Representación de Datos Alfanuméricos

Elementos de Informática

Representación de Datos

Las limitaciones de la electrónica convencional imponen una sola forma práctica de representación: la lógica de dos estados "0" y "1".

La "representación binaria de datos" significa representar datos por medio de estos ceros y unos, denominados "dígitos binarios" o "bits".

El bit es la unidad mínima de información. El problema a resolver es como los bits pueden representar los datos de una manera útil, y como recuperar posteriormente esos datos.

Representación de Datos

Si vamos a representar datos de forma normal y que resulte comprensible, necesitaremos grupos de bits que identifiquen al menos.

- los números del 0 al 9 (10 combinaciones de bits)
- las letras del alfabeto (26 combinaciones de bits)
- otros símbolos (\$ # & * @ % ^ < =)

Representación de Datos

Cantidad de bits = x	2 ^x	
1	2	0 y 1
2	4	00, 01, 10, 11
3	8	000, 001, 010, 011, 100, 101, 110, 111
4	16	0000, 0001, 0010, 0011, 0100, 0101, 0110, 0111,
5	32	•••••
6	64	•••••

Utilizando grupos de 7 bits se obtienen $2^7 = 128$ combinaciones posibles de bits que nos permiten representar 10 números, letras y símbolos.

Codificación y Decodificación

Los datos que ingresan a una computadora deben ser convertidos a una "representación interna" para que el computador pueda procesarlos. Después de que este proceso es llevado a cabo, los resultados obtenidos son convertidos nuevamente a una "representación externa" para que el usuario reciba datos provistos de significado.

Todo este proceso implica una *codificación* y posterior *decodificación* de la información.

Codificación

Codificar es aplicar un conjunto de reglas inequívocas que especifiquen la forma en que pueden representarse los datos, de forma tal que sea posible la decodificación ulterior.

La elección de un sistema de codificación adecuado es importante, ya que de esto dependerá la fiabilidad, universalidad y velocidad del sistema.

Códigos

Existen en la actualidad una infinidad de códigos, algunos que son propios de una máquina o marca específica y otros que son universalmente aceptados.

Algunos ejemplos de estos códigos internacionales son: BCD, EBCDIC, ASCII, etc.

Códigos

Normalmente, en el almacenamiento de una computadora, se pueden almacenar dos clases de datos:

- ✓ Alfanuméricos (letras, caracteres y nros.)
- ✓ Datos numéricos.

La mayoría de las computadoras se manejan con códigos de 8 bits para la representación de un carácter.

Representación de Datos Alfanuméricos

En informática hay dos códigos de uso general para representación de caracteres:

<u>ASCII</u>: American Standard Code for Information Interchange, de uso universal en microprocesadores.

EBCDIC: Extended Binary Coded Decimal Interchange Code; que usan principalmente algunos modelos de "IBM".

Esquema de Representación

Esquema de representación de un byte en ambos códigos:

La representación de cada carácter se divide en dos partes, una de zona a la izquierda y una numérica a la derecha.

8 bits

zona numéricos

ASCII - Un Poco de Historia

El código ASCII proviene de las siglas en ingles para American Standard Code for Information Interchange, es decir Código Americano (estadounidense) Estándar para el intercambio de Información. (se pronuncia Aski).

Fue creado en 1963 por el Comité Estadounidense de Estándares o "ASA", organismo que desde 1969 se llama "Instituto Estadounidense de Estándares Nacionales" o "ANSI" como se lo conoce desde entonces.

Este código nació a partir de reordenar y expandir el conjunto de símbolos y caracteres ya utilizados por ese entonces en telegrafía por la compañía Bell.

ASCII - Un Poco de Historia

En un primer momento solo incluía las letras mayúsculas, pero en 1967 se agregaron las letras minúsculas y algunos caracteres de control, formando así lo que se conoce como US-ASCII, es decir los códigos del 0 al 127.

Así con este conjunto de solo 128 caracteres fue publicado en 1967 como estándar, conteniendo todos lo necesario para escribir en idioma ingles.

ASCII - 7

Este código utiliza 7 bits que proporciona $2^7 = 128$ combinaciones diferentes.

El octavo bit o bit mas significativo, es decir el de la izquierda, es el bit de paridad; utilizado para controlar si la transmisión del byte tiene errores.



Bit de Paridad

El bit de paridad se utiliza como un método de detección de errores.

La paridad puede ser par o impar. Esto implica que la cantidad de 1's en el paquete de datos, para nuestro caso el byte, debe tener una cantidad par o impar de unos, según la paridad elegida.

Generalmente se utiliza paridad par. Ejemplos:





Bit de Paridad

ASCII - 7

- Cada dígito (0 al 9) tiene 011 como la parte de zona del byte y su representación binaria como la parte numérica.

Número decimal	Zona	Numérica
0	011	0000
1	011	0001
2	011	0010
3	011	0011
4	011	0100
5	011	0101
6	011	0110
7	011	0111
8	011	1000
9	011	1001

ASCII - 7

Los caracteres alfabéticos (mayúsculas) están distribuidos en dos grupos según si sus bytes tienen 100 o 101 como la parte de zona.

Las tablas de códigos nos muestran las equivalencias entre los caracteres y su representación binaria, decimal, hexadecimal u octal.

Por lo tanto debemos plantear una manera de poder encontrar representaciones equivalentes a partir de la representación mostrada en la tabla. Bits menos significativos

Tabla del Código ASCII 7 bits

Representación Decimal Zona Bits más significativos

	(000	001	010		011		100		101		111	0	11	1
		0	1		2		3		4		5		6		- 7
0000	0	NUL 0	DLE 16	SP	32	0	48	@	64	P	80		96	p	112
0001	1	SOH 1	DCl 17	!	33	1	49	A	65	Q	81	a	97	q	113
0010	2	STX 2	DC2 18	et	34	2	50	В	66	R	82	Ъ	98	r	114
0011	3	ETX 3	DC3	#	35	3	51	С	67	S	83	С	99	s	115
0100	4	EOT 4	DC4 20	\$	36	4	52	D	68	T	84	d	100	t	116
0101	5	ENQ 5	NAK 21	%	37	5	53	E	69	U	85	8	101	u	117
0110	6	ACK 6	SYN 22	&	38	б	54	F	70	V	36	f	102	v	118
0111	7	BEL 7	ETB 23	,	39	7	55	G	71	W	87	g	103	w	119
1000	8	BS 8	CAN 24	(40	8	56	H	72	X	88	h	104	x	120
1001	9	HT 9	EM 25)	41	9	57	I	73	Y	89	i	105	у	121
1010	A	LF 10	SUB 26	sk	42	:	58	J	74	Z	90	j	106	z	122
1011	В	VT 11	ESC 27	+	43	;	59	K	75	[91	k	107	{	123
1100	С	FF 12	FS 28		44	<	60	L	76	1	92	1	108	1	124
1101	D	CR 13	GS 29	-	45	=	61	M	77]	93	m	109	}	125
1110	E	SO 14	RS 30	÷	46	>	62	N	78	^	94	n	110	~	126
1111	F	SI 15	US 31	1	47	?	63	0	79	-	95	0	111	DE	

ASCII Hay Simbola	ASCII Hay Simbola	ASCII Hay Simbola	ASCII Hay Simbola
ASCII Hex Símbolo 0 0 NUL 1 1 SOH 2 2 STX 3 3 ETX 4 4 EOT 5 5 ENQ 6 6 ACK 7 7 BEL 8 8 BS 9 9 TAB 10 A LF 11 B VT 12 C FF 13 D CR 14 E SO	ASCII Hex Simbolo 16 10 DLE 17 11 DC1 18 12 DC2 19 13 DC3 20 14 DC4 21 15 NAK 22 16 SYN 23 17 ETB 24 18 CAN 25 19 EM 26 1A SUB 27 1B ESC 28 1C FS 29 1D GS 30 1E RS	ASCII Hex Símbolo 32 20 (space) 33 21 ! 34 22 " 35 23 # 36 24 \$ 37 25 % 38 26 & 39 27 ' 40 28 (41 29) 42 2A * 43 2B + 44 2C , 45 2D - 46 2E	ASCII Hex Símbolo 48 30 0 49 31 1 50 32 2 51 33 3 52 34 4 53 35 5 54 36 6 55 37 7 56 38 8 57 39 9 58 3A : 59 3B ; 60 3C < 61 3D = 62 3E >
ASCII Hex Símbolo 64 40 @ 65 41 A 66 42 B 67 43 C 68 44 D 69 45 E 70 46 F 71 47 G 72 48 H 73 49 I 74 4A J 75 4B K 76 4C L 77 4D M 78 4E N 79 4F O	31 1F US ASCII Hex Simbolo	47 2F / ASCII Hex Símbolo 96 60 . 97 61 a 98 62 b 99 63 c 100 64 d 101 65 e 102 66 f 103 67 g 104 68 h 105 69 i 106 6A j 107 6B k 108 6C I 109 6D m 110 6E n 111 6F o	63 3F ? ASCII Hex Simbolo 112 70 p 113 71 q 114 72 r 115 73 s 116 74 t 117 75 u 118 76 v 119 77 w 120 78 x 121 79 y 122 7A z 123 7B { 124 7C 125 7D } 126 7E ~ 127 7F

		000	001	010		01	1	100)	10	1	11	0	11	1
		0	1		2		3		4		5	1	6		- 7
0000	0	NUL 0	DLE 16	SP	32	0	48	@	64	P	80		96	p	112
0001	1	SOH 1	DCl 17	!	33	1	49	A	65	Q	81	a	97	q	113
0010	2	STX 2	DC2 18	ce	34	2	50	В	00	R	82	р	98	r	114
0011	3	ETX 3	DC3 19	#	35	3	51	С	67	S	83	c	99	s	115
0100	4	EOT 4	DC4 20	\$	36	4	52	D	68	T	84	ď	100	t	116
0101	5	ENQ 5	NAK 21	%	37	5	53	E	69	U	85	8	101	u	117
0110	6	ACK 6	SYN 22	&	38	б	54	F	70	V	86	f	102	v	118
0111	7	BEL 7	ETB 23	,	39	7	55	G	71	W	87	g	103	w	119
1000	8	BS 8	CAN 24	(40	8	56	Н	72	X	88	h	104	x	120
1001	9	HT 9	EM 25)	41	9	57	I	73	Y	89	i	105	у	121
1010	A	LF 10	SUB 26	*	42	:	58	J	74	Z	90)	106	z	122
1011	В	VT 11	ESC 27	+	43	;	59	K	75	1	91	k	107	{	123
1100	С	FF 12	FS 28		44	٧	60	L	76	1	92	1	108	1	124
1101	D	CR 13	GS 29	-	45	=	61	M	77]	93	m	109	}	125
1110	E	SO 14	RS 30		46	>	62	N	78	^	94	n	110	~	126
1111	F	SI 15	US 31	1	47	?	63	0	79	-	95	0	111	DE	

Tabla del código ASCII 7 bits

		000	001	010		011		100	1	101		11	0	11	1
		0	1	_	2		3		4		5	_	6		7
100	0	NUL 0	DLE 16	SP	32	0	48	@	64	P	80		96	p	112
101	1	SOH 1	DCl 17	!	33	1	49	A	65	Q	81	a	97	q	113
110	2	STX 2	DC2 18	cc	34	2	50	В	66	R	82	Ъ	98	r	114
11	3	ETX 3	DC3	#	35	3	51	С	67	S	83	С	99	s	115
00	4	EOT 4	DC4 20	\$	36	4	52	D	68	T	84	ď	100	t	116
01	5	ENQ 5	NAK 21	%	37	5	53	E	69	U	85	8	101	u	117
10	6	25	24	0		4		7		1		0	2	v	118
	_	Z°	Δ.		<u></u>		7	_	4	1	4	4	3	w	119
6	4	32	16)	8		4		2		1	1		x	
0.1	-	TITE	123.4		17	0		*	7.00	37			1		120
E	54	0	16	5	0		0		0)	5_	8	0
	Α	10	26		42		58		74		90	1	106		122
	T.							Tr	-	1		k		{	123
11	В	VT 11	ESC 27	+	43	;	59	K	75	•	91		107	8	
11		VT 11		+	43 O	;	59		75)	91	1	107	6	5
0 0	В	VT 11 0	27 0	+		;		***	75 77)	91 93	1	107		
11	в 64	VT 11 0	0		0	>	C))		1			5

	111-64= 47
5964=1 45	47-32= 15
5 13 20=13	15-8= 7
3-8= 5	7-4=3
4= 1	3-2= 1

65964 = 1 45
4 5 1 3€13
13-8= 5
5-4= 1

		000	001	010)	01	1	100	0	101	1	110	11	1
		0	1		2		3		4		5	1 6		7
0000	0	NUL 0	DLE 16	SP	32	0	48	@	64	P	80	96	p	112
0001	1	SOH 1	DCl 17	!	33	1	49	A	65	Q	81	a 97	q	113
0010	2	STX 2	DC2 18	64	34	2	50	В	66	R	82	b 98	r	114
0011	3	ETX 3	DC3 19	#	35	3	51	С	67	S	83	c 99	s	115
0100	4	EOT 4	DC4 20	\$	36	4	52	D	68	T	84	d 100	t	116
0101	5	ENQ 5	NAK 21	%	37	5	53	E	69	U	85	e 101	u	117
0110	6	ACK 6	SYN 22	&	38	б	54	F	70	V	86	f 102	v	118
0111	7	BEL 7	ETB 23	,	39	7	55	G	71	W	87	g 103	w	119
1000	8	BS 8	CAN 24	(40	8	56	Н	72	X	88	h 104	x	
1001	9	HT 9	EM 25)	41	9	57	I	73	Y	89	i 105	у	121
1010	A	LF 10	SUB 26	n	42	:	58	J	74	Z	90	j 106	z	122
1011	В	VT 11	ESC 27	+	43	;	59	K	75	1	91	k 107	{	123
1100	С	FF 12	FS 28		44	<	60	L	76	1	92	1 108	1	124
1101	D	CR 13	GS 29	-	45	=	61	М	77]	93	m 109	}	125
1110	E	SO 14	RS 30	s.	46	>	62	N	78	^	94	n 110	~	
1111	F	SI 15	US 31	1	47	?	63	0	79	-	95	0	DI	EL 127

Tabla del código ASCII 7 bits

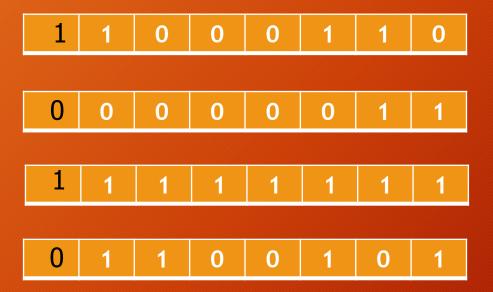
k Símbolo	ASCI	Hex	Símbolo	ASCII	Hex	Símbolo	ASCII	Hex	Símbolo
NUL SOH STX ETX EOT ENQ ACK BEL BS TAB LF VT FF	16 17 18 19 20 21 22 23 24 25 26 27 28	10 11 12 13 14 15 16 17 18 19 1A 1B 1C	DLE DC1 DC2 DC3 DC4 NAK SYN ETB CAN EM SUB ESC FS GS	32 33 34 35 36 37 38 39 40 41 42 43 44 45	20 21 22 23 24 25 26 27 28 29 2A 2B 2C 2D	(space) ! # \$ % & ' () *	48 49 50 51 52 53 54 55 56 57 58 59 60 61	30 31 32 33 34 35 36 37 38 39 3A 3B 3C	0 1 2 3 4 5 6 7 8 9
1	0		0	0		0	0		1
x Símbolo	ASCI	Hex	Símbolo	ASCII	Hex	Símbolo	ASCII	Hex	Símbolo
x Símbolo	ASCII	l Hex	Símbolo	ASCII	Hex	Símbolo 1	ASCII 0	Hex	Símbolo
		52 53	O K S T	1 98 99	62 63	Símbolo 1	O 114 115	/2 73	1 r
1 B C	82 83	52 53	0 K S	1 98 99	62 63	1 c	0 114 115	12 73	1
C C	82 83	52 53	0 K S	1 98 99	62 63	1 0 0	0 114 115	12 73	1
1	1 82 83 84 1	52 53 54	0 8 9 0 x	1 99 100 1 104	63	1 1 h	1114 115 146 1 120	73 74	1 s 1

A m 0

Ob ca re

Ejercicio

Completa con el bit de paridad, teniendo en cuenta la paridad par en los siguientes bytes.



Existe también el código ASCII - 8; este código utiliza los 8 bits para almacenar un carácter y el computador almacena un bit extra o noveno bit llamado bit de paridad.

En 1981, la empresa IBM desarrolló una extensión de 8 bits del código ASCII, en esta versión se reemplazaron algunos caracteres de control obsoletos.

Se incorporaron 128 caracteres nuevos, tales como símbolos, signos, gráficos adicionales y letras latinas, necesarias para la escrituras de textos en otros idiomas, como por ejemplo el español. Así fue como se agregaron los caracteres que van del ASCII 128 al 255.

Ca		res ASCII de ontrol
00	NULL	(carácter nulo)
01	SOH	(inicio encabezado)
02	STX	(inicio texto)
03	ETX	(fin de texto)
04	EOT	(fin transmisión)
05	ENQ	(consulta)
06	ACK	(reconocimiento)
07	BEL	(timbre)
08	BS	(retroceso)
09	HT	(tab horizontal)
10	LF	(nueva línea)
11	VT	(tab vertical)
12	FF	(nueva página)
13	CR	(retorno de carro)
14	SO	(desplaza afuera)
15	SI	(desplaza adentro)
16	DLE	(esc.vínculo datos)
17	DC1	(control disp. 1)
18	DC2	(control disp. 2)
19	DC3	(control disp. 3)
20	DC4	(control disp. 4)
21	NAK	(conf. negativa)
22	SYN	(inactividad sínc)
23	ETB	(fin bloque trans)
24	CAN	(cancelar)
25	EM	(fin del medio)
26	SUB	(sustitución)
27	ESC	(escape)
28	FS	(sep. archivos)
29	GS	(sep. grupos)
30	RS	(sep. registros)
31	US	(sep. unidades)
127	DEL	(suprimir)

			res AS mibles		
32	espacio	64	@	96	•
33		65	Ä	97	а
34		66	В	98	b
35	#	67	С	99	С
36	\$	68	D	100	d
37	%	69	E	101	е
38	&	70	F	102	f
39		71	G	103	g
40	(72	Н	104	h
41)	73	- 1	105	i
42	*	74	J	106	j
43	+	75	K	107	k
44	,	76	L	108	- 1
45	-	77	M	109	m
46		78	N	110	n
47	1	79	0	111	0
48	0	80	Р	112	р
49	1	81	Q	113	q
50	2	82	R	114	r
51	3	83	S	115	s
52	4	84	Т	116	t
53	5	85	U	117	u
54	6	86	٧	118	٧
55	7	87	W	119	w
56	8	88	Х	120	х
57	9	89	Υ	121	У
58	:	90	Z	122	z
59	;	91	[123	{
60	<	92	1	124	- 1
61	=	93]	125	}
62	>	94	^	126	~
63	?	95	-		

ASCII extendido (Página de código 437)							
128	Ç	160	á	192	L	224	Ó
129	ű	161	í	193		225	ß
130	é	162	ó	194	-	226	ô
131	â	163	ú	195	Ţ	227	Ò
132	ä	164	ñ	196		228	Ö
133	à	165	Ñ	197	+	229	Õ
134	å	166	a	198	ä	230	μ
135	ç	167	•	199	Ã	231	þ
136	ê	168	ż	200	L	232	Þ
137	ë	169	®	201	1	233	Ú
138	è	170	7	202	T	234	Û
139	ï	171	1/2	203	ΤF	235	Ù
140	î	172	1/4	204	F	236	Ý
141	ì	173	i	205	=	237	Ý
142	Ä	174	«	206	#	238	-
143	Α	175	>>	207	п	239	
144	É	176	#	208	õ	240	≡
145	æ	177		209	Ð	241	±
146	Æ	178		210	Ê	242	_
147	ô	179		211	Ë	243	3√4
148	Ö	180	-{	212	È	244	¶
149	Ò	181	Á	213	ļ.	245	§
150	û	182	Â	214	ĺ	246	÷
151	ù	183	À	215	Î	247	
152	ÿ	184	©	216	Ţ	248	۰
153	Ö	185	1	217	J	249	
154	Ü	186		218	Т	250	
155	Ø	187]	219		251	1
156	£	188		220		252	3
157	Ø	189	¢	221		253	2
158	×	190	¥	222	į	254	
159	f	191	7	223	-	255	nbsp

Byte	Cod.	Char									
10000000	128	C	10100000	160	á	11000000	192	+	11100000	224	Ó
10000001	129	ü	10100001	161	i	11000001	193	-	11100001	225	ß
10000010	130	é	10100010	162	6	11000010	194	4	11100010	226	Ô
10000011	131	â	10100011	163	ú	11000011	195	+	11100011	227	Ò
10000100	132	ä	10100100	164	ñ	11000100	196	+	11100100	228	õ
10000101	133	à	10100101	165	Ñ	11000101	197	+	11100101	229	ŏ
10000110	134	å	10100110	166	a	11000110	198	ä	11100110	230	ш
10000111	135	c	10100111	167		11000111	199	Ä	11100111	231	b
10001000	136	ê	10101000	168	2	11001000	200	+	11101000	232	D
10001001	137	ě	10101001	169	00	11001001	201	+	11101001	233	Ü
10001010	138	è	10101010	170	-	11001010	202	21	11101010	234	Û
10001011	139	ĭ	10101011	171	3/2	11001011	203	-	11101011	235	Ü
10001100	140	î	10101100	172	3/4	11001100	204	1	11101100	236	ý
10001101	141	1	10101101	173	1	11001101	205		11101101	237	Y
10001110	142	Ã	10101110	174	*	11001110	206	+	11101110	238	-
10001111	143	Å	10101111	175	30	11001111	207	0	11101111	239	
10010000	144	Ė	10110000	176	- 22	11010000	208	ð	11110000	240	
10010001	145	æ	10110001	177	-	11010001	209	Đ	11110001	241	±
10010010	146	Æ	10110010	178	-	11010010	210	Ê	11110010	242	1000
10010011	147	ô	10110011	179	7	11010011	211	Ë	11110011	243	3/4
10010100	148	ö	10110100	180	1	11010100	212	E	11110100	244	4
10010101	149	ò	10110101	181	À	11010101	213	i	11110101	245	8
10010110	150	û	10110110	182	Â	11010110	214	I	11110110	246	+
10010111	151	ù	10110111	183	Â	11010111	215	Î	11110111	247	
10011000	152	ÿ	10111000	184	0	11011000	216	Ĩ	11111000	248	0
10011001	153	ŏ	10111001	185	1	11011001	217	+	11111001	249	-19
10011010	154	Ü	10111010	186	1	11011010	218	+	11111010	250	-
10011011	155	a	10111011	187	+	11011011	219	2000	11111011	251	1
10011100	156	£	10111100	188	+	11011100	220		11111100	252	3
10011101	157	Ø	10111101	189	e	11011101	221	7	11111101	253	2
10011110	158	×	10111110	190	¥	11011110	222	İ	11111110	254	
10011111	159	f	10111111	191	+	11011111	223		11111111	255	-

Ejercicio

Convierte a hexadecimal el siguiente mensaje y luego tradúcelo al lenguaje español:

Carácter	ASCII	Carácter	ASCII	
Α	0100 0001	W	0101 0111	
В	0100 0010	X	0101 1000	
C	0100 0011	Υ	0101 1001	
D	0100 0100	Z	0101 1010	
E	0100 0101	0	0011 0000	
F	0100 0110	1	0011 0001	
G	0100 0111	2	0011 0010	
Н	0100 1000	3	0011 0011	
T	0100 1001	4	0011 0100	
J	0100 1010	5	0011 0101	
K	0100 1011	6	0011 0110	
L	0100 1100	7	0011 0111	
M	0100 1101	8	0011 1000	
N	0100 1110	9	0011 1001	
0	0100 1111	+	0010 1011	
Р	0101 0000	-	0010 1101	
Q	0101 0001	*	0010 1010	
R	0101 0010	. 13	0011 1010	
S	0101 0011	=	0011 1101	
Т	0101 0100	<	0011 1100	
U	0101 0101	;	0011 1011	
٧	0101 0110	411	100000000000000000000000000000000000000	

EBCDIC

El código ampliado decimal codificado en binario es un código de 8 bits utilizado principalmente por los equipos "IBM" y compatibles.

Se forma:

- Cada **dígito numérico** tiene **1111** como la parte de zona del byte y su representación binaria, como la parte numérica.
- Los **caracteres alfabéticos** (mayúsculas) están distribuidos en tres grupos, ya sea que los bytes tengan 1100; 1101 o 1110 como bits de zona.

EBCDIC

Ejemplos:

Carácter	zona	numérica
A	1100	0001
В	1100	0010
1	1111	0001
2	1111	0010

La codificación se comprende observando el siguiente diagrama.

EBCDIC

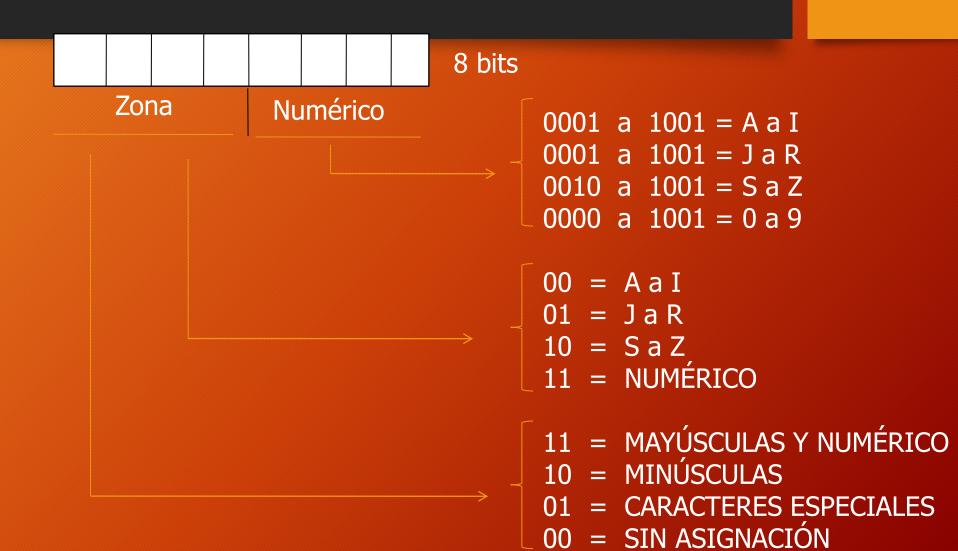
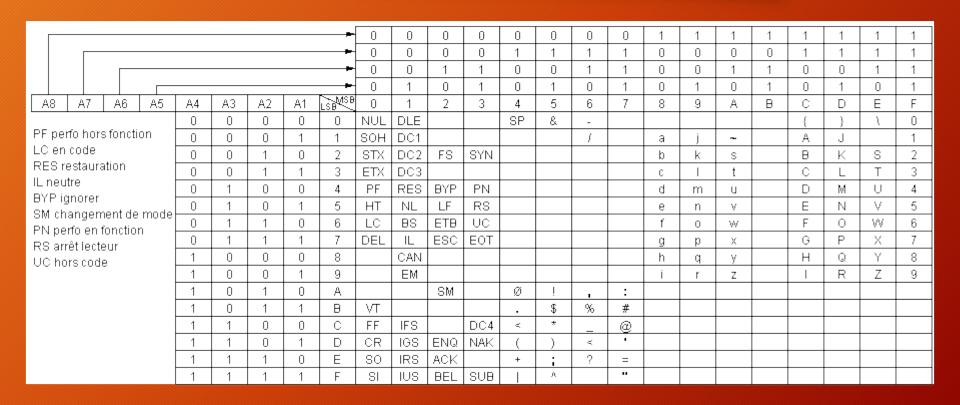


Tabla EBCDIC



Ejercicio

Escribe el día de la semana que es hoy utilizando el código EBCDIC