

➤ **American Standard Code for Information Interchange : ASCII**

Différentes normes ont été inventées pour communiquer à distance avec uniquement deux caractères : Le morse en 1844, le code Baudot (ou Murray code) en 1874, le code ASCII en 1963 puis normalisé ANSI (American National Standards Institute) en 1986.

Le code ASCII est codé sur 7 bits, il permet donc de coder 128 caractères différents :

Extrait :

Dec	Hex	Char	Dec	Hex	Char	Dec	Hex	Char	Dec	Hex	Char
0	00	Null	32	20	Space	64	40	@	96	60	`
1	01	Start of heading	33	21	!	65	41	A	97	61	a
2	02	Start of text	34	22	"	66	42	B	98	62	b
3	03	End of text	35	23	#	67	43	C	99	63	c
4	04	End of transmit	36	24	\$	68	44	D	100	64	d
5	05	Enquiry	37	25	%	69	45	E	101	65	e
6	06	Acknowledge	38	26	&	70	46	F	102	66	f
7	07	Audible bell	39	27	'	71	47	G	103	67	g
8	08	Backspace	40	28	(72	48	H	104	68	h
9	09	Horizontal tab	41	29)	73	49	I	105	69	i

Remarque :

de 0 à 31 : caractères de contrôle

de 48 à 57 : chiffres

de 65 à 90 : lettres majuscules

de 97 à 122 : lettres minuscules

Astuce :

On peut accéder directement à un caractère en appuyant sur Alt puis son code ASCII en décimal...

➤ **ASCII étendu**

Le code ASCII standard sur 7 bits ne comporte aucun caractère accentué et peu de caractères spéciaux, aussi il a été étendu sur 8 bits pour coder 128 caractères supplémentaires.

Mais le code ASCII étendu possède de nombreuses variantes suivant le pays ou le système sur lequel il est utilisé, les plus courants étaient OEM au début puis ANSI.

Mais il existe beaucoup d'autres versions, elles sont répertoriées dans la norme ISO 8859

(ISO 8859-1 : latin-1 pour l'Europe occidentale par exemple est le plus utilisé)

<u>Extrait</u> :				DEC	HEX	CHAR	DEC	HEX	CHAR	DEC	HEX	CHAR
		OEM	ANSI			OEM	ANSI			OEM	ANSI	
160	A0	á		192	C0	Ł	À	224	E0	α	à	
161	A1	í	ı	193	C1	ł	Á	225	E1	β	á	
162	A2	ó	ç	194	C2	Ť	Â	226	E2	Γ	â	
163	A3	ú	£	195	C3	Ŧ	Ã	227	E3	Π	ã	

➤ **Unicode**

L'Unicode est un codage sur 24 bits (soit 3 octets) qui pourrait coder jusqu'à 16⁶ = 16 777 216 caractères mais dans la pratique on ne va que de U+0000 à U+10FFFF (ceux de l'ISO 10 646 avec des règles d'écritures supplémentaires). Il est compatible avec tous les pays et tous les systèmes et répertorie actuellement plus de 100 000 caractères différents...

➤ **Universal character set Transformation Format 8 bits : UTF-8**

Le codage Unicode est un peu lourd à utiliser, aussi il existe une autre norme plus légère et très répandue : l'UTF-8. (Utilisé par 92,3 % des sites web d'après les chiffres d'octobre 2018 sur w3techs.com)

Cette norme est compatible avec l'ASCII et l'Unicode et utilise de 2 à 4 octets pour coder les caractères.

taille	valeurs autorisées	significatifs
1 octet	0xxxxxxx	7 bits
2 octets	110xxxxx 10xxxxxx	11 bits
3 octets	1110xxxx 10xxxxxx 10xxxxxx	16 bits
4 octets	11110xxx 10xxxxxx 10xxxxxx 10xxxxxx	21 bits