



① Table d'encodage UTF8 (Universal Transformation Format)

Point de code			Encodage du point de code en UTF8
Compris entre (décimal)	Compris entre (hexadécimal)	écriture binaire (séparée par quartets)	
0 et 127	00 et 7F	xxxxxxx (7 bits)	0xxxxxxx
128 et 2047	80 et 7FF	xxxxxxxxxxx (11 bits)	110xxxxx 10xxxxxx
2048 et 65535	800 et FFFF	xxxxxxxxxxxxxxx (16 bits)	1110xxxx 10xxxxxx 10xxxxxx
65536 et 1114111	10 000 et 10 FFFF	xxxxxxxxxxxxxxxxxxx (21 bits)	11110xxx 10xxxxxx 10xxxxxx 10xxxxxx

② ENCODAGE : UNICODE vers UTF8

Caractère	Point de code hexadécimal	Point de code décimal	Point de code binaire
Euro €	0x20AC	8364	10000010101100

Pour savoir le nombre d'octets nécessaires à l'encodage UTF8 (quelle ligne de la table d'encodage UTF8 ?),

☒ On compte le nombre de bits dans le point de code binaire du caractère : **10 0000 1010 1100**

→ il y a 14 bits donc il faut 3 octets pour l'encodage UTF8. On utilise donc la 3^e ligne de la table d'encodage UTF8.

☒ On prépare les entêtes UTF8 des octets et on remplit avec les bits du point de code binaire.

Octet 1								Octet 2								Octet 3							
1	1	1	0	0	0	1	0	1	0	0	0	0	0	1	0	1	0	1	1	0	0		
E				2				8				2				A				C			

☒ On peut alors convertir l'encodage binaire UTF8 en hexadécimal.

Caractère	Encodage UTF8 binaire	Encodage UTF8 hexadécimal
Euro €	111000101000001010101100	0xE282AC

③ DECODAGE : UTF8 vers UNICODE

☒ On convertit l'encodage UTF8 en hexadécimal en binaire (l'encodage binaire UTF8).

☒ On retrouve le point de code binaire en supprimant les bits d'entête de l'encodage binaire UTF8.

☒ On groupe les bits par 4 (en partant de la droite) et on convertit chaque quartet en hexadécimal.

Caractère : π

Encodage UTF-8 : CF 80 >> 11001111 10000000

>> 11001111 10000000

Point de code (binaire) : >> 0111100000 >> 011 1100 0000

Point de code (hexadécimal) : >> 3 C 0

Point de code (hexadécimal) : U+3C0

Point de code (décimal) : 960

Caractère	Encodage UTF8 binaire	Encodage UTF8 hexadécimal
pi π	11001111 10000000	0x CF 80

Exercices : unicode vers utf8

Caractère	Point de code hexadécimal	Point de code décimal	Point de code binaire
Ensemble vide ∅	0xD8	216	1101 1000

Octet 1								Octet 2							
1	1	0	0	0	0	1	1	1	0	0	1	1	0	0	0
C				3				9				8			

Caractère	Encodage UTF8 binaire	Encodage UTF8 hexadécimal
Ensemble vide ∅	11000011 10011000	0xC398

Caractère	Point de code hexadécimal	Point de code décimal	Point de code binaire
Danger ⚠	0x26A0	9888	10 0110 1010 0000

Octet 1								Octet 2								Octet 3							
1	1	1	0	0	0	1	0	1	0	0	1	1	0	1	0	1	0	0	0	0	0	0	
E				2				8				2				A				C			

Caractère	Encodage UTF8 binaire	Encodage UTF8 hexadécimal
Danger ⚠	11100010 10011010 10100000	0xE29AA0

Caractère	Point de code hexadécimal	Point de code décimal	Point de code binaire
Atome ⚗	0x269B	9883	10 0110 1001 1011

Octet 1								Octet 2								Octet 3							
1	1	1	0	0	0	1	0	1	0	0	1	1	0	1	0	1	0	0	1	1	0	1	1
E				2				8				2				A				C			

Caractère	Encodage UTF8 binaire	Encodage UTF8 hexadécimal
Atome ⚗	11100010 10011010 10011011	0xE29A9B

Caractère	Point de code hexadécimal	Point de code décimal	Point de code binaire
Drapeau 🏳️	0x1F3F2	127986	1 1111 0011 1111 0010

Octet 1								Octet 2								Octet 3								Octet 4							
1	1	1	1	0	0	0	0	1	0	0	1	1	1	1	1	1	0	0	0	0	1	1	1	1	1	0	0	0	0		
F				0				9				F				8				F				B				2			

Caractère	Encodage UTF8 binaire	Encodage UTF8 hexadécimal
Drapeau 🏳️	11110000 10011111 10001111 10110010	0x F09F8FB2