

# Découvrir et approfondir python



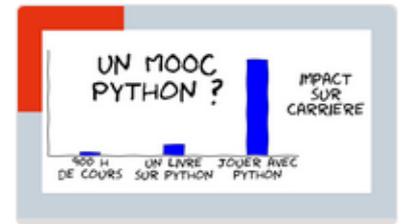
## Codage, décodage, ASCII, UTF-8, chaîne de caractères :

-  V W2-01 Codage jeux de caractères et Unicode.mp4
-  V W2-02 Les chaînes de caractères.mp4

Vidéos proposées par l'INRIA sur



Python 3 : des fondamentaux aux concepts avancés du langage



Thierry PARMENTELAT et Arnaud Legoud



## Pour bien comprendre et approfondir :

### Le jeu de caractères ASCII

American Standard Code for Information Interchange

MSB \ LSB	0	1	2	3	4	5	6	7	
	000	001	010	011	100	101	110	111	
0	0000	NUL	DLE	SP	0	@	P	`	p
1	0001	SOH	DC1	!	1	A	Q	a	q
2	0010	STX	DC2	"	2	B	R	b	r
3	0011	ETX	DC3	#	3	C	S	c	s
4	0100	EOT	DC4	\$	4	D	T	d	t
5	0101	ENQ	NAK	%	5	E	U	e	u
6	0110	ACK	SYN	&	6	F	V	f	v
7	0111	BEL	ETB	'	7	G	W	g	w
8	1000	BS	CAN	(	8	H	X	h	x
9	1001	HT	EM	)	9	I	Y	i	y
A	1010	LF	SUB	*	:	J	Z	j	z
B	1011	VT	ESC	+	;	K	[	k	}
C	1100	FF	FS	,	<	L	\	l	
D	1101	CR	GS	-	=	M	]	m	{
E	1110	SO	RS	.	>	N	^	n	~
F	1111	SI	US	/	?	O	_	o	DEL

Décoder les caractères ci-dessous encodés en ASCII :

0b0100101 :

0b1111101 :

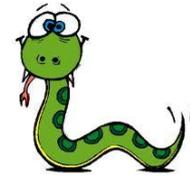
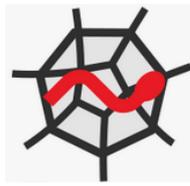
Coder les caractères ci-dessous avec la table de caractère ASCII :

A :

+ :

Décoder le message ci-dessous encodés en ASCII :

0x56 0x69 0x76 0x65 0x20  
0x4E 0x53 0x49 0x21:



Cocher les affirmations correctes :

- Le code ASCII standard est sur 8 bits
- Le code ASCII étendu permet de coder 256 caractères
- Il existe environ 1 million de caractères à coder dans le monde
- Le jeu de caractère UNICODE est compatible avec le jeu de caractère ASCII sur 7 bits
- Une fois décodé un caractère peut être affiché directement sur un ordinateur
- Un glyphe est une rayure sur un CD-ROM

**UTF-8** (abréviation de l'anglais Universal Character Set Transformation Format - 8 bits) est un codage de caractères informatiques conçu pour coder l'ensemble des caractères du « répertoire universel de caractères codés ». Le codage est de taille variable entre 1 et 4 octets. (Source Wikipédia)

Number of bytes	Bits for code point	First code point	Last code point	Byte 1	Byte 2	Byte 3	Byte 4
1	7	U+0000	U+007F	0xxxxxxx			
2	11	U+0080	U+07FF	110xxxxx	10xxxxxx		
3	16	U+0800	U+FFFF	1110xxxx	10xxxxxx	10xxxxxx	
4	21	U+10000	U+10FFFF	11110xxx	10xxxxxx	10xxxxxx	10xxxxxx

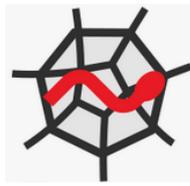
Dans la table résumée des codages en UTF-8 les 'x' représente des valeurs possibles pour le codage, chaque 'x' peut contenir 0 ou 1. Combien peut-on coder de caractères avec cette organisation ?

- 592000
- $2^{7+11+16+21}$
- 2164864
- beaucoup

En Python, les objets de type `bool`, `int`, `str`, `bytes`, `tuple`, `range` et `frozenset` sont immutables. Tous les autres types, les listes, les dictionnaires, ou les instances de vos propres classes sont des objets mutables.

Un objet chaine de caractères en Python :

- Ne peut jamais être modifiée
- Pour obtenir la liste des méthodes associées à l'objet de type chaine de caractère str la commande est : `dir(str)`
- Pour obtenir la liste des méthodes associées à l'objet de type chaine de caractère str la commande est : `help(str)`
- Seul l'administrateur du poste peut modifier une chaine de caractères



## Corrigé :

Décoder les caractères ci-dessous encodés en ASCII :

0b0100101 : %

0b1111101 : {

Coder les caractères ci-dessous avec la table de caractère ASCII :

A : 1000001

+ : 0101011

Décoder le message ci-dessous encodés en ASCII :

0x56 0x69 0x76 0x65 0x20 0x4E 0x53 0x49 0x21: **Vive NSI !**

Dans la table résumée des codages en UTF-8 les 'x' représente des valeurs possibles pour le codage, chaque 'x' peut contenir 0 ou 1. Combien peut-on coder de caractères avec cette organisation ?

- 592000
- $2^{7+11+16+21}$
- 2164864
- beaucoup

Un objet chaîne de caractères en Python :

- Ne peut jamais être modifiée
- Pour obtenir la liste des méthodes associées à l'objet de type chaîne de caractère str la commande est : dir(str) **On obtient uniquement la liste des noms de méthodes**
- Pour obtenir la liste des méthodes associées à l'objet de type chaîne de caractère str la commande est : help(str) **On obtient la liste des noms de méthodes avec leurs descriptions**
- Seul l'administrateur du poste peut modifier une chaîne de caractères