



TANKERINO.com

Esercizi rappresentazione del testo

Esercizio 1: Converti in binario "? - Mi presti 10€ Giò? ?? " con le 2 codifiche unico

Nota: Al posto dal simbolo ? ci dovrebbe essere un simbolo cinese e al posto di ?? c'è un simbolo cinese antico.

UTF-8

Carattere	Esadecimale UTF-8
?	E6 BC A2
M	4D
i	69
(spazio)	20
p	70
r	72
e	65
s	73
t	74
i	69
(spazio)	20
l	31
o	30
€	E2 82 AC
(spazio)	20
G	47
i	69
ò	C3 B2

?

3F

??

F0 A0 80 80

Nella tabella, ogni carattere è associato al suo codice esadecimale equivalente in UTF-8.

Alcuni caratteri, come "?", "€", "ò", e "??", sono rappresentati da più byte in UTF-8, e quindi hanno cod esadecimali più lunghi.

Svolgimento:

1. ?

Codifica UTF-8:

- E6: 11100110
- BC: 10111100
- A2: 10100010

2. Mi presti 10€ Giò?

- M: 01001101
- i: 01101001
- (spazio): 00100000
- p: 01110000
- r: 01110010
- e: 01100101
- s: 01110011
- t: 01110100
- i: 01101001
- (spazio): 00100000
- l: 00110001
- 0: 00110000
- €: 11100010 10000010 10101100
- (spazio): 00100000
- G: 01000111
- i: 01101001
- ò: 11000011 10110010
- ?: 00111111

3. ??

- Codifica UTF-8: 11110000101000001000000010000000

Concatenando tutti i vari bit si ottiene una codifica UTF-8

UTF-32

In UTF-32, ogni carattere è rappresentato da 4 byte (32 bit). Ecco la codifica UTF-32 della tua frase "?]
presti 10€ Giò? ??":

1. ?: 00004E2D
2. M: 0000004D
3. i: 00000069
4. spazio: 00000020
5. p: 00000070
6. r: 00000072
7. e: 00000065
8. s: 00000073
9. t: 00000074
10. i: 00000069
11. spazio: 00000020
12. l: 00000031
13. o: 00000030
14. €: 00020AC
15. spazio: 00000020
16. G: 00000047
17. i: 00000069
18. ò: 000000F2
19. ?: 0000003F
20. ??: 00020080

Ogni carattere è rappresentato da un codice esadecimale di 4 byte. In UTF-32, la codifica è più diretta rispetto a UTF-8, poiché ogni carattere ha una lunghezza fissa di 32 bit.

Esercizio2: Converti da binario in utf-8

11000011101100010100001011100110101111001010001011110000100000001000000010000000010

Ricordiamo la regola generale:

1. Byte singolo (ASCII): I caratteri che rientrano nel range ASCII (0-127) vengono codificati con un byte. Questo byte inizia sempre con lo 0 di bit di alto ordine. 0xxxxxxx
2. 2 Byte: Se un carattere richiede due byte in UTF-8, il primo byte inizierà con 110, e il secondo byte inizierà con 10. 110xxxxx 10xxxxxx
3. 3 Byte: Per i caratteri che richiedono tre byte, il primo byte inizia con 1110, e i due byte successivi iniziano entrambi con 10. 1110xxxx 10xxxxxx 10xxxxxx
4. 4 Byte: Infine, i caratteri che richiedono quattro byte avranno il primo byte che inizia con 11110, seguito da tre byte che iniziano tutti con 10. 11110xxx 10xxxxxx 10xxxxxx 10xxxxxx

Simbolo	Binario	UTF-8
A (1)	01000001	41
B (2)	01000010	42
ñ (3)	11000011 10110001	C3 B1
? (4)	11100110 10111100 10100010	E6 BC A2
?? (5)	11110000 10000000 10000000 10000000	F0 80 80 80

Soluzione 3 - 2 - 4 - 5 - 1

(CC BY-NC-SA 3.0) lezione - by tankerino.com

<https://www.tankerino.com>

Questa lezione e' stata realizzata grazie al contributo di:



Risorse per la scuola

<https://www.baobab.school>



Siti web a Varese

<https://www.francescobelloni.it>