

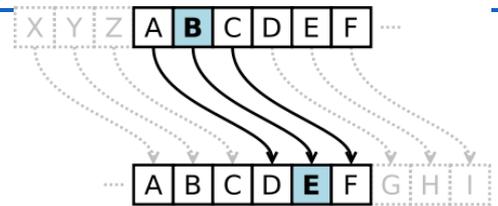
Chiffrage - Déchiffrage - Analyse de textes

Traitement des textes en python

Voir le document apprendre_python3.pdf de
Gérard Swinnen



Chaîne de caractères
ch5 p.41
Listes
ch5 p.45 et ch10 p.146
Dictionnaires
ch10 p.160



1 Les chaînes de caractères

Saisie d'une chaîne

```
monTexte="Bonjour et bienvenue, bonne et et journée en ICN"
```

La variable monTexte est de type string. C'est une suite de caractères. Chaque caractère peut être éventuellement codé sur plusieurs octets selon la norme Unicode employée voir ch10.

Dans une chaîne de caractère la séquence \n provoque un passage à la ligne.

Les délimiteurs d'une chaîne sont : une quote simple ' , double " et triple ""

On peut accéder à un caractère particulier, à un tronçon :

On peut concaténer deux chaînes avec le

+

On peut connaître le nombre de caractères avec

len(monTexte)

On peut convertir une chaîne en entier

int(monTexteNumerique)

On peut convertir une chaîne en réel (float)

float(monTexteNumerique)

On peut convertir une chaîne complète en MAJUSCULES : monTexte.upper()

ou en minuscules : monTexte.lower()

```
In [5]: monTexte="Bonjour le monde"
```

```
In [6]: monTexte[3]
Out[6]: 'j'
```

```
In [7]: monTexte[7:12]
Out[7]: ' le m'
```

```
In [8]: monTexte[:14]
Out[8]: 'Bonjour le mon'
```

```
In [9]: monTexte[12:]
Out[9]: 'onde'
```

2 Les listes

Une liste est une collection d'éléments séparés par une virgule et entre crochets. On peut également accéder à un élément quelconque. Voir ici :

```
>>> maliste = [0, "un", 2] # on définit une liste pour faire des tests
>>> len(maliste)          # on demande sa longueur (length en anglais)
3
>>> maliste[0]           # on demande le premier élément (rang ou index 0)
0
>>> maliste[1]           # le deuxième (index 1)
'un'
>>> "un" in maliste      # on demande si le texte "un" est un élément de la liste
True
>>> "deux" in maliste    # on demande si le texte "deux" est un élément de la liste
False
>>> maliste.index("un")  # on demande l'index (rang) de l'élément "un"
1
>>> maliste.count("un") # on demande le nombre d'apparitions de l'élément "un"
1
```

3 Les dictionnaires¹

Les dictionnaires sont des types composites qui utilisent des clés pour accéder aux différents éléments.

Création :

```
>>> a = {}
```

Tester la présence d'une valeur :

```
>>> a.has_key("nom")
True
```

Ajout de valeurs :

```
>>> a = {}
>>> a["nom"] = "engel"
>>> a["prenom"] = "olivier"
>>> a
{'nom': 'engel', 'prenom': 'olivier'}
```

Récupérer les clés avec une boucle :

```
>>> fiche = {"nom": "engel", "prenom": "olivier"}
>>> for cle in fiche.keys():
...     print cle
...
nom
prenom
```

Retrouver une valeur :

```
>>> data = {"name": "Olivier", "age": 30}
>>> data.get("name")
'Olivier'
>>> data.get("adresse", "Adresse inconnue")
'Adresse inconnue'
```

Récupérer les valeurs avec une boucle :

```
>>> fiche = {"nom": "engel", "prenom": "olivier"}
>>> for valeur in fiche.values():
...     print valeur
...
engel
olivier
```

¹ Voir ici <http://apprendre-python.com/page-apprendre-dictionnaire-python>

Récupérer les clés et les valeurs par une boucle :

```
>>> fiche = {"nom":"engel","prenom":"olivier"}
>>> for cle,valeur in fiche.items():
...     print cle, valeur
...
nom engel
prenom olivier
```

4 Analyse détaillée d'un code

```
1  #!/usr/bin/env python
2  # -*- coding: utf-8 -*-
3
4  from os import chdir
5  path = r'G:\GPRO\USB\ICN2nd\Preparation ICN ISN 2017 2018\Cryptographi
6  chdir(path)
7
8  # Saisie d'une chaine
9  monTexte="Bonjour et bienvenue, bonne et et journée en ICN"
10
11
12 # Travail sur les chaines
13 nombreDeCaractere = len(monTexte)
14 print("Nombre de caractères : ",nombreDeCaractere)
15
16
17 # Sélection d'un seul caractère
18 numero = input('Entrez votre choix : ')
19 numero = int(numero) - 1
20 while numero > nombreDeCaractere:
21     print('Entrez une valeur comprise entre 1 et ',nombreDeCaractere)
22     numero = input('Entrez votre choix : ')
23     numero = int(numero) - 1
24
25 print(monTexte)
26 for i in range(numero):
27     print(' ',end='')
28     print('^')
29
30 print(end='\n')
```

```
Nombre de caractères : 49
Entrez votre choix : 8
Bonjour et bienvenue, bonne et et journée en ICN
|
^
```

```

32 # Travail avec une liste
33 listeDesMots = monTexte.split()
34 print('La liste des mots du texte : ',listeDesMots)
35
36 motChoisit = input('Entrer le mot à rechercher : ')
37 nombreOccurence = monTexte.count(motChoisit)
38 print("Nombre d'occurence du mot = ",nombreOccurence,'\n')
39
40 # Création d'un dictionnaire avec les mots du texte
41 dicoPython = {}
42 for mot in listeDesMots:
43     if mot in dicoPython:
44         dicoPython[mot]+=1
45     else:
46         dicoPython[mot]=1
47 print('Le dictionnaire')
48 print(dicoPython)
49 print()
50 print('Liste des valeurs')
51 for cle in dicoPython.keys():
52     print("Mot ",cle," : ",dicoPython[cle])

```

La liste des mots du texte : ['Bonjour', 'et', 'bienvenue,', 'bonne', 'et', 'et', 'journée', 'en', 'ICN']
Entrer le mot à rechercher : et
Nombre d'occurence du mot = 3

Le dictionnaire
{'en': 1, 'journée': 1, 'bienvenue,': 1, 'Bonjour': 1, 'ICN': 1, 'bonne': 1, 'et': 3}

Liste des valeurs
Mot en : 1
Mot journée : 1
Mot bienvenue, : 1
Mot Bonjour : 1
Mot ICN : 1
Mot bonne : 1
Mot et : 3

5 Exercices à faire à la maison (Source Swinnen)

- 5.7 Écrivez un script qui compte le nombre d'occurrences du caractère « e » dans une chaîne.
- 5.15 Écrivez un programme qui analyse un par un tous les éléments d'une liste de mots (par exemple: ['Jean', 'Maximilien', 'Brigitte', 'Sonia', 'Jean-Pierre', 'Sandra']) pour générer deux nouvelles listes. L'une contiendra les mots comportant moins de 6 caractères, l'autre les mots comportant 6 caractères ou davantage.