

Ejemplos UTF-8

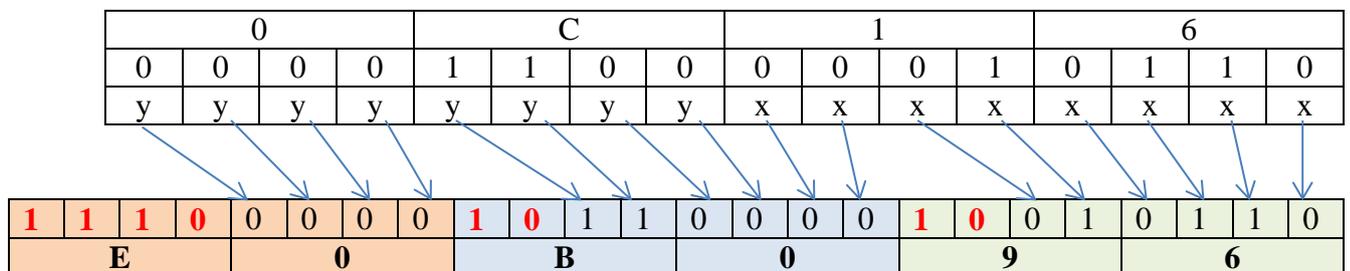
1. Obtener la codificación UTF-8 del carácter 'ఖ' (TELUGU LETTER KHA), con código 0x0C16.

- a. Identificamos el rango:

Rango UTF-32	Formato binario	UTF-8	
0x00 a 0x7F	xxxxxxx ₂	0xxxxxxx ₂	ASCII
0x80 a 0x7FF	yyyxxxxxxxx ₂	110yyyxx ₂ 10xxxxx ₂	Caracteres occidentales
0x800 a 0xFFFF	yyyyyyyyxxxxxxxx ₂	1110yyyy ₂ 10yyyyxx ₂ 10xxxxx ₂	Caracteres orientales
0x1000 a 0x10FFFF	zzzzzyyyyyyyyyxxxxxxxx ₂	11110zzz ₂ 10zzyyyy ₂ 10yyyyxx ₂ 10xxxxx ₂	Caracteres inusuales

1 1 1 0 y y y y 1 0 y y y y x x 1 0 x x x x x x

- b. Pasamos a binario y rellenamos: 0x00000C16



- c. Obtenemos el valor en hexadecimal: 0xE0B096

U+0C16	ఖ	e0 b0 96	TELUGU LETTER KHA
--------	---	----------	-------------------

<http://www.utf8-chartable.de/>

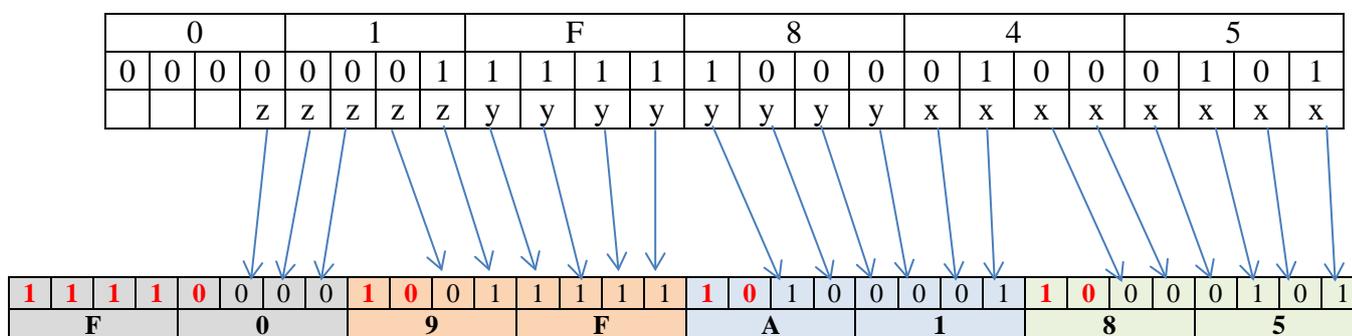
2. Obtener la codificación UTF-8 del carácter (UPWARDS HEAVY ARROW), con código U+1F845.

a. Identificamos el rango:

Rango UTF-32	Formato binario	UTF-8	
0x00 a 0x7F	xxxxxxx ₂	0xxxxxxx ₂	ASCII
0x80 a 0x7FFF	yyyxxxxxxxx ₂	110yyyxx ₂ 10xxxxx ₂	Caracteres occidentales
0x800 a 0xFFFF	yyyyyyyyxxxxxxxx ₂	1110yyyy ₂ 10yyyyxx ₂ 10xxxxx ₂	Caracteres orientales
0x1000 a 0x10FFFF	zzzzzyyyyyyyxxxxxxxx ₂	11110zzz ₂ 10zzyyy ₂ 10yyyyxx ₂ 10xxxxx ₂	Caracteres inusuales

1 1 1 1 0 z z z 1 0 z z y y y y 1 0 y y y y x x 1 0 x x x x x x

b. Pasamos a binario y rellenamos: 0x0001F845



c. Obtenemos el valor en hexadecimal: 0xF09FA185

U+1F845	↑	0xf0 0x9f 0xa1 0x85	UPWARDS HEAVY ARROW
---------	---	---------------------	---------------------

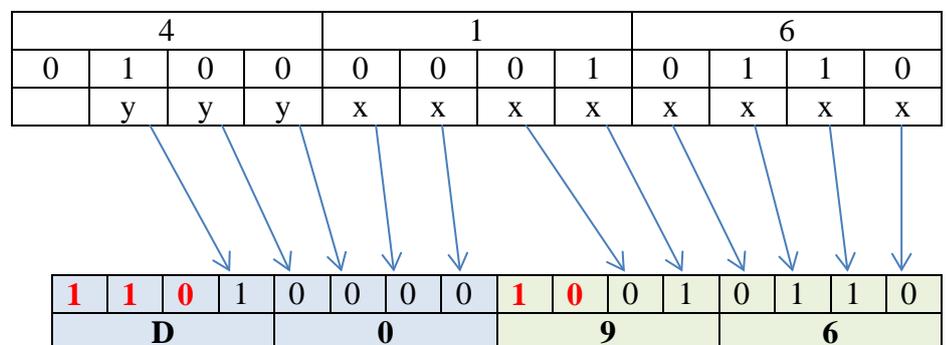
3. Obtener la codificación UTF-8 del carácter 'Ж' (CYRILLIC CAPITAL LETTER ZHE), con código 0x0416.

a. Identificamos el rango:

Rango UTF-32	Formato binario	UTF-8	
0x00 a 0x7F	xxxxxx₂	0xxxxxx₂	ASCII
0x80 a 0x7FF	yyyxxxxxxxx ₂	110yyyxx ₂ 10xxxxxx ₂	Caracteres occidentales
0x800 a 0xFFFF	yyyyyyyyxxxxxxxx ₂	1110yyyy ₂ 10yyyyxx ₂ 10xxxxxx ₂	Caracteres orientales
0x1000 a 0x10FFFF	zzzzzyyyyyyyxxxxxxxx ₂	11110zzz ₂ 10zzyyyy ₂ 10yyyyxx ₂ 10xxxxxx ₂	Caracteres inusuales

1 1 0 y y y x x 1 0 x x x x x x

b. Pasamos a binario y rellenamos: 0x00000416



c. Obtenemos el valor en hexadecimal: 0xD096

U+0416	Ж	0xd0 0x96	CYRILLIC CAPITAL LETTER ZHE
--------	---	-----------	-----------------------------

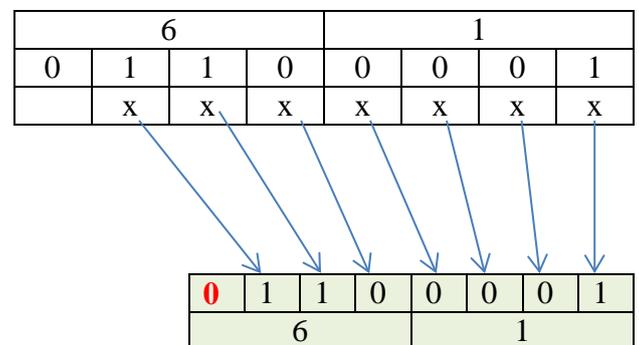
4. Obtener la codificación UTF-8 del carácter 'a' (LATIN SMALL LETTER A), con código 0x0061.

a. Identificamos el rango:

Rango UTF-32	Formato binario	UTF-8	
0x00 a 0x7F	xxxxxxx ₂	0xxxxxxx ₂	ASCII
0x80 a 0x7FF	yyyyzzzzzzzz₂	110yyyyz₂ 10xxxxxx ₂	Caracteres occidentales
0x800 a 0xFFFF	yyyyyyyyxxxxxxxx ₂	1110yyyy ₂ 10yyyyxx ₂ 10xxxxxx ₂	Caracteres orientales
0x1000 a 0x10FFFF	zzzzzyyyyyyyyyxxxxxxxx ₂	11110zzz ₂ 10zzyyyy ₂ 10yyyyxx ₂ 10xxxxxx ₂	Caracteres inusuales

0	x	x	x	x	x	x	x
---	---	---	---	---	---	---	---

b. Pasamos a binario y rellenamos: 0x00000061



c. Obtenemos el valor en hexadecimal: 0x61

U+0061	a	0x61	LATIN SMALL LETTER A
--------	---	------	----------------------