

# Fiche de révision 5:

## Chaînes de caractères

Thème 1 ♦ Fiche 5

### Tableau de correspondance caractère-nombre.

Chaque caractère est associé à un nombre. On parle de jeu de caractères codés (en anglais **charset**).

Les nombres correspondants à des caractères sont ensuite codés en binaire.

### Code ASCII

Le code **ASCII** (American Standard Code for Information Interchange).

(créé en 1961). Il permet de coder **sur 7 bits 128** symboles numérotés de 0 à 127.

	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F	
0	0 ^@ NUL NULL	1 ^A SOH START OF HEADING	2 ^B STX START OF TEXT	3 ^C ETX END OF TEXT	4 ^D EOT END OF TRANSM.	5 ^E ENQ ENQUIRY	6 ^F ACK ACKNOWLEDGE	7 ^G BEL BELL	8 ^H BS BACKSP.	9 ^I HT CHARACT. TABULATION	10 ^J LF LINE FEED	11 ^K VT LINE TABULATION	12 ^L FF FORM FEED	13 ^M CR CARRIAGE RETURN	14 ^N SO SHIFT OUT	15 ^O SI SHIFT IN	
1	16 ^P DLE DATA LINK ESCAPE	17 ^Q DC1 DEVICE CONTROL 1	18 ^R DC2 DEVICE CONTROL 2	19 ^S DC3 DEVICE CONTROL 3	20 ^T DC4 DEVICE CONTROL 4	21 ^U NAK NEG. ACKNOWLEDGE	22 ^V SYN SYNCHRONOUS IDLE	23 ^W ETB END OF TRANSM.	24 ^X CAN CANCEL	25 ^Y EM END OF MEDIUM	26 ^Z SUB SUBSTITUTE	27 ^[ ESC ESCAPE	28 ^\ FS INFO. SEP. 4	29 ^] GS INFO. SEP. 3	30 ^^ RS INFO. SEP. 2	31 ^_ US INFO. SEP. 1	
2	32 SPACE	33 ^@ ! EXCLAM. MARK	34 ^B " QUOT. MARK	35 ^# # NUMBER SIGN	36 ^\$ \$ DOLLAR SIGN	37 %^ % PERCENT SIGN	38 ^& & AMPERSAND	39 ^' ' APOS-TROPHE	40 ^( ( LEFT PAREN.	41 ^) ) RIGHT PAREN.	42 ^* * ASTERISK	43 ^+ + PLUS SIGN	44 ^, , COMMA	45 ^- - HYPHEN-MINUS	46 ^. . FULL STOP	47 ^/ / SOLIDUS	
3	48 0 DIGIT ZERO	49 1 DIGIT ONE	50 2 DIGIT TWO	51 3 DIGIT THREE	52 4 DIGIT FOUR	53 5 DIGIT FIVE	54 6 DIGIT SIX	55 7 DIGIT SEVEN	56 8 DIGIT EIGHT	57 9 DIGIT NINE	58 : COLON	59 ; SEMI-COLON	60 < < LS. THAN SIGN	61 = = EQUALS SIGN	62 > > GR. THAN SIGN	63 ? ? QUESTION MARK	
4	64 ^@ COMMERCIAL AT	65 A	66 B	67 C	68 D	69 E	70 F	71 G	72 H	73 I	74 J	75 K	76 L	77 M	78 N	79 O	
5	80 P	81 Q	82 R	83 S	84 T	85 U	86 V	87 W	88 X	89 Y	90 Z	91 [ LEFT SQ. BRACKET	92 \ REVERSE SOLIDUS	93 ] RT. SQ. BRACKET	94 ^ CIRCUMFLEX ACCENT	95 _ LOW LINE	
6	96 ^` GRAVE ACCENT	97 a	98 b	99 c	100 d	101 e	102 f	103 g	104 h	105 i	106 j	107 k	108 l	109 m	110 n	111 o	
7	112 p	113 q	114 r	115 s	116 t	117 u	118 v	119 w	120 x	121 y	122 z	123 { L. CURLY BRACKET	124   VERTICAL LINE	125 } R. CURLY BRACKET	126 ~ TILDE	127 ^? DEL DELETE	

## La norme ISO-8859-1

Cette norme étend l'ASCII à **8 bits** pour pouvoir coder plus de caractères en particulier les caractères accentués.

Elle permet de coder **256 caractères**. Les 128 premiers sont ceux de l'ASCII.

## L'unicode ou UTF-8

Cette norme prolonge la norme ISO-8859-1.

Elle vise à coder tout caractère de n'importe quel langage. La place attribuée au codage du caractère est variable selon sa fréquence d'utilisation. Les caractères de la norme ISO-8859-1 sont codés sur **1 octet**, mais les caractères plus rares peuvent nécessiter **jusqu'à 4 octets**.

## Le code d'un caractère en Python

La fonction *ord* en Python renvoie le nombre entier représentant le code Unicode du caractère représenté par la chaîne donnée.

```
>>ord("a")  
97
```

La fonction *chr* en Python renvoie le caractère correspondant au code Unicode donné.

```
>>chr(97)  
a
```