

Codificación y decodificación de caracteres

ANSI (8 bits) que sustituye a ASCII (7 bits)

The screenshot shows the Immunity Debugger's memory dump window. The file being analyzed is 'ANSI 8 bits.txt'. The dump displays the text 'HOLA MUNDO..hola mundo..0123456789' in ASCII format. The address range shown is 00000000 to 00000020. The dump is configured with 'offset 0', 'absolute' addressing, and 'Ascll' encoding.

address	00	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	Ascll	unsigned	bigendian
00000000	48	4f	4c	41	20	4d	55	4e	44	4f	0d	0a	68	6f	6c	61	HOLA MUNDO..hola		
00000010	20	6d	75	6e	64	6f	0d	0a	30	31	32	33	34	35	36	37	mundo..01234567		
00000020	38	39															89		

UTF-8 (Plano básico multilingüe)

<https://www.charset.org/utf-8>

Google Traductor

Texto

Documentos

Sitios web

DETECTAR IDIOMA

CHINO

ESPAÑOL

INGLÉS



CHINO (TRADICIONAL)


hola mundo



你好世界

Los números en chino del 1 al 10

Número	Chino	Pinyin
0	零 / 〇	líng
1	一	yī
2	二	èr
3	三	sān
4	四	sì
5	五	wǔ
6	六	liù
7	七	qī
8	八	bā
9	九	jiǔ
10	十	shí

 Iniciar

≡ ANSI 8 bits.txt ●

≡ ANSI 8 bits.txt

≡ UTF-8.txt ●

≡ UTF-8.txt

≡ UTF-8.txt

1

你好世界

2

零一二三四五六七八九十

Iniciar

ANSI 8 bits.txt

ANSI 8 bits.txt

UTF-8.txt

UTF-8.txt

offset

0

absolute

relative

address	00	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	UTF8	<input type="checkbox"/> unsigned	<input type="checkbox"/> bigendian
00000000	e4	bd	a0	e5	a5	bd	e4	b8	96	e7	95	8c	0d	0a	e9	9b	你好世界..零		
00000010	b6	e4	b8	80	e4	ba	8c	e4	b8	89	e5	9b	9b	e4	ba	94	.一二三四五		
00000020	e5	85	ad	e4	b8	83	e5	85	ab	e4	b9	9d	e5	8d	81	20	六七八九十		

E4	1110 0100
	Implica que son tres bytes
E4	1110 0100
BD	10 11 1101
A0	10 10 0000
0100111101100000 - 0X4F60	你

E5	1110 0101
	Implica que son tres bytes
E5	1110 0101
A5	10 10 0101
BD	10 11 1101
0101 1001 0111 1101 - 0X597D	钹

E4	1110 0100
	Implica que son tres bytes
E4	1110 0100
B8	10 11 1000
96	1001 0110
0100 1110 0001 0110 - 0X4E16	世


E7	1110 0111
	Implica que son tres bytes
E7	1110 0111
95	1001 0101
86	1000 1100
0111 0101 0100 1100- 0X754C	界


Unicode	Carácter	Byte 1	Byte 2	Byte 3	Byte 4
U+4F60	你	1110 0100	10 11 1101	10 10 0000	
U+597D	钹	1110 0101	10 10 0101	10 11 1101	
U+4E16	世	1110 0100	10 11 1000	10 01 0110	
U+754C	界	1110 0111	10 01 0101	10 00 1100	

UTF-8 Resto de planos (pares subrogados) - 010000-10FFFF

Las direcciones van entre 0x 0000

Dec	Hex	UTF-8 Hex	Char	Unicode description
74773	U+12415	F0 92 90 95	𐎧	Cuneiform Numeric Sign One Gesh2
74767	U+1240F	F0 92 90 8F	𐎦	Cuneiform Numeric Sign Four U

F0	1111 0 000
	Implica que son cuatro bytes
F0	1111 0 000
92	1001 0010
90	1001 0000
95	1001 0101
0 0001 0010 0100 0001 0101 0x 12415	 Cuneiform Numeric Sign One Gesh2

F0	1111 0 000
92	1001 0010
90	1001 0000
8f	1000 1111
0 0001 0010 0100 0000 1111 0x 1240F	 Cuneiform Numeric Sign Four U