



1. Les tables de caractères

A chaque caractère, on associe un code numérique, mais il existe de nombreuses normes...

Dans les tables qui vont suivre, il faut ajouter le nombre en ligne avec le nombre en colonne pour obtenir le code correspondant à la case située à l'intersection de la ligne et de la colonne.

Ces nombres sont écrits en hexadécimal.

Par exemple dans le tableau ci-dessous, le caractère A est située dans la case à l'intersection de la **ligne 40** et de la **colonne 1** ce qui fait $40 + 1 = 41$. Le code hexadécimal du caractère A est donc 41.

PDF : en [archive] v · d · m	0	1	2	3	4	5	6	7	8
U+0000	NUL	SOH	STX	ETX	EOT	ENQ	ACK	BEL	BS
U+0010	DLE	DC1	DC2	DC3	DC4	NAK	SYN	ETB	CAN
U+0020	SP	!	"	#	\$	%	&	'	(
U+0030	0	1	2	3	4	5	6	7	8
U+0040	@	A	B	C	D	E	F	G	H

2. ASCII

L'American Standard Code for Information Interchange est une norme informatique de codage de caractères apparue dans les années 1960. Elle définit 128 codes à 7 bits, comprenant 95 caractères imprimables : les chiffres arabes de 0 à 9, les lettres minuscules et capitales de A à Z, et des symboles mathématiques et de ponctuation.

PDF : en [archive] v · d · m	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
U+0000	NUL	SOH	STX	ETX	EOT	ENQ	ACK	BEL	BS	HT	LF	VT	FF	CR	SO	SI
U+0010	DLE	DC1	DC2	DC3	DC4	NAK	SYN	ETB	CAN	EM	SUB	ESC	FS	GS	RS	US
U+0020	SP	!	"	#	\$	%	&	'	()	*	+	,	-	.	/
U+0030	0	1	2	3	4	5	6	7	8	9	:	;	<	=	>	?
U+0040	@	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
U+0050	P	Q	R	S	T	U	V	W	X	Y	Z	[\]	^	_
U+0060	`	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o
U+0070	p	q	r	s	t	u	v	w	x	y	z	{		}	~	DEL

3. ISO/CEI 8859-1

Cette norme apparue le 1^{er} février 1986 définit l'alphabet latin numéro 1, qui consiste en 191 caractères de l'alphabet latin, chacun d'entre eux étant codé sur un octet.

	x0	x1	x2	x3	x4	x5	x6	x7	x8	x9	xA	xB	xC	xD	xE	xF
0x	positions inutilisées															
1x																
2x	SP	!	"	#	\$	%	&	'	()	*	+	,	-	.	/
3x	0	1	2	3	4	5	6	7	8	9	:	;	<	=	>	?
4x	@	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
5x	P	Q	R	S	T	U	V	W	X	Y	Z	[\]	^	_
6x	`	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o
7x	p	q	r	s	t	u	v	w	x	y	z	{		}	~	
8x	positions inutilisées															
9x																
Ax	NBSP	¡	¢	£	¤	¥	¦	§	¨	©	ª	«	¬	­	®	¯
Bx	°	±	²	³	´	µ	¶	·	¸	¹	º	»	¼	½	¾	¿
Cx	À	Á	Â	Ã	Ä	Å	Æ	Ç	È	É	Ê	Ë	Ì	Í	Î	Ï
Dx	Ð	Ñ	Ò	Ó	Ô	Õ	Ö	×	Ø	Ù	Ú	Û	Ü	Ý	Þ	ß
Ex	à	á	â	ã	ä	å	æ	ç	è	é	ê	ë	ì	í	î	ï
Fx	ð	ñ	ò	ó	ô	õ	ö	÷	ø	ù	ú	û	ü	ý	þ	ÿ

4. UNICODE

Unicode est un standard informatique qui permet des échanges de textes dans différentes langues, à un niveau mondial. Il donne à tout caractère de n'importe quel système d'écriture un nom et un identifiant numérique, et ce de manière unifiée, quels que soient la plate-forme informatique ou le logiciel utilisés.

UTF-8 est un codage de longueur variable : 1, 2 ou 4 octets selon les besoins.

UTF-16 utilise 2 ou 4 octets et UTF-32 utilise 4 octets.

Par exemple le caractère A a pour code 41_h en UTF-8 et 0041_h en UTF-16 et 00000041_h en UTF-32.

5. Encodage et javascript

L'instruction javascript `String.fromCharCode(code)` permet d'obtenir le caractère Unicode correspondant au code écrit en base 10.

On peut écrire plusieurs codes à la suite, comme ceci : `String.fromCharCode(code1, code2, code3...)`

Réciproquement, l'instruction javascript `"A".charCodeAt()` permet d'obtenir le code Unicode correspondant au caractère A (par exemple).

Remarque 1 : `chaîne_de_caracteres.charCodeAt(position)` permet d'obtenir le code Unicode correspondant au caractère de la chaîne de caractères à la position donnée.

Remarque 2 : `chaîne_de_caracteres.charAt(position)` permet d'obtenir le caractère de la chaîne de caractères à la position donnée, c'est donc la même instruction que `chaîne_de_caracteres[position]` à ne pas confondre avec `charCodeAt` ...

Rappel : `chaîne_de_caracteres.length` permet d'obtenir la taille d'une chaîne de caractères.

6. Questions

Répondez aux questions dans une page HTML, au besoin, utiliser une console javascript

- question 1. Quelles sont les limites du code ASCII ?
- question 2. Quel est l'avantage du code ISO/CEI 8859-1 sur le code ASCII ?
- question 3. L'euro € fait-il partie du code ISO/CEI 8859-1 ? Pourquoi à votre avis ?
- question 4. Quel est le code du caractère A dans la norme ISO/CEI 8859-1 ? Donner la réponse en hexadécimal, puis en décimal.
- question 5. Quels sont les caractères présent dans le code ASCII qui sont également présents dans le code ISO/CEI 8859-1 ?
- question 6. Quel est l'inconvénient de l'encodage UTF-32 par rapport à UTF-8 ?
- question 7. Combien de caractères sont théoriquement possibles en unicode ? Le Grand dictionnaire de la langue chinoise compte plus de 56 000 caractères. Unicode est-il suffisant pour le chinois et pour toutes les langues de tous les pays ? (justifier les réponses par des calculs)
- question 8. Quel est le caractère correspondant à l'unicode 916 ?
- question 9. Quel est l'unicode de l'idéogramme chinois 人 qui signifie l'être humain ? (copier ce caractère et le coller dans la console)
- question 10. A quel texte correspond la suite de code décimal 89,101,115 ?
- question 11. Donner la suite de code UTF-8 correspondant au texte "Parc".
- question 12. Quelle est la longueur de ce texte ? :

Exercice n°1

Créer une boucle qui permet d'afficher l'alphabet (au choix minuscule ou majuscule).

Exercice n°2

Créer page HTML contenant une balise textarea identifiée.

Coller le texte suivant dans le formulaire correspondant :

Voici que la saison décline,
L'ombre grandit, l'azur décroît,
Le vent fraîchit sur la colline,
L'oiseau frissonne, l'herbe a froid.

Août contre septembre lutte ;
L'océan n'a plus d'aleçon ;
Chaque jour perd une minute,
Chaque aurore pleure un rayon.

La mouche, comme prise au piège,
Est immobile à mon plafond ;
Et comme un blanc flocon de neige,
Petit à petit, l'été fond.

Victor Hugo, Dernière gerbe

Ouvrir la console et récupérer le texte dans une variable.

Quelle est la longueur de ce texte (espaces et ponctuations compris) ?

Faire une boucle qui permet d'afficher la position du caractère, le caractère et le code du caractère, comme ceci :

0	"V"	86
1	"o"	111
2	"i"	105
3	"c"	99
4	"i"	105
5	" "	32
6	"q"	113
7	"u"	117
8	"e"	101
9	" "	32

Montrer le résultat à votre professeur.

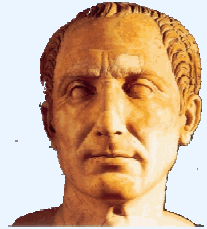
Quel est le code correspondant au caractère espace ? Quelle est son abréviation dans le code ASCII ?

A quoi correspond le caractère en position 28 ? Quelle est son abréviation dans le code ASCII ?

A quelles autres position trouve-t-on ce caractère ?

Exercice n°3

Le chiffre de César consiste en un décalage de deux caractères par exemple.
Ainsi dans un message chiffré, la lettre a devient c.



Créer une page HTML avec une zone de saisie pour le message original et un conteneur pour le message chiffré.

Dans un premier temps on se limitera à un message contenant un seul caractère.

Le message chiffré doit apparaître dès que le message original est écrit : pour cela utiliser l'évènement oninput.

- ✓ Agrémenter votre page avec la photo du buste de César et le nom de Jules César chiffré.
- ✓ Que se passe-t-il si on tape un message commençant par la lettre y ou z ?
- ✓ Améliorer votre script pour tenir compte de ce cas de figure.
- ✓ Améliorer votre script pour coder un message de taille quelconque.
- ✓ Améliorer votre script pour ne pas toucher au caractère espace.

Placer des commentaires (précédés de `//`) dans votre script correspondant à ces trois améliorations.