

RND. Texte. Exercices

Exercice 1 : Vrai / Faux

	VRAI	FAUX
Le code ASCII est suffisant pour la langue française.		
La norme UTF-8 inclut la norme ASCII.		
Unicode peut être utilisé directement.		
Avec 4 bits, on peut représenter les 26 lettres majuscules de l'alphabet.		
Choisir l'encodage UTF-8 dans son navigateur permet de bien lire les caractères accentués.		
Le caractère @ est présent dans le code ASCII.		

Exercice 2 : QCM

Pour chaque question, une seule réponse est correcte parmi les quatre proposées.

Question 1 : Le code ASCII permet de représenter en binaire les caractères alphanumériques. Quel est son principal inconvénient ?

1. Il utilise beaucoup de bits.
2. Il ne différencie pas les majuscules des minuscules.
3. Il ne représente pas les caractères accentués.
4. Il n'est pas compatible avec la plupart des systèmes informatiques.

Question 2 : Quel est un avantage du codage UTF-8 par rapport au codage ASCII ?

1. Il permet de coder un caractère sur un octet au lieu de deux.
2. Il permet de coder les majuscules.
3. Il permet de coder tous les caractères.
4. Il permet de coder différentes polices de caractères.

Question 3 : On considère les codes ASCII en écriture hexadécimale (en base 16). Le code ASCII de la lettre A est 0x41, celui de la lettre B est 0x42, celui de la lettre C est 0x43, etc. Quel est le code ASCII, en hexadécimal, de la lettre X (c'est la 24e lettre de l'alphabet usuel) ?

1. 0x58.
2. 0x64.
3. 0x7A.
4. 0x88.

Question 4 : Laquelle de ces affirmations concernant le codage UTF-8 des caractères est vraie ?

1. Le codage UTF-8 est sur 7 bits.
2. Le codage UTF-8 est sur 8 bits.
3. Le codage UTF-8 est sur 1 à 4 octets.
4. le codage UTF-8 est sur 8 octets.

Question 5 : Parmi les noms suivants, lequel **n'est pas** celui d'une méthode d'encodage des caractères ?

1. UTF-16.
2. ASCII.
3. Arial.
4. Unicode.

Exercice 3 : Décoder les caractères ASCII.

La table ci-après donne le code associé à chacun ses caractères ASCII imprimables (les cases vides correspondent à des caractères de contrôle comme « espace », « retour à la ligne », « tabulation » etc.) Le code du caractère en hexadécimal s'obtient en écrivant le numéro de la ligne suivi du numéro de la colonne. Par exemple, la lettre M a pour code hexadécimal 4D₁₆, c'est-à-dire 77₁₀ en décimal.

Code ASCII

Dec	Hx	Oct	Char	Dec	Hx	Oct	Html	Chr	Dec	Hx	Oct	Html	Chr	Dec	Hx	Oct	Html	Chr
0	0	000	NUL (null)	32	20	040	 	Space	64	40	100	@	@	96	60	140	`	`
1	1	001	SOH (start of heading)	33	21	041	!	!	65	41	101	A	A	97	61	141	a	a
2	2	002	STX (start of text)	34	22	042	"	"	66	42	102	B	B	98	62	142	b	b
3	3	003	ETX (end of text)	35	23	043	#	#	67	43	103	C	C	99	63	143	c	c
4	4	004	EOT (end of transmission)	36	24	044	$	\$	68	44	104	D	D	100	64	144	d	d
5	5	005	ENQ (enquiry)	37	25	045	%	%	69	45	105	E	E	101	65	145	e	e
6	6	006	ACK (acknowledge)	38	26	046	&	&	70	46	106	F	F	102	66	146	f	f
7	7	007	BEL (bell)	39	27	047	'	'	71	47	107	G	G	103	67	147	g	g
8	8	010	BS (backspace)	40	28	050	((72	48	110	H	H	104	68	150	h	h
9	9	011	TAB (horizontal tab)	41	29	051))	73	49	111	I	I	105	69	151	i	i
10	A	012	LF (NL line feed, new line)	42	2A	052	*	*	74	4A	112	J	J	106	6A	152	j	j
11	B	013	VT (vertical tab)	43	2B	053	+	+	75	4B	113	K	K	107	6B	153	k	k
12	C	014	FF (NP form feed, new page)	44	2C	054	,	,	76	4C	114	L	L	108	6C	154	l	l
13	D	015	CR (carriage return)	45	2D	055	-	-	77	4D	115	M	M	109	6D	155	m	m
14	E	016	SO (shift out)	46	2E	056	.	.	78	4E	116	N	N	110	6E	156	n	n
15	F	017	SI (shift in)	47	2F	057	/	/	79	4F	117	O	O	111	6F	157	o	o
16	10	020	DLE (data link escape)	48	30	060	0	0	80	50	120	P	P	112	70	160	p	p
17	11	021	DC1 (device control 1)	49	31	061	1	1	81	51	121	Q	Q	113	71	161	q	q
18	12	022	DC2 (device control 2)	50	32	062	2	2	82	52	122	R	R	114	72	162	r	r
19	13	023	DC3 (device control 3)	51	33	063	3	3	83	53	123	S	S	115	73	163	s	s
20	14	024	DC4 (device control 4)	52	34	064	4	4	84	54	124	T	T	116	74	164	t	t
21	15	025	NAK (negative acknowledge)	53	35	065	5	5	85	55	125	U	U	117	75	165	u	u
22	16	026	SYN (synchronous idle)	54	36	066	6	6	86	56	126	V	V	118	76	166	v	v
23	17	027	ETB (end of trans. block)	55	37	067	7	7	87	57	127	W	W	119	77	167	w	w
24	18	030	CAN (cancel)	56	38	070	8	8	88	58	130	X	X	120	78	170	x	x
25	19	031	EM (end of medium)	57	39	071	9	9	89	59	131	Y	Y	121	79	171	y	y
26	1A	032	SUB (substitute)	58	3A	072	:	:	90	5A	132	Z	Z	122	7A	172	z	z
27	1B	033	ESC (escape)	59	3B	073	;	;	91	5B	133	[[123	7B	173	{	{
28	1C	034	FS (file separator)	60	3C	074	<	<	92	5C	134	\	\	124	7C	174	|	
29	1D	035	GS (group separator)	61	3D	075	=	=	93	5D	135]]	125	7D	175	}	}
30	1E	036	RS (record separator)	62	3E	076	>	>	94	5E	136	^	^	126	7E	176	~	~
31	1F	037	US (unit separator)	63	3F	077	?	?	95	5F	137	_	_	127	7F	177		DEL

Source: www.asciitable.com

Question : A quel nom connu correspond la liste des codes ASCII suivants, donnés en hexadécimal ?

47 65 6F 72 67 65 20 42 6F 6F 6C 65.

Exercice 4 : Décoder les caractères UTF-8.

En UTF-8, le codage des 128 premiers caractères est celui de l'ASCII et donc sur un octet. Les autres caractères sont représentés sur plusieurs octets (entre 2 et 4).

La série d'octets suivants, donnés en hexadécimal, a été relevée dans un fichier codé en UTF-8.

43 6F 64 C3 A9 20 65 6E 20 55 54 46 2D 38.

Il contient uniquement des caractères de la table ASCII à l'exception d'un « é ».

- Quelle est la **séquence d'octets** qui représente le « é » ? Qu'est-il inscrit dans le fichier ?
- Si le fichier avait été interprété en **latin-1** (tableau ci-dessous) que serait-il alors affiché ?

Code latin-1

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

Exercice 5 (*) : Coder en UTF-8 (réponses <https://fr.wikipedia.org/wiki/UTF-8#Exemples>)

Regarder la vidéo suivante : <https://www.youtube.com/watch?v=uxUohPrZ3n4>

Auteur : Christian Queinnec, durée : 15 min 57 sec.

Coder selon la norme UTF-8 les caractères suivants :

- A (nombre associé : 41₁₆)
- é (nombre associé : E9₁₆)
- 💎 (nombre associé : FFD₁₆)