

# Chaînes de caractères

## – Analyse d'un texte

### Caractère et chaîne

- Un **caractère** est un symbole unique, par exemple une lettre minuscule "a", une lettre majuscule "B", un symbole spécial "&", un symbole représentant un chiffre "7", une espace " " que l'on notera aussi "␣".
- Pour désigner un caractère, il faut le mettre entre guillemets simples 'z' ou entre guillemets doubles "z".
- Une **chaîne de caractères** est une suite de caractères, comme un mot "Bonjour", une phrase 'Il fait beau.', un mot de passe "N[w5ms}e!".
- Le type d'un caractère ou d'une chaîne est `str` (pour *string*).

## Opérations sur les chaînes

- La **concaténation**, c'est-à-dire la mise bout à bout de deux chaînes, s'effectue à l'aide de l'opérateur `+`. Par exemple `"para"+"pluie"` donne la chaîne `"parapluie"`.
- La chaîne vide `" "` est utile lorsque l'on veut initialiser une chaîne avant d'y ajouter d'autres caractères.
- La **longueur** d'une chaîne est le nombre de caractères qu'elle contient. Elle s'obtient par la fonction `len()`. Par exemple `len("Hello World")` renvoie 11 (l'espace compte comme un caractère).

## Opérations sur les chaînes

Si `mot` est une chaîne alors on peut récupérer chaque caractère par `mot[i]`. Par exemple si `mot = "avion"` alors :

- `mot[0]` est le caractère "a",
- `mot[1]` est le caractère "v",
- `mot[2]` est le caractère "i",
- `mot[3]` est le caractère "o",
- `mot[4]` est le caractère "n".

Lettre	a	v	i	o	n
Rang	0	1	2	3	4

Si `mot` est une chaîne, les caractères s'obtiennent par `mot[i]` pour `i` variant de 0 à `len(mot) - 1`.

## Sous-chaînes

On peut extraire plusieurs caractères d'une chaîne à l'aide de la syntaxe `mot[i:j]` qui renvoie une chaîne formée des caractères numéro  $i$  à  $j - 1$  (attention le caractère numéro  $j$  n'est pas inclus!).

Par exemple si `mot = "vendredi"` alors :

- `mot[0:4]` renvoie la sous-chaîne "vend" formée des caractères de rang 0, 1, 2 et 3 (mais pas 4),
- `mot[3:6]` renvoie "dre" correspondant aux rangs 3, 4 et 5.

Lettre	v	e	n	d	r	e	d	i
Rang	0	1	2	3	4	5	6	7

Autre exemple : `mot[1:len(mot)-1]` renvoie le mot privé de sa première et dernière lettre.

## Un peu plus sur les chaînes

- Une boucle `for ... in ...` permet de parcourir une chaîne, caractère par caractère :

```
for caractere in mot:
    print(caractere)
```

- On peut tester si un caractère appartient à une certaine liste de caractères. Par exemple :

```
if caractere in ["a", "A", "b", "B", "c", "C"]:
```

permet d'exécuter des instructions si le caractère `caractere` est l'une des lettres `a`, `A`, `b`, `B`, `c`, `C`.

- Pour éviter certaines lettres, on utiliserait :

```
if caractere not in ["X", "Y", "Z"]:
```

# Codage des caractères

Un caractère est stocké par l'ordinateur sous la forme d'un entier. Pour le codage ASCII/unicode, la lettre majuscule « A » est codé par 65, la lettre minuscule « h » est codée par 104, le symbole « # » par 35.

Voici la table des premiers caractères. Les numéros 0 à 32 ne sont pas des caractères imprimables. Cependant le numéro 32 est le caractère espace " ".

33	!	43	+	53	5	63	?	73	I	83	S	93	]	103	g	113	q	123	{
34	"	44	,	54	6	64	@	74	J	84	T	94	^	104	h	114	r	124	
35	#	45	-	55	7	65	A	75	K	85	U	95	_	105	i	115	s	125	}
36	\$	46	.	56	8	66	B	76	L	86	V	96	'	106	j	116	t	126	~
37	%	47	/	57	9	67	C	77	M	87	W	97	a	107	k	117	u	127	-
38	&	48	0	58	:	68	D	78	N	88	X	98	b	108	l	118	v		
39	'	49	1	59	;	69	E	79	O	89	Y	99	c	109	m	119	w		
40	(	50	2	60	<	70	F	80	P	90	Z	100	d	110	n	120	x		
41	)	51	3	61	=	71	G	81	Q	91	[	101	e	111	o	121	y		
42	*	52	4	62	>	72	H	82	R	92	\	102	f	112	p	122	z		

# Codage des caractères

La fonction `chr()` est une fonction Python qui renvoie le caractère associé à un code.

python : `chr()`

Usage : `chr(code)`

Entrée : un code (un entier)

Sortie : un caractère

Exemple :

- `chr(65)` renvoie "A"
- `chr(123)` renvoie "{"

La fonction `ord()` est une fonction Python correspondant à l'opération inverse : elle renvoie le code associé à un caractère.

python : `ord()`

Usage : `ord(carac)`

Entrée : un caractère (une chaîne de longueur 1)

Sortie : un entier

Exemple :

- `ord("A")` renvoie 65
- `ord("*")` renvoie 42