

NSI

03



REPRÉSENTATION D'UN TEXTE

[Stéphane BEAUDET – Frédéric PEURIERE]
D'après l'ouvrage de Brahim Dahlab et Cedric Bony

Connaître différentes méthodes d'encodage des caractères

Activité 1: À la découverte de la table « à ski »...avec M. Padchance !

Regarde la vidéo de M. Padchance #1

<https://drive.google.com/open?id=1XTVIn4e6TUtj5SXtkOh3jma9Vhm6NRSv>

1. Aide M Padchance, en suivant les instructions notées sur les touches manquantes, que se passe-t-il?....

(DANS UN EDITEUR DE TEXTE, comme Notepad++ à télécharger sur

<https://notepad-plus-plus.org/downloads/v7.8.2/>

FIXE : maintiens le bouton Alt et sans relâcher compose les numéros avec le clavier numérique, puis relâche
PORTABLE : selon les modèles, utilises le pavé numérique, qui est souvent mis en bleu sur les touches {è_ç}, {uio} et {jkl}. Activer (verr. num) Appuyer simultanément sur (alt) + (code ASCII) OU appuie simultanément sur (fn) + (alt) + (code ASCII)



Code	Il apparaît a l'écran
Alt + 065	
Alt + 066	
Alt + 067	

2. À ton avis, quel code dois-je saisir pour avoir Z ? Complète le tableau pour répondre.

Tester maintenant dans la console de Python les instructions suivantes

```
>>> ord('A') >>> ord('a') >>> ord('U') >>> ord('B') >>>ord(a) >>>ord(67)
```

```
>>> chr(65) >>> chr(90) >>>chr(123) >>> chr('A') >>> chr(A)
```

3. Fort de cette découverte, réalise un programme python qui permet de coder ce message « PADCHANCE PLEURE » en créant un décalage de lettres : par exemple, « BAC » deviendrait « CBD » avec un décalage. Pour info, on appelle ce codage le code de César.....

(pour vérifier : PADCHANCE PLEURE devient ainsi avec un décalage de 5 : UFIHMFESHJ%UQJZWJ)

Regarde la video de M. Padchance #2

<https://drive.google.com/open?id=1pxSotVQdwWd0-o0Ws8YhJGRBM-qazWnb>

4. Au regard du document « **Ressource01 - Table ASCII** » qui se prononce « *aski* », situé sur le site, explique les raisons de sa mésaventure, puis à ton tour fais évoluer ton programme pour qu'il permette de transformer des majuscules en minuscules et inversement au choix de l'utilisateur. Pour info, cela revient à recréer la méthode upper et lower de Python.

Regarde la vidéo de M. Padchance #3

https://drive.google.com/open?id=1nBU1rP3tbQBsP2zG5_8rbsT4zft4cig1

5. Explique ce qui s'est passé, à l'aide de la table ASCII et du document « **Ressource02 Les caractères de contrôle** ».

Activité 2: Le ski, ce n'est qu'en hiver !

1. À l'aide de la table ASCII, code cette phrase « Un âne éternue » en langage binaire et décimal, que remarques-tu ? Conclue sur l'utilisation de la table ASCII pour coder du texte en français ou en allemand.

2. Au final, combien de caractères peut-on encoder grâce à la table ASCII ? En déduire le nombre de bits utilisés par cette méthode d'encodage.

3. D'autres encodages sont progressivement apparus pour répondre à l'expansion d'internet à travers le monde et donc aux échanges de documents entre différents pays et donc différents caractères. On trouve par exemple, l'encodage ISO 8859-1 ainsi que l'ISO 8859-15 mais également l'UTF-8, chacun apportant des évolutions.

Sur Notepad++ , ouvre le document *phrase*

Change l'encodage comme indiqué dans le tableau ci dessous avec l'aide du document « **Ressource03 – Notepad ++** » et réponds aux questions posées :

ENCODE en	Question
ANSI	Indique si avec cet encodage le « é » s'affiche correctement Y a t il d'autres caractères erronés ?
8859-15	Quels sont les changements ?
Convertir en UTF8	Quels sont les changements ?
Ouvre http:// xinhuanet.com (Ctrl U pour le code source)	Quel encodage est choisi pour afficher les caractères de cette page ?
Ouvre http:// russian.news.cn (Ctrl U pour le code source)	Quel encodage est choisi pour afficher les caractères de cette page ? Que peux-tu en conclure ?

4. Ces encodages diffèrent de l'ASCII en ajoutant (au moins) un bit supplémentaire. Combien le fait d'avoir 8 bits amène-t-il de nouvelles possibilités ?

5. Quelques précisions sur l'UTF-8

L'encodage UTF-8 utilise 1, 2, 3 ou 4 octets en respectant certaines règles :

*Un texte en ASCII de base (appelé aussi US-ASCII) est codé de manière identique en UTF-8. On utilise un octet commençant par 0.

Caractère	Point de code (hexadécimal)	Valeur scalaire		Codage UTF-8
		décimal	binaire	binaire
A	U+0041	65	1000001	01000001

*Les octets ne sont pas remplis entièrement.

* Les bits de poids fort du premier octet forment une suite de 1 indiquant le nombre d'octets utilisés pour coder le caractère. Les octets suivants commencent tous par 10.

Définition du nombre d'octets utilisés

Représentation binaire UTF-8	Signification
<i>0xxxxxxx</i>	1 octet codant 1 à 7 bits
<i>110xxxxx 10xxxxxx</i>	2 octets codant 8 à 11 bits
<i>1110xxxx 10xxxxxx 10xxxxxx</i>	3 octets codant 12 à 16 bits
<i>11110xxx 10xxxxxx 10xxxxxx 10xxxxxx</i>	4 octets codant 17 à 21 bits

* Dans la norme ISO 8859-1 le « é » est codé 1110 1001, en UTF-8 on le codera **11000011 10101001**. On pourra remarquer deux choses : le codage ISO s'inscrit dans le codage UTF-8, on a comblé le premier octet avec des zéros (en italique).

Activité 3 : Les impacts de l'encodage...

1. Ouvre python, puis trouve astucieusement le code qui va te permettre de découvrir quel est son encodage par défaut. Ecris-le ci-dessous :

2. Vérifie en saisissant :

```
import sys
print(sys.getdefaultencoding())
```

3. Dans les deux extraits de code source de pages web ci-dessous, identifie, en les encerclant, dans quel encodage chacune s'affiche, et grâce à quel balise/attribut

```
2 <!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.0 Transitional//EN"
3
4 <html xmlns="http://www.w3.org/1999/xhtml">
5
6 <head>
7 <meta http-equiv="Content-Type" content="text/html; utf-8" />
8 <title>Home</title>
9 <link rel="stylesheet" type="text/css" href="styl1.css" />
```

```
<!DOCTYPE html>
  ▲ <html>
    ▲ <head>
      <meta http-equiv="Content-Type" content="text/html;charset=iso-8859-15" />
      <title>iso-8859-15 Encoded Page</title>
    </head>
```

Regarde la vidéo: M. Padchance #4

<https://drive.google.com/open?id=1OuwJqFu3vf2LruGfNanqrvQEs6S-7lu8>

4. Es-tu surpris du résultat ? Explique ce qui s'est passé

5. Soit deux listes d'éléments :

Liste1 = ['a', 'Z', 'Zoo', '4', 'fin'] et

Liste2 = ['34', 'motlongenminuscule', 'MOTTRESPETIT', '?', '0', 'motenminuscule']

En t'aidant du document ressource approprié, **écris tes pronostics** sur le résultat du tri de chacune de ces listes.

Puis à l'aide d'un script Python, trie les éléments de ces listes, comme M. Padchance, et vérifie tes hypothèses. *Indice : utilise la méthode `sort()` de la manière suivante : `liste.sort()`*