

# Representación interna da información

#### Bit

**Bit** (contracción de díxito binario, "Bl/nary digi/T" en inglés)

É a menor unidade de información.

Un bit ten un único valor, 0 ou 1, tamén interpretado como verdadeiro ou falso.

El **bit** es la unidad mínima de información; con él podemos representar dos valores cualesquiera, como verdadero o falso, abierto o cerrado, blanco o negro, norte o sur, rojo o azul... Basta con asignar uno de esos valores al estado de «apagado» (0) y el otro al estado de «encendido» (1).

## Nibble / Byte

#### Nibble («mordisco») = 4 bits

Un nibble pode representar un díxito hexadecimal.

#### Byte ou octeto («bite», bocado) = 8 bits

Tamén se lle chama «carácter» porque con 8 bits pódense representar tódalas letras. (verémolo máis adiante).

# Byte / Palabra

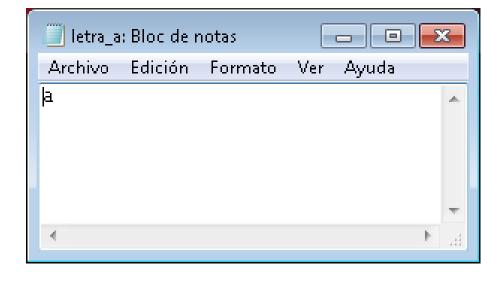
Cuando se almacena la información no se trabaja a nivel de bit, sino que se trabaja a nivel de carácter (letra, número o signo de puntuación), que ocupa lo que se denomina un **byte**, que a su vez está compuesto de 8 **bits**. El ordenador trabaja con agrupaciones de bits fáciles de manipular y suelen ser múltiplos de 2, la base del sistema binario. Los tamaños más comunes son:

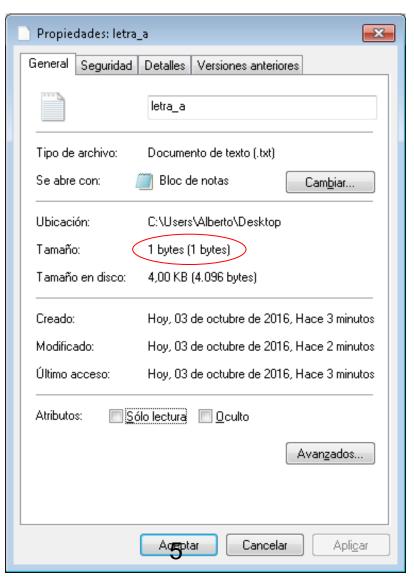
- Octeto, carácter o byte: es la agrupación de 8 bits, el tamaño típico de información; con él se puede codificar el alfabeto completo (ASCII estándar).
- Palabra: tamaño de información manejada en paralelo por los componentes del sistema, como la memoria, los registros o los buses. Son comunes las palabras de 8, 32, 64, 128 y 256 bits: 1 byte, 4, 8, 16, 32 bytes. A mayor tamaño de palabra, mayor es la precisión y la potencia de cálculo del ordenador.

Así, cuando decimos que un archivo de texto ocupa 5 000 bytes, queremos decir que contiene el equivalente a 5 000 letras o caracteres (entre dos y tres páginas de texto sin formato).

## Exemplo 1 byte = 1 caracter

Crea un ficheiro de texto coa letra «a». Comproba o tamaño que ocupa.





# Múltiplos de byte

Lo normal es utilizar los múltiplos del byte: el kilobyte (KB), el megabyte (MB), el gigabyte (GB), etc.

En informática se utilizan las potencias de 2 (23, 210, 220...) para representar las medidas de la información; sin embargo se ha extendido el uso de las potencias de 10 (uso decimal), debido a que se ha impuesto el uso del *Sistema Internacional de Medidas* (SI), o sistema métrico. Así pues, el primer término de medida que se utilizó fue el **kilobyte** (KB), y se eligió este porque 210 es aproximadamente 1000, que se asocia con el kilo (1000 gramos); en realidad debería ser 1024 bytes, ya que 210 son 1024.

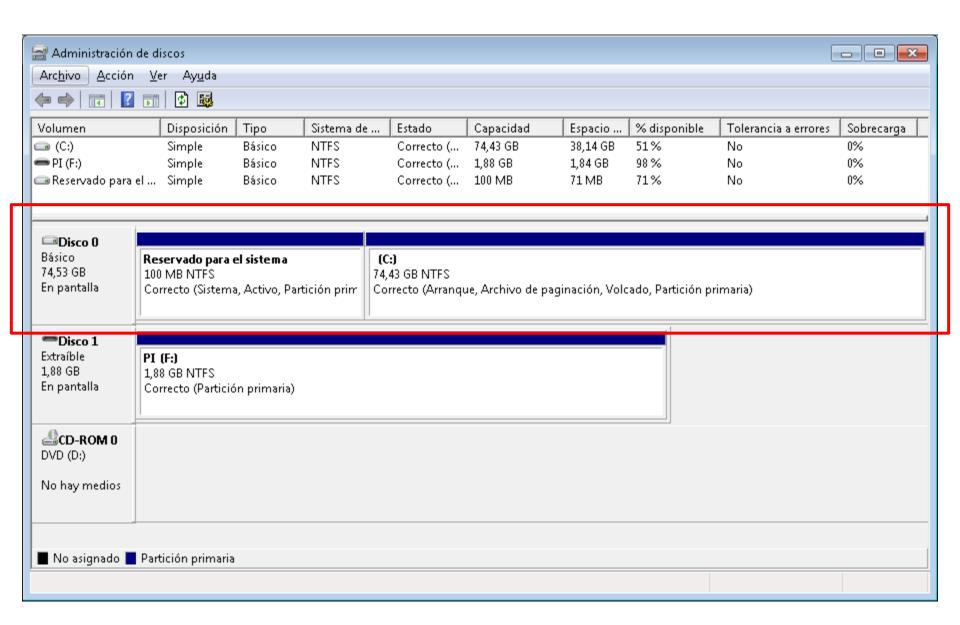
Nombre (símbolo)	Sistema Internacional de Unidades (SI) Estándar (uso decimal)	Prefijo binario (uso binario)	Nombre (símbolo)	
Kilobyte (KB)	$1000^1 = 10^3 \text{ bytes}$	$1024^1 = 2^{10}$ bytes	Kibibyte (kib)	
Megabyte (MB)	1000 <sup>2</sup> = 10 <sup>6</sup> bytes	$1024^2 = 2^{20}$ bytes	Mebibyte (Mib)	
Gigabyte (GB)	1000 <sup>3</sup> = 10 <sup>9</sup> bytes	$1024^3 = 2^{30}$ bytes	Gibibyte (Gib)	
Terabyte (TB)	$1000^4 = 10^{12}$ bytes	1024 <sup>4</sup> = 2 <sup>40</sup> bytes	Tebibyte(Tib)	
Petabyte (PB)	$1000^5 = 10^{15}$ bytes	$1024^5 = 2^{50}$ bytes	Pebibyte (Pib)	
Exabyte (EB)	1000 <sup>6</sup> = 10 <sup>18</sup> bytes	1024 <sup>6</sup> = 2 <sup>60</sup> bytes	Exbibyte (Eib)	
Zettabyte (ZB)	$1000^7 = 10^{21}$ bytes	$1024^7 = 2^{70}$ bytes	Zebibyte (Zib)	
Yottabyte (YB)	1000 <sup>8</sup> = 10 <sup>24</sup> bytes	1024 <sup>8</sup> = 2 <sup>80</sup> bytes	Yobibyte (Yib)	

Tabla 1.15. Unidades de medida de información en decimal y en binario.

Unidades de información (del byte)							
Sistema Internaciona (decimal)	al	ISO/IEC 80000-13 (binario)					
Múltiplo SI (símbolo)		Múltiplo (símbolo)	ISO/IEC				
kilobyte (kB)	10 <sup>3</sup>	kibibyte (KiB)	2 <sup>10</sup>				
megabyte (MB)	10 <sup>6</sup>	mebibyte (MiB)	2 <sup>20</sup>				
gigabyte (GB)	10 <sup>9</sup>	gibibyte (GiB)	2 <sup>30</sup>				
terabyte (TB)	10 <sup>12</sup>	tebibyte (TiB)	2 <sup>40</sup>				
petabyte (PB)	10 <sup>15</sup>	pebibyte (PiB)	2 <sup>50</sup>				
exabyte (EB)	10 <sup>18</sup>	exbibyte (EiB)	2 <sup>60</sup>				
zettabyte (ZB)	10 <sup>21</sup>	zebibyte (ZiB)	2 <sup>70</sup>				
yottabyte (YB)	10 <sup>24</sup>	yobibyte (YiB)	2808				

Número de bytes	Múltiplo	Equivalencia aproximada
1	1 B	Una letra.
10	10 B	Una o dos palabras.
100	100 B	Una o dos frases.
1000	1 kB	Una historia muy corta.
10 000	10 kB	Una página de enciclopedia, tal vez con un dibujo simple.
100 000	100 kB	Una fotografía de resolución mediana.
1 000 000	1 MB	Una novela.
10 000 000	10 MB	Dos copias de la obra completa de Shakespeare.
100 000 000	100 MB	Un estante de 1 metro de libros.
1 000 000 000	1 GB	Una furgoneta llena de páginas con texto.
1 000 000 000 000	1 TB	Todas las Paginas de texto elaboradas de 50 000 árboles.
10 000 000 000 000	10 TB	La colección impresa de la biblioteca del congreso de EEUU.
1 000 000 000 000 000	1 PB	Los datos que maneja Google cada hora.
1 000 000 000 000 000 000	1 EB	El peso de todos los datos en Internet para finales del año 2001.





# Representación de datos alfabéticos e alfanuméricos

- ASCII
- UNICODE
- BCD
- EBCDIC

```
, ViWkkkGPPPPPPGGkiw,
                            ifXXXXXXXXffXXXXAAAAAffffXXXXXXXXXXk
                      WXXXXXXXfAPPPPAXXXfAAXXPPAAPPPAPPPPPPPAXXXXXXX
                  WXXXXXLAPAAPPGAXLXXXXXXXPGXAGAPPAAAPAAPPPPPPGAXXXXXX
                1XXXXfPPAPAPPPPAfffXXXk
                                         AXGEEGAAPPPGPAPPPPPPXXXXXXX
VXfaaaaXXXXXXXXXXXAPaaaaaPPPPPPPGfX.
                                        : uXGPXPkPPPPPPPPPPAXXXXf
VXffAAAAAAfAAfXfAAAAPAAPPPPPPGGGGGfu
                                       VW1XGGXXGPPPAPPPAXXXX
.XfffaffXfPPPAPPPAAAAPPPPPAfffffPPXXXf
                                      , WWXGGXXkPPPPPPAXXXA
 XXfffXXPPAAPAAAAAPAPPPAfXXXXXXffXfPAXXXu; PXGAXPGAAAAfXXX
 XXXXXAGPPPPPPPAPPAPPAXXXXXXXXXXXXA£A£APPXXXXAGXXkPPAAAXX
  XXfffAPPPPPPPPPPPAXXXXW.
                             wwGXXAAPAAPPPPPPXPPAAAAPPX,;
  XGXXXfPPPPPPPPPPAXXA:
                               wiXXAAPAPAAAAAAAAAAAAXXXXXXXXXXXXX
  PXXk. AXAAAPPPPPXXX.
                                uWXPAPPPAAAAAAAAAAAAAAAAAAAAAAAAA
                                ,uXAAAPAAAAAffffffAAfAAAAAAAAPAPPPfXXXXX,
        XXAPPPXXk
                                 iffPAPAAAAffffffffffkfAfffkAAAAAAAAAPPfXXXP
        XXAPPAXP
                                 1fXAAAAAAAfffXXXffffffffAAAfffAAAPAAPAPXXXV
         XAPPXX,
                                WIXTPAAAAATTTTTTTXXTTTTTTTTTATAAAAAAAAAPPXXX
        XXPGAX:
                                XXV.
        X XAXX
                                VGXfafffffaffXXXXXXfffXXXXXXXXX,
    XG,
        XXXf
                                WXXXXXXXffXXXXXXXXXXXXX;
         XXX
                           XXX wXXXXXXXXXXXXXXXXXXXXX
         GXX
     RXXXXXXX
                                . WXP
        XXG
         kXXXV
               :fXXXXXfGiuuVViWG&XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXff&&ffXX
                   ; kXXXXXXXXXXXXXXX .,
                                            ,.. xxxxxxxxxxxxxxxxxx
                                                                  uXX
```

#### **ASCII**

- American Standard Code for Information Interchange (Código Estadounidense Estándar para el Intercambio de Información)
- Recomendación X3.4-1977 do Instituto Estadounidense de Normas Nacionais (ANSI).
- ASCII é unha correspondencia entre cadeas de bits e unha serie de símbolos (alfanuméricos e outros), permitindo desta forma a comunicación entre dispositivos dixitais así como o seu procesamento e almacenamento.
- **7 bits**  $\rightarrow$  2<sup>7</sup> caracteres diferentes (128)

### **ASCII**

- Caracteres de control (33): nulo, timbre, salto de liña, retorno de carro, ...
- Caracteres imprimibles (95): 0-9, a-z, A-Z, !, ?, #, @, etc

https://es.wikipedia.org/wiki/ASCII#Los\_caracteres\_de\_control\_ASCII

https://es.wikipedia.org/wiki/ASCII#Caracteres\_imprimibles\_ \_ASCII

Binario	Decimal	Hex	Abreviatura	Repr AT		Nombre/Significado
0000 0000	0	00	NUL	NUL	^@	Carácter Nulo
0000 0001	1	01	SOH	SOH	^A	Inicio de Encabezado
0000 0010	2	02	STX	STX	^B	Inicio de Texto
0000 0011	3	03	ETX	ЕΤХ	^C	Fin de Texto
0000 0100	4	04	EOT	EOT	^D	Fin de Transmisión
0000 0101	5	05	ENQ	ENQ	^E	Consulta
0000 0110	6	06	ACK	ACK	^F	Acuse de recibo
0000 0111	7	07	BEL	BEL	^G	Timbre
0000 1000	8	08	BS	BS	^H	Retroceso
0000 1001	9	09	HT	нт	^	Tabulación horizontal
0000 1010	10	0A	LF	LF	^J	Salto de línea

Binario	Dec	Hex	Representación
0100 0000	64	40	@
0100 0001	65	41	Α
0100 0010	66	42	В
0100 0011	67	43	С
0100 0100	68	44	D
0100 0101	69	45	E
0100 0110	70	46	F
0100 0111	71	47	G

Binario	Dec	Hex	Representación		
0110 0000	96	60	,		
0110 0001	97	61	а		
0110 0010	98	62	b		
0110 0011	99	63	С		
0110 0100	100	64	d		
0110 0101	101	65	е		
0110 0110	102	66	f		
0110 0111	103	67	g		

### **ASCII**

#### ASCII Extendido → 8 bits (256 caracteres)

Cód.	Caráct.	Cód.	Caráct.	Cód.	Caráct.	Cód.	Caráct.	Cód.	Caráct.
176	0	192	À	208	Ð	224	à	240	ð
177	±	193	Á	209	Ñ	225	á	241	ñ
178	2	194	Â	210	Ò	226	â	242	ò
179	3	195	Ã	211	Ó	227	ã	243	ó
180	•	196	Ä	212	Ô	228	ä	244	ô
181	h	197	Å	213	Õ	229	å	245	õ
182	¶	198	Æ	214	Ö	230	æ	246	ö
183		199	ç	215	×	231	ç	247	÷
184	,	200	È	216	Ø	232	è	248	ø
185	1	201	É	217	Ù	233	é	249	ù
186	٧	202	Ê	218	Ú	234	ê	12/50	Ú

#### UNICODE

Intento de crear un estándar universal, uniforme e único (o máis usado)

#### Formatos:

- UTF-8: 8 bits.
- UTF-16: 16 bits.
- UTF-32: 32 bits.

UTF: Unicode Transformation Format

Versión 6 (ano 2010) : 109.449 caracteres

#### **BCD**

Binary Coded Decimal (decimal codificado en binario).

Cada díxito decimal representase con 4 bits.

```
Decimal: 0 1 2 3 4 5 6 7 8 9 BCD: 0000 0001 0010 0011 0100 0101 0110 0111 1000 1001
```

#### Exemplo:

```
Decimal: 5 9 2 3 7 BCD: 0101 1001 0010 0011 0111
```

Binario puro:

```
1110011101100101
```

#### **EBCDIC**

#### **Extended BCD Interchange Code**

Código BCD extendido de caracteres decimales codificados en binario para intercambio de información.

8 bits → 256 caracteres.

EBCDIC = BCD + caracteres alfanuméricos