```

5. Affichez le ou les mots les plus longs du dictionnaire. Vous devriez en trouver trois.

```

1 l_max=longueur_max(dico)
2 for m in dico :
3     if len(m) == l_max :
4         print(m)

```

*anticonstitutionnellement désoxyribonucléoprotéique oligoasthénotératospermie*

**Exercice 3** — Le jeu du pendu

Le jeu du pendu est un jeu entre deux joueurs, appelés ci-dessous *Alice* et *Bob*. Au début, Alice choisit un mot secret *m* et le publie en ne divulguant que sa première et dernière lettre, les autres étant remplacées par la symbole « \_ ». À chaque tour, Bob propose une lettre : si elle fait partie du mot, Alice complète l'affichage en l'ajoutant autant de fois qu'elle apparaît, sinon Bob perd une vie. L'objectif est pour Bob de deviner le mot sans perdre toute ses vies.

**Exemple de partie** ou Alice est joué par l'ordinateur et perd face à Bob (l'humain)

```

1 il te reste 5 vies
2 e___t
3
4 Quelle lettre demandes-tu, humain ? e
5 Bien joué !
6 e__et
7
8 Quelle lettre demandes-tu, humain ? m
9 Raté... Il te reste 4 vies
10 e__et
11
12 Quelle lettre demandes-tu, humain ? f
13 Bien joué !
14 effet
15 GAGNÉ ! On rejoue ?

```

Dans votre répertoire, créez un fichier *pendu.py*. Au début du fichier, commencez par créer la liste *dico* à partir du fichier *dictionnaire.txt* comme dans l'exercice 2.

- Écrivez une fonction *mot\_au\_hasard()* sans arguments qui renvoie un mot tiré au hasard. Vous devrez utiliser la variable globale *dico* et la fonction *randint* du module *random* (si vous ne vous souvenez plus comment fonctionne *randint*, faites *help(randint)*). Testez votre fonction en affichant dix mots tirés au hasard

```

1 def mot_au_hasard():
2     # dico est une variable globale
3     n=len(dico)
4     aléa=randint(0,n-1)
5     return dico[aléa]

```

*Si on veut être sûr de ne jamais tirer de mots de deux ou une lettre, on peut rajouter un test et un appel récursif.*

```

1 def mot_au_hasard():
2     # dico est une variable globale
3     n=len(dico)
4     aléa=randint(0,n-1)
5     if len(dico[aléa]) >=3:
6         return dico[aléa]
7     else:
8         return mot_au_hasard()

```

- Écrivez une fonction *cache\_lettre(mot)*, qui renvoie une chaîne de caractère obtenue en remplaçant dans *mot* toutes les lettres (sauf la première et la dernière) par le symbole : \_. Par exemple *cache\_lettre('inspiration')* renverra 'insporaton'.

```

1 def cache_lettre(mot):
2     n=len(mot)
3     assert(n>=3)
4     res=mot[0]
5     for i in range(1,n-1):
6         res = res + '_'
7     return res + mot[n-1]

```

3. Écrivez une fonction `ajoute_lettre(mot,secret,lettre)` qui prends trois arguments :

- `secret` est un mot (une chaîne de caractère)
- `mot` est une version incomplète de la chaîne `secret`
- `lettre` est une des 26 lettres de l'alphabet.

Si `lettre` est un caractère de `secret`, la fonction `ajoute_lettre` doit renvoyer une version complétée de `mot` auquel on a ajouté la lettre « `lettre` », sinon, on renvoie la chaîne `mot` sans modification. Dans cette question, on suppose que les mots ne contiennent pas d'accent.

```

1 >>> ajoute_lettre('e_s__g__m__t','enseignement','e')
2   'e_se_g_eme_t'
3 >>> ajoute_lettre('e_s__g__m__t','enseignement','z')
4   'e_s__g__m__t'
```

```

1 def ajoute_lettre(mot,secret,lettre):
2     n=len(mot)
3     assert(n==len(secret))
4     res=''''
5     for i in range(n):
6         if secret[i]==lettre:
7             res = res+secret[i]
8         else:
9             res = res+mot[i]
10    return res
```

4. Écrire une fonction `pluriel` qui prend un nombre et un mot et qui renvoie la chaîne correspondante en accordant au pluriel si nécessaire.

```

1 >>> (pluriel(3,"chien") , pluriel(1,"chat"))
2 ('3 chiens', '1 chat')
```

```

1 def pluriel(k,mot):
2     if k==0 or k==1:
3         return str(k) + ' ' + mot
4     else:
5         return str(k) + ' ' + mot + 's'
```

5. Il est temps de voir comment nous allons utiliser ces fonctions. La fonction principale sera la suivante :

```

1 def nouvelle_partie(vies) :
2     secret = mot_au_hasard() # le mot à deviner
3     mot = cache_lettre(secret) # le mot partiellement découvert
4     print('il te reste' , vies , 'vies')
5     while vies > 0 and '_' in mot :
6         print(mot)
7         c = input('Quelle lettre demandes-tu, humain ? ')
8         ancien_mot=mot
9         mot=ajoute_lettre(mot,secret,c)
10        if mot != ancien_mot :
11            print('Bien joué !')
12        else :
13            vies = vies-1
14            print('Raté.. Il te reste ' + pluriel(vies,'vie')+'.')
15        if vies == 0 :
16            print('PENDU! Le mot secret était' , secret + '. On rejoue ?')
17        else :
18            print(mot)
19            print('GAGNÉ ! On rejoue ?')
```

Copier la et testez la.

6. On va maintenant gérer correctement les caractères accentués en utilisant une fonction.

```
1 from unicodedata import normalize
2 def sans_accents(s):
3     return normalize('NFKD', s).encode('ascii', 'ignore').decode('ascii')
```

Cette fonction permet de renvoyer une chaîne obtenue en ôtant les cédilles et les accents en les remplaçant par leur version ASCII. Modifiez votre fonction `ajoute_lettre` pour qu'elle gère correctement les accents.

```
1 >>> ajoute_lettre('d___f___nç___nt', 'déréférencassent', 'é')
2 'dé_efé_enç__ent'
```

```
1 def ajoute_lettre(mot, secret, lettre):
2     n=len(mot)
3     assert(n==len(secret))
4     res=""
5     for i in range(n):
6         if sans_accents(secret[i])==lettre:
7             res = res+secret[i]
8         else:
9             res = res+mot[i]
10    return res
```

#### Exercice 4 — Record battu!

Modifier votre programme précédent pour qu'il ajoute votre score (nombre de vies restantes) dans un fichier `score.txt`. Le fichier devra se présenter sous cette forme :

```
1 Vies restantes (meilleur score)
2 =====
3 6 : Olivier le 19/02/2021 à 12:30:1
4 5 : Sandrine le 23/03/2021 à 11:11:11
5 1 : Pierre le 15/02/2021 à 09:03:51
```

Lorsque le joueur gagnera on lui demandera son nom, on lira le fichier `score.txt`, on stockera les anciens résultats dans une liste que l'on mettra à jour avec le nouveau score puis on réécritra le tout dans le fichier initial. Si le fichier n'existe pas, il faudra le créer. On fera attention à ce que les scores soient toujours classés par ordre décroissant. On pourra s'aider de la fonction suivante pour afficher la date.

```
1 from time import localtime
2 def date():
3     t=localtime()
4     heure=f"{t.tm_hour:0>2}:{t.tm_min:0>2}:{t.tm_sec:0>2}"
5     jour =f"{t.tm_mday:0>2}/{t.tm_mon:0>2}/{t.tm_year:0>2}"
6     return 'le ' + jour + ' à ' + heure
```

*La fonction `découpe` transforme une chaîne de la forme '6 : Olivier le 19/02/2021 à 12:30:1' en un couple contenant un entier et une chaîne : (6, 'Olivier le 19/02/2021 à 12:30:1')*

```
1 def découpe(ligne):
2     k = ligne.index(" : ")
3     return (int(ligne[:k]),ligne[k+3:])
```

*La fonction `ajouter` permet d'insérer notre couple au bonne endroit parmis une liste de couple déjà trié. On commence à ajouter les scores qui sont supérieurs, puis notre nouveau score puis les scores inférieurs. La liste obtenue est donc toujours triée.*

```

1 def ajoute(nouvelle_ligne, liste):
2     (score, chaîne)=nouvelle_ligne
3     rés=[]
4     i=0
5     # On ajoute les scores supérieurs
6     while i<len(liste) and liste[i][0]>score:
7         rés.append(liste[i])
8         i=i+1
9     # On ajoute le nouveau score
10    rés.append(nouvelle_ligne)
11    # On ajoute les scores inférieurs
12    while i<len(liste):
13        rés.append(liste[i])
14        i=i+1
15    return rés

```

*La fonction pour afficher les meilleurs scores (en supposant que le fichier existe déjà)*

```

1 def afficher_score():
2     assert(os.path.isfile('score.txt'))
3     f_in = open('score.txt', 'r', encoding = 'utf-8')
4     contenu=f_in.read()
5     f_in.close()
6     print(contenu)

1 import os.path
2
3 def ajout_score(vies,nom):
4     if os.path.isfile('score.txt'):# On vérifie si le fichier existe déjà.
5         f_in = open('score.txt', 'r', encoding = 'utf-8')
6         liste=f_in.readlines()
7         f_in.close()
8         liste=liste[2:] # on supprime les deux première ligne
9         liste= [ découpe(ligne) for ligne in liste ]
10    else:
11        liste=[]
12    nouvelle_ligne = (vies, nom+' '+date()+'\n')
13    liste = ajoute(nouvelle_ligne,liste)
14    f_out = open('score.txt', 'w', encoding = 'utf-8')
15    f_out.write('Vies restantes (meilleur score)\n')
16    f_out.write('=====+\n')
17    for couple in liste:
18        (v,c)=couple
19        f_out.write(str(v) + ' : ' + c)
20    f_out.close()

```

### Exercice 5 — L'ordinateur sauve sa peau

On veut maintenant échanger les rôles, et faire chercher le mot secret à l'ordinateur. On commence par implémenter une première stratégie utilisant du hasard, puis on étudie une stratégie plus efficace.

- Écrivez une fonction `candidats(début,fin,longueur)` qui prend en argument deux caractères `début` et `fin` et un entier `longueur` et qui renvoie la liste de tous les mots du dictionnaire qui commencent par `début`, finissent par `fin`, et sont de longueur `longueur`. Par exemple :

```

1 >>> candidat('h','s',6)
2 ['habeas', 'hachis', 'haggis', 'hermès', 'herpès', 'hiatus', 'hormis', 'hybris']

```

*Version classique*

```

1 def candidats(deb,fin,l) :
2     L=[]
3     for m in dico:
4         if len(m) == 1 and m[0] == deb and m[-1] == fin:
5             L.append(m)
6     return L

```

*Version plus courte*

```

1 def candidats(deb,fin,l) :
2     return [ m for m in dico if len(m) == 1 and m[0] == deb and m[-1] == fin]

```

2. Écrivez une fonction `choix_lettre`(mot,possibilités) qui prend en argument un mot partiellement révélé et une liste de mots possibilités et qui renvoie au hasard une lettre qui apparaît dans au moins un mot de possibilités à une position où on n'a encore rien révélé. Par exemple, si `mot = 'h___is'` et `possibilités = ['hachis', 'haggis', 'hormis', 'hybris']`, un appel à `choix_lettre(mot,possibilités)` renverra une lettre au hasard parmi achaggormybr.

```

1 def choix_lettre(s,L):
2     lettres = []
3     l = len(s)
4     for m in L :
5         assert(len(m) == 1)
6         for i in range(l) :
7             if s[i] == '_' :
8                 lettres.append(sans_accents(m[i]))
9     x = randint(0, len(lettres) - 1)
10    return lettres[x]

```

3. Écrivez une fonction `filtre_lettre`(lettre,possibilités) qui prend en argument un caractère lettre et une liste de mots possibilités et qui renvoie la sous-liste des mots qui ne contiennent pas lettre, sauf éventuellement comme première ou dernière lettre. Par exemple,

```

1 >>> filtre_lettre('e',['avion','état','route','été','enquête'])
2 ['avion', 'état', 'route', 'été']

```

```

1 def filtre_lettre(c,L) :
2     F=[]
3     for m in L:
4         if c not in sans_accents(m[1:-1]):
5             F.append(m)
6     return F
7
8 # Version plus courte
9 def filtre_lettre(c,L) :
10    return [m for m in L if c not in sans_accents(m[1:-1])]

```

4. Écrivez une fonction `est_compatible`(mot,secret) qui prend en arguments deux chaînes de caractères correspondant à un mot incomplet (mot) et un mot complet secret, et qui renvoie `True` si mot peut se compléter en secret. Par exemple,

```

1 >>> est_compatible('h___is','hachis')
2 True
3 >>> est_compatible('h___is','hermès')
4 False
5 >>> est_compatible('he___s','hermès')
6 False

```

```

1 def est_compatible(s,m) :
2     if len(s) != len(m) :
3         return False
4     for i in range(len(s)) :
5         if s[i] != '_' and s[i] != m[i] :
6             return False
7         if s[i] == '_' and m[i] != '_' and m[i] in s[1:-1] :
8             return False
9     return True

```

5. Écrivez une fonction `joue_chercheur(vies)` qui démarre une partie où vous devez faire deviner un mot à l'ordinateur. Ce dernier donnera des informations sur l'état d'avancement de ses réflexions. On aura par exemple

```

1 Quel est ton indice de départ, humain ? h___s
2 J'hésite entre : habeas hachis haggis hermès herpès hiatus hormis hybris
3 Je demande le a. Nouvel indice ? h___s
4 Il me reste 4 vies.
5 J'hésite entre : hermès herpès hormis hybris
6 Je demande le i. Nouvel indice ? h___is
7 J'hésite entre : hormis hybris
8 Je demande le y. Nouvel indice ? hy__is
9 J'ai trouvé : hybris.

```

```

1 def joue_chercheur(vies) :
2     s = input('Quel est ton indice, humain? ')
3     L = candidats(s[0],s[-1],len(s))
4
5     while vies > 0 and len(L) > 1 :
6
7         print("J'hésite entre" , L)
8         lettre = choix_lettre(s,L)
9         print('Je demande le ' , lettre + '.')
10
11        while True :
12            s2 = input('Nouvel indice? ')
13            if est_compatible(s,s2) : break
14            print('erreur de saisie. Ancien indice :',s)
15
16        if s2 == s :
17            # rate
18            L = filtre_lettre(lettre,L)
19            vies = vies - 1
20            print('Il me reste' , vies , 'vie' + ('s' if vies > 2 else '') + '.')
21        else :
22            # bonne lettre
23            s = s2
24            L = [m for m in L if est_compatible(s,m)]
25
26        if len(L) == 1 :
27            print("J'ai trouvé :" , L[0])
28        elif len(L) == 0 :
29            print("Tricheur!")
30        else :
31            print("J'ai perdu, tu avais bien choisi ton mot...")

```

6. Améliorez la fonction `choix_lettre(s,L)` pour accroître les chances de gagner de l'ordinateur.

*Indication* Une possibilité est de s'inspirer de la dichotomie : on peut penser que le choix de la lettre c sera un bon choix si il permet de garder autant de candidats qu'il permet d'en éliminer, autrement dit si le nombre de mots de

possibilité qui contiennent c est le plus proche possible de la moitié de la longueur de possibilité. Ainsi quelque soit la réponse, on éliminera environ la moitié des possibilités.

Par exemple, si `mot = 'h_____s'` et `L = ['habeas', 'hachis', 'haggis', 'hermès', 'herpès', 'hiatus', 'hormis', 'hybris']`, le choix de '`i`' permet de garder 5 candidats et d'en éliminer 3, tandis que le choix de '`y`' permet de garder 1 candidat et d'en éliminer 7; le choix de `i` est donc meilleur que le choix de `y`. Mais le choix optimal, sur cet exemple, est '`a`', car il permet de garder 4 candidats et d'en éliminer tout autant.

Vous pouvez cependant réfléchir à d'autres stratégies (maximiser les chances de réduire le nombre de lettres à deviner, stratégie « petit joueur » pour prendre le moins de risque possible, etc), et comparer expérimentalement l'efficacité de ces stratégies.