

Introducción a los sistemas informáticos



El sistema informático: definiciones

- Hardware

- La palabra hardware en informática se refiere a las partes físicas, tangibles, de un sistema informático, sus componentes eléctricos, electrónicos, electromecánicos y mecánicos.
- Según la RAE, conjunto de aparatos de una computadora.

- Software

- Según la RAE, el software es un conjunto de programas, instrucciones y reglas informáticas que permiten ejecutar distintas tareas en una computadora. ... Se considera que el software es el equipamiento lógico e intangible de un ordenador.



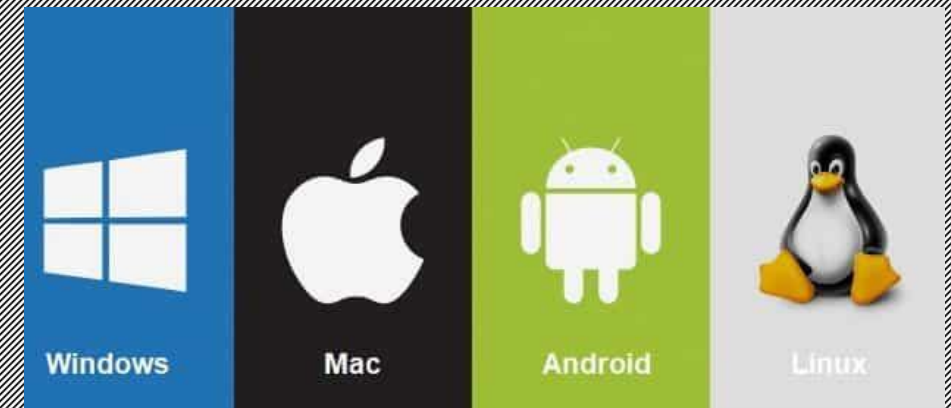
El sistema informático: definiciones

- Programa
 - Según la RAE, Conjunto unitario de instrucciones que permite a una computadora realizar funciones diversas, como el tratamiento de textos, el diseño de gráficos, la resolución de problemas matemáticos, el manejo de bancos de datos, etc.
- Aplicación informática
 - Conjunto de varios programas



El sistema informático: definiciones

- Sistema operativo
 - Conjunto de órdenes y programas que controlan los procesos básicos de una computadora y permiten el funcionamiento de otros programas.
 - El sistema operativo es el componente software de un sistema informático capaz de hacer que los programas (software) procesen información (datos) sobre los componentes electrónicos de un ordenador o sistema informático (hardware).



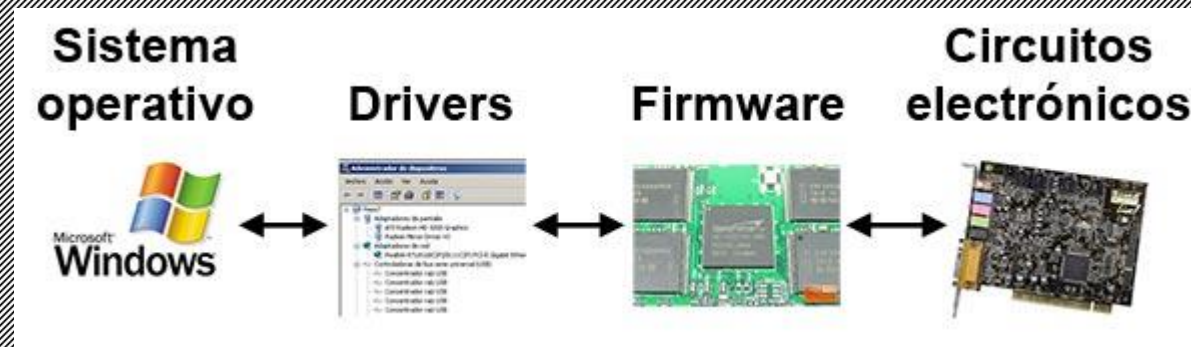
El sistema informático: definiciones

- Firmware (I)
 - El firmware, también conocido como soporte lógico inalterable, es el programa básico que controla los circuitos electrónicos de cualquier dispositivo.



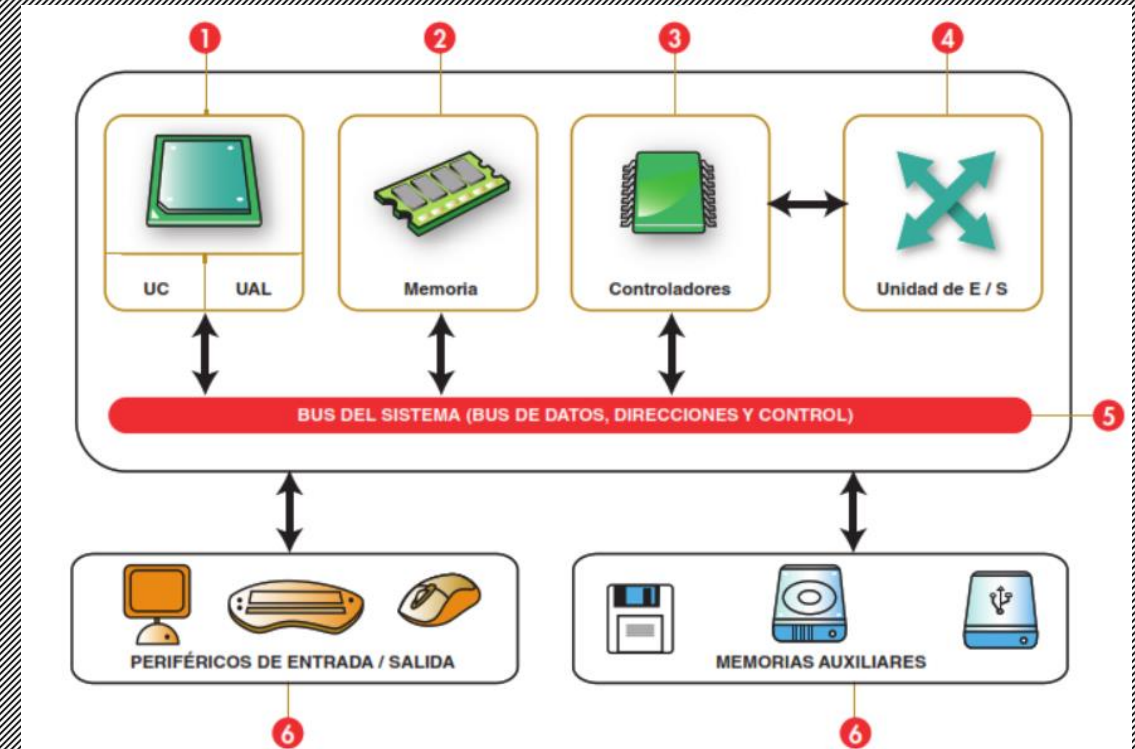
El sistema informático: definiciones

- Firmware (II)
 - El firmware es un bloque de instrucciones de programa para propósitos específicos, grabado en una memoria tipo ROM, que establece la lógica de mas bajo nivel que controla los circuitos electrónico de un dispositivo de cualquier tipo.



Componentes físicos: el hardware

1. Unidad central de proceso (UCP)
 - Unidad aritmético-lógica (UAL)
 - Unidad de control (UC)
2. Memoria central
3. Controladores
4. Unidad de entrada/salida (E/S)
5. Buses
6. Unidad periféricas



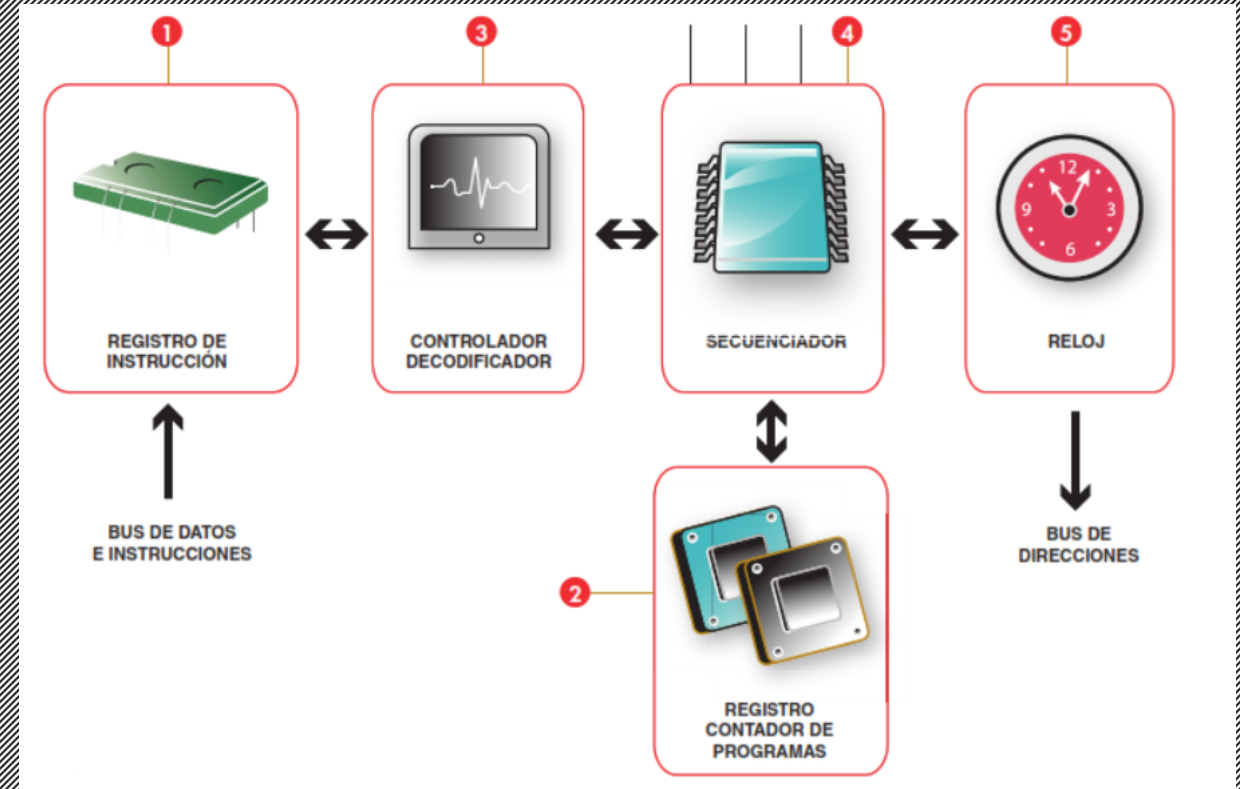
Unidad central de proceso (UCP) o procesador

- Es el elemento encargado del control y ejecución de las operaciones que se efectúan dentro del ordenador con el fin de realizar el tratamiento automático de la información.
- Es la cabeza pensante del sistema informático y se encarga de controlar todas las tareas y procesos que se realizan dentro de él.



Unidad de control (UC)

- La función principal de la unidad de control de la CPU es dirigir la secuencia de pasos de modo que la computadora lleve a cabo un ciclo completo de ejecución de una instrucción, y hacer esto con todas las instrucciones de que conste el programa.
- Componentes
 1. Registro de instrucción
 2. Registro de contador de programa
 3. Controlador y decodificador
 4. Secuenciador
 5. Reloj



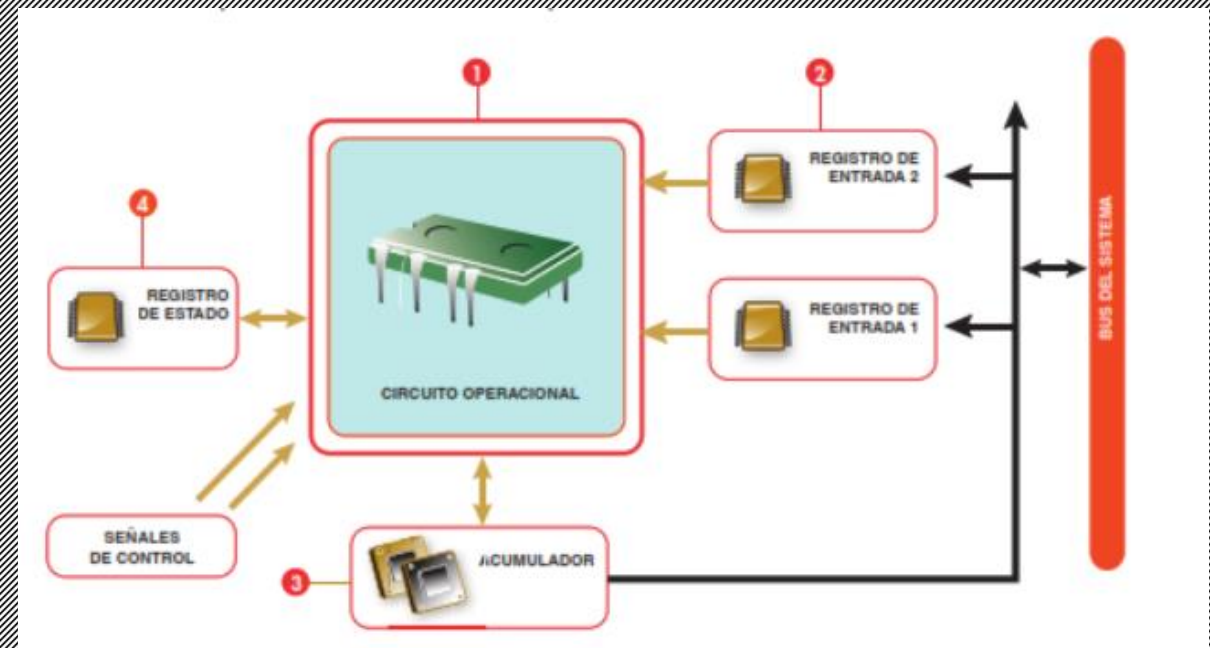
Unidad aritmético-lógica (UAL) (I)

- La unidad aritmético-lógica es la parte de la UCP encargada de realizar operaciones aritméticas y lógicas sobre la información.
- Op. Aritméticas → Sumas, restas, multiplicaciones, divisiones...
- Op. Lógicas → operaciones de comparación (>, <, >=, ...)

Operación	Operador
Mayor que	>
Menor que	<
Mayor o igual	>=
No mayor	NOT > (<=)
Y lógico	AND
O lógico	OR
Álgebra de Boole.	

Unidad aritmético-lógica (UAL) (II)

- Componentes de la UAL
 - Operacional o circuito operacional
 - Realiza las operaciones con los datos de los registros de entrada
 - Registros de entrada
 - Contiene los operandos de la operación.
 - Acumulador
 - Almacena los resultados de las operaciones.
 - Registro de estado
 - Registra las condiciones de la operación anterior



La memoria. Funciones, tipos y características.

- El ordenador almacena dentro de su memoria interna todos los programas y datos con los que va a trabajar y van a ser procesados.
- Tipos de memoria
 - Memorias de almacenamiento externo
 - Memoria interna
 - RAM
 - ROM

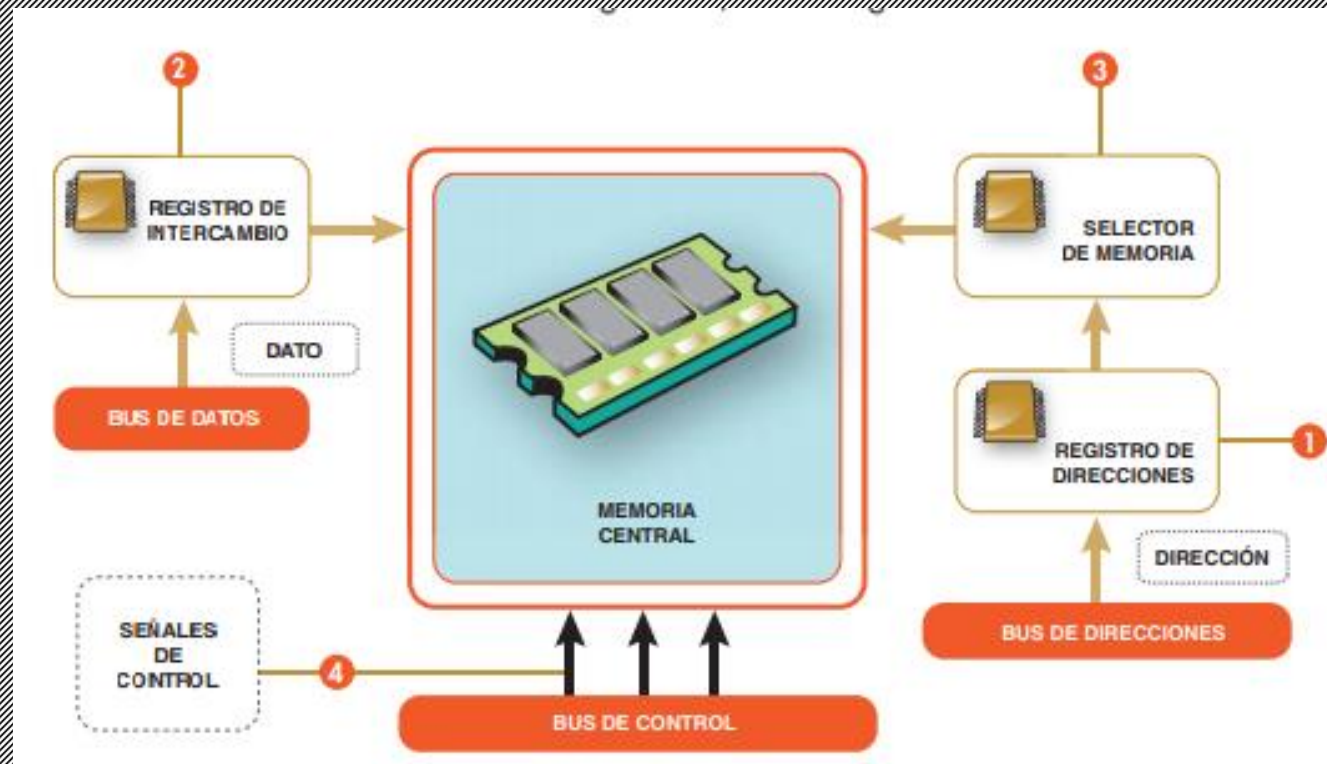
Jerarquías de memoria



La RAM

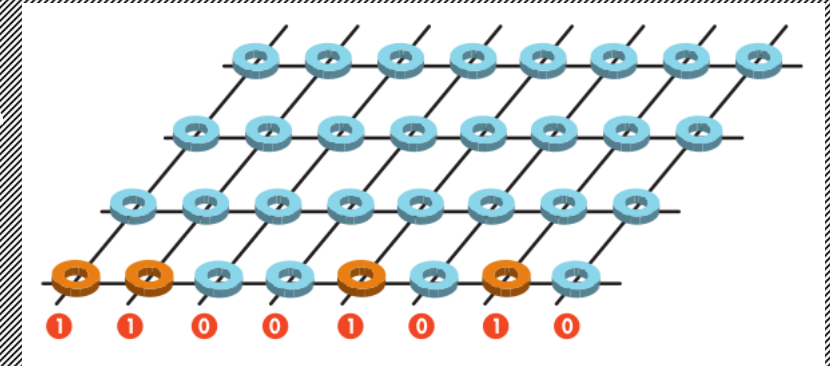
- La memoria RAM, también denominada memoria de acceso aleatorio o memoria de acceso directo, se compone de uno o más chips y se utiliza como memoria principal para programas y datos.
- Los elementos que componen la memoria principal son:
 - Registro de direcciones
 - Contiene la dirección de la celda o posición de memoria a la que se va a acceder.
 - Registro de intercambio
 - Recibe los datos de en operaciones de lectura y almacena los datos en las operaciones de escritura.
 - Selector de memoria
 - Se activa cada vez que hay que leer o escribir conectando la celda o posición de memoria con el registro de intercambio.
 - Señales de control
 - Indica si una operación es de lectura o escritura.

La RAM



La RAM

- Esta formada de componentes electrónicos (biestables) capaces de almacenar información en forma de ceros y unos (sistema binario).
- Para almacenar información en la memoria esta suele ser referenciada por bloques. Estos bloques suelen ser de ocho celdillas o biestables (8 bits \rightarrow 1 byte). Cada conjunto de ellos representa un carácter, es decir, cualquier letra o número como combinación de 8 bits.
- Para no perder información de la memoria los biestables son recargados constantemente, lo que se denomina refresco de memoria.

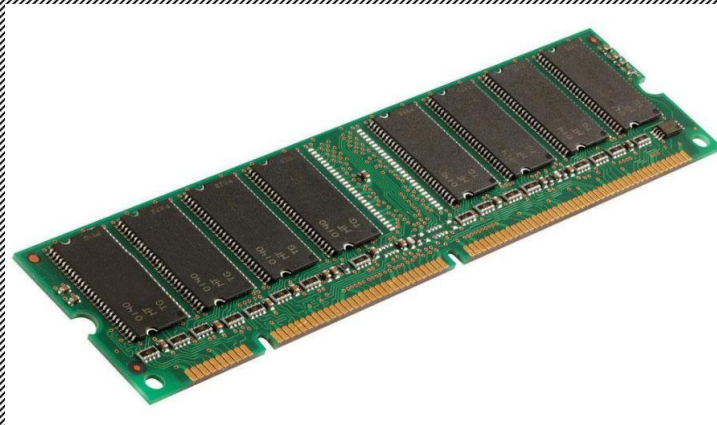


Tipos de memoria

- DRAM (Dynamic RAM)
 - Construida mediante condensadores. Necesita refrescarse cada cierto tiempo. Es más lenta que la SRAM.
- SRAM (Static RAM)
 - Alternativa a la DRAM que no necesita refrescarse.
 - Debido a su alto coste se suele utilizar como memoria cache.
- SDRAM (Synchronous Dynamic RAM)
 - Tiene la capacidad de la DRAM y la velocidad de la SRAM.
 - Necesita refresco de sus celdas.
- DDRAM (Double Data Rate)
 - Doble recarga.
 - Doble tasa de transferencia.
 - Refresca dos veces por impulso de reloj.

Tipos de memoria

DRAM



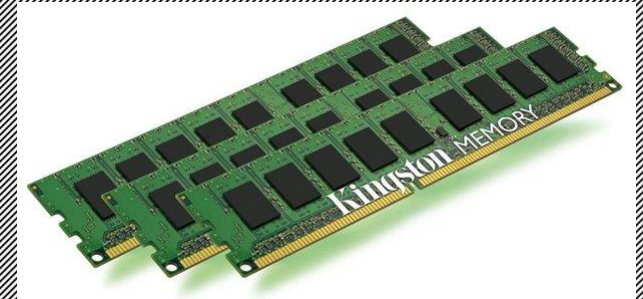
SRAM



SDRAM



DDRAM



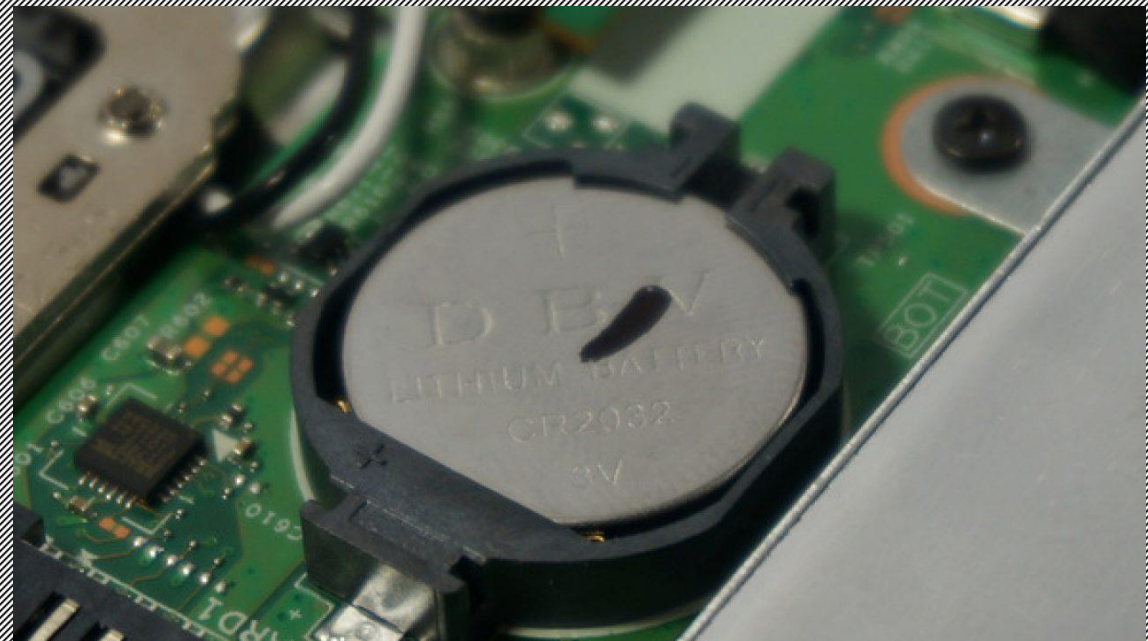
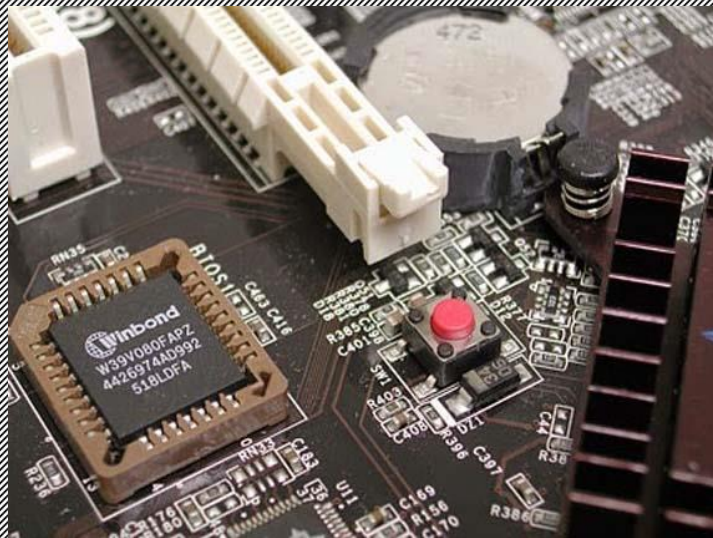
ROM (read-only memory)

- La memoria ROM o memoria de solo lectura contiene programas especiales que sirven para cargar e iniciar el arranque del ordenador.
- Se encuentra almacenada toda la información referente a los componentes hardware.
- El software que integra la ROM forma el BIOS (Basic Input Output System)
- Tipos de memorias:
 - PROM (Programmable Read Only Memory)
 - EPROM (Erasable Programmable Read Only Memory)



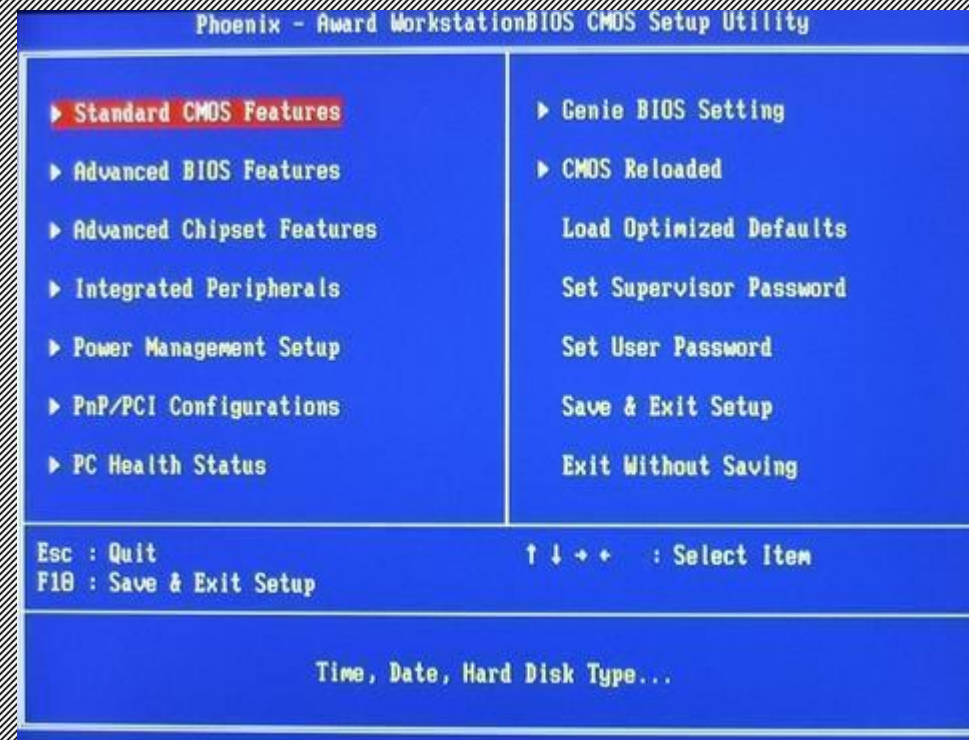
CMOS (Complementary metal–oxide–semiconductor)

- Tipo de memoria interna del ordenador que se caracteriza por consumir muy poca energía eléctrica, lo que le hace idónea para almacenar datos de la BIOS.
- Se utiliza para guardar los datos básicos de hardware y de configuración.
 - Ejemplo. Fecha y hora.



ACCESO A LA BIOS

- Se suele utilizar las teclas F2 o SUPR.



Otros tipos de memoria

- VRAM o memoria de video
 - Actualmente se utiliza DDR SGRAM (GDDR4 (2005), GDDR5 (2008), GDDR5X (2016) y GDDR6 (2018))
- CDRAM (Cache DRAM)
 - es una mezcla de memoria estática (SRAM) y memoria dinámica (DRAM).
 - Similar a la caché de los modernos procesadores, en la CDRAM los datos frecuentemente usados se almacenan en la rápida SRAM, lo que incrementa el rendimiento.
 - Suelen ir asociadas a determinados dispositivos como CD-ROM

Unidades de E/S y el BUS

- Unidad de E/S
 - Sirve para comunicar el procesador y el resto de componentes internos del ordenador con los periféricos de E/S y las memorias de almacenamiento externo o auxiliares.
- BUS
 - Es el elemento responsable de establecer una correcta interacción entre los diferentes componentes del ordenador.
 - Bus de datos
 - Bus de direcciones
 - Bus de control

LOS PERIFÉRICOS

- Dispositivos hardware con los cuales el usuario puede interactuar.
- Los periféricos se conectan con el ordenador a través de los denominados puertos.
- El componente hardware usado para la gestión de los periféricos es la unidad de E/S
- Muchos de los periféricos de E/S necesitan un tipo de software especial para ser configurados denominados drivers o controladores.
- Podemos distinguir entre
 - Periféricos de entrada
 - Periféricos de salida
 - Periféricos de E/S
 - Periféricos de almacenamiento
 - Periféricos de comunicación.

Componentes lógicos

Tipos de datos

- Según que podemos hacer
 - Datos de entrada
 - Datos intermedios
 - Datos de Salida
- Según varíen o no
 - Datos fijos
 - Datos variables

Sistemas de codificación

- La codificación de caracteres es el método que permite convertir un carácter de un lenguaje natural (como el de un alfabeto o silabario) en un símbolo de otro sistema de representación, como un número o una secuencia de pulsos eléctricos en un sistema electrónico, aplicando normas o reglas de codificación.

El sistema binario

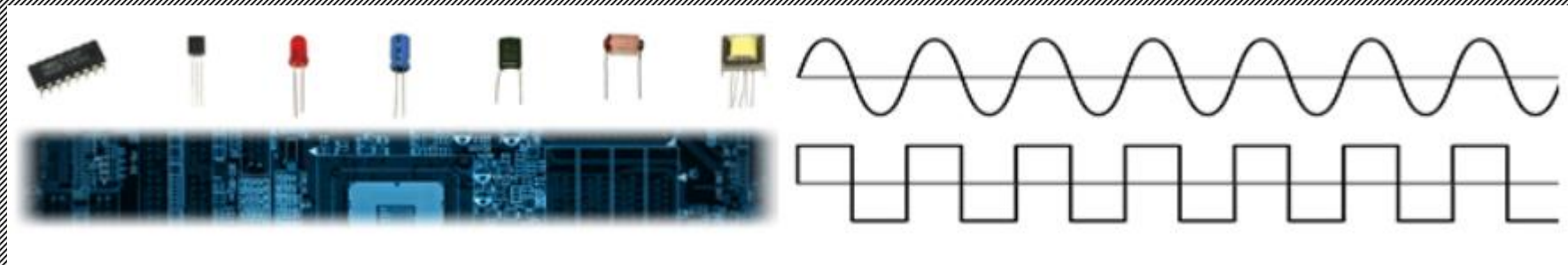
Un **componente electrónico** es un dispositivo capaz de manipular electrones o sus campos asociados. Algunos ejemplos son transistores, diodos, y condensadores.

Los componentes electrónicos se combinan y conectan mediante cables para formar **circuitos electrónicos** capaces de realizar operaciones complejas. La radio es un circuito electrónico que transforma una señal electromagnética en una señal sonora.

A un circuito electrónico lo llamamos **analógico** si opera con rangos continuos de señales eléctricas. Por ejemplo, un altavoz amplifica la señal a un número indefinido de valores entre un volumen mínimo y máximo.

El sistema binario

En cambio, si opera con señales discretas, decimos que el circuito electrónico es **digital**. Un ordenador es un circuito digital que opera con dos voltajes de valores exactos: cero voltios, y el número de voltios que proporcione la fuente de voltaje, por ejemplo 1.4 voltios. Esto es así porque están contruidos con **transistores**, que son componentes electrónicos cuyo funcionamiento es más fiable cuando representa solo dos estados. A grandes rangos, cuantos más estados representas con un rango finito de voltaje, más posibilidades existen de confundir un estado con el otro.



El sistema binario

En resumen, los ordenadores expresan información usando solo dos estados porque así sus componentes son más fiables.

El sistema de numeración más fácil de implementar con dos estados es el **sistema numérico binario**. El binario es similar al decimal, pero solo utiliza los dígitos uno y cero.

Por ejemplo:

decimal	0	1	2	3	4	...
binario	0	1	10	11	100	...

Bits y bytes

Cuando transferimos texto usando un ordenador, existen componentes electrónicos y cables que leen y escriben estas cadenas de ceros y unos. Cada uno de los dígitos 0 o 1 que circulan por el ordenador, se considera la unidad mínima de información en informática, y se conoce como **bit**. Para acelerar la transferencia, los bits se transfieren en grupos de 8, 16, 32, o 64 denominados palabras. Cuantos más bits lees a la vez, más cables y componentes necesitas para transferirlos. El equilibrio entre miniaturización, coste de construcción, y otras consideraciones prácticas, hace que los ordenadores actuales operen con bloques de 64 bits. Sin embargo hace 50 años, el límite práctico estaba en 8. Un grupo de 8 bits se conoce como **byte**.

Bits y bytes

Las unidades de información son múltiplos de dos por dos motivos:

El ordenador usa el sistema de numeración binario tanto para expresar datos, como para referenciar la posición en memoria de esos datos.

El número de valores expresable con un dígito es $\text{base}^{\text{número de dígitos}}$ (que es un múltiplo de la base).

En consecuencia, la capacidad expresiva de un dígito es completamente aprovechada si referencia una cantidad de información que múltiplo de su base. Por ejemplo con un dígito decimal (base 10) puedes expresar 10 valores (0, 1, 2, 3, 4, 5, 6, 7, 8, 9). Por tanto si usáramos la base decimal, lo más eficiente sería referenciar información en múltiplos de diez, de otro modo desperdiciaríamos la capacidad expresiva de un dígito.

El motivo por el que además son múltiplos de 8 es histórico. Han existido computadores en los que un byte tenía 1, 6, 7, 8, 9, 12, 18, y 20 bits, pero la popularidad del **chip 8080** (el cual usaba 8-bit) hizo que byte fuera sinónimo de 8 bits.

unidades de medida de información más utilizadas

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

Conversión de un número decimal a binario

Dividir sucesivamente el número decimal y los cocientes que se van obteniendo por 2 hasta que el cociente sea menor de 2. La unión del último cociente y todos los restos obtenidos, escritos en orden inverso, será el número expresado en binario.

25	2			
1	12	2		
Quinto	0	6	2	
	Cuarto	0	3	2
		Tercero	1	1
			Segundo	Primero

el número decimal 25 será el 11 001 en el sistema binario

Conversión de un número decimal a binario

Pesos asociados							
2^7	2^6	2^5	2^4	2^3	2^2	2^1	2^0
128	64	32	16	8	4	2	1

Conversión de un número binario A decimal

i = posición empezando en 0.

D = número de dígitos a la derecha de la coma.

N = número de dígitos.

dígito = cada uno de los que componen el número.

base = base del sistema de numeración.

1011

$$(1 \cdot 2^3) + (0 \cdot 2^2) + (1 \cdot 2^1) + (1 \cdot 2^0) = 11$$

$$N_i = \sum_{i=-d}^n (\text{dígito})_i \cdot (\text{base})^i$$

OPERACIONES

Suma binaria

Suma binaria

$$0 + 0 = 0$$

$$0 + 1 = 1$$

$$1 + 0 = 1$$

$$1 + 1 = 0, \text{ acarreo } 1$$

Resta binaria

$$0 - 0 = 0$$

$0 - 1 = 1$, acarreo 1, que
se suma al siguiente sustraendo

$$1 - 0 = 1$$

$$1 - 1 = 0$$

Multiplicación binaria

$$0 \cdot 0 = 0$$

$$0 \cdot 1 = 0$$

$$1 \cdot 0 = 0$$

$$1 \cdot 1 = 1$$

25

			1	0	0	0	0	→ 16
+	1	0	1	0	0	1	→ 41	
	1	1	1	0	0	1	→ 57	

				1	1	1		Acarreos
				↓	↓	↓		
	1	0	1	0	1	1	1	→ 87
+		1	0	0	0	0	1	→ 33
<hr style="border: 0.5px solid black;"/>								
	1	1	1	1	0	0	0	→ 120

	1	1	1	0	1	0	1	→ 117
-	↓1	↓1	↓1		↓1			Acarreos
		1	1	1	0	1	0	→ 58
	0	1	1	1	0	1	1	→ 59

$$\begin{array}{r}
 \rightarrow 25 \\
 \rightarrow 5 \\
 \hline
 \\
 + \\
 \\
 \hline
 \rightarrow 125
 \end{array}$$

$$\begin{array}{r}
 1010 \mid 10 \\
 00 101 \quad \text{Cociente} \rightarrow 5 \\
 10 \\
 00 \quad \text{Resto} \rightarrow 0
 \end{array}$$

El sistema octal

El sistema octal tiene como base de numeración 8, es decir, utiliza ocho símbolos para representar las cantidades. Estos símbolos son 0, 1, 2, 3, 4, 5, 6, 7.

Conversión de un número decimal a octal

925	8		
920	115	8	
45	35	14	8
5	3	6	1
Cuarto ↑	Tercero ↑	Segundo ↑	Primero ↑

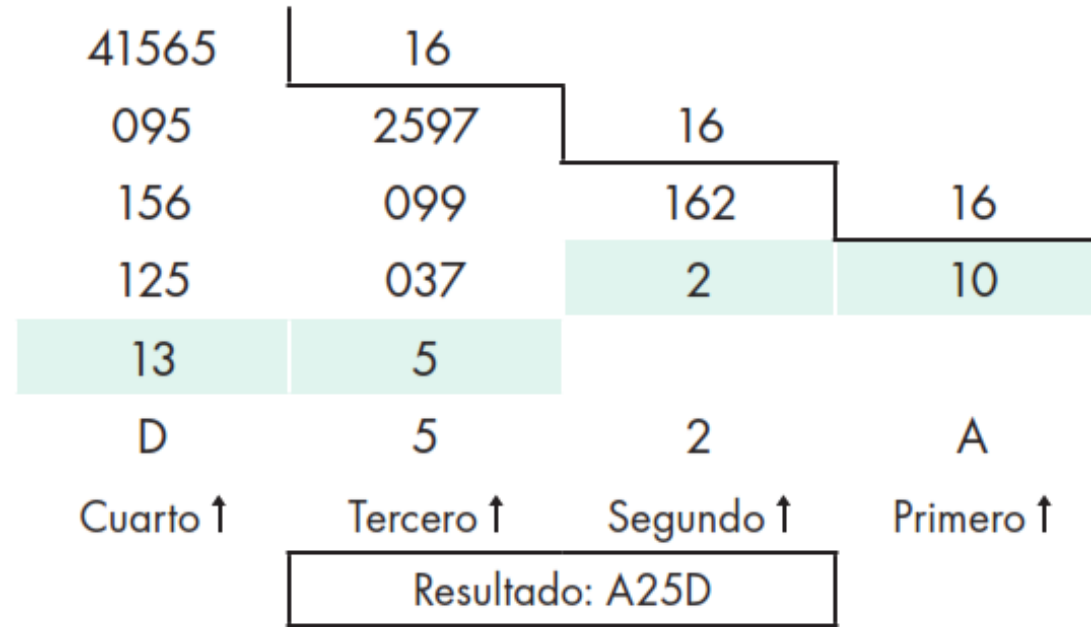
Pesos asociados en el sistema octal			
8^3	8^2	8^1	8^0
512	64	8	1
1	6	3	5

El sistema hexadecimal

El sistema hexadecimal tiene como base de numeración 16, es decir, utiliza dieciséis símbolos para representar las cantidades. Estos símbolos son 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, A, B, C, D, E, F.

Símbolo	Valor asignado
A	10
B	11
C	12
D	13
E	14
F	15

El sistema hexadecimal



Pesos asociados en el sistema hexadecimal			
16^3	16^2	16^1	16^0
4096	256	16	1
	A	1	D

Otras conversiones

Conversión hexadecimal-binario

Pasar a binario 73B

7	-	3	-	B
↓		↓		↓
0111		0011		1011

Conversión binario-hexadecimal

Pasar a hexadecimal 101011011

0001	-	0101	-	1011
↓		↓		↓
1		5		B

Conversión octal-binario

Pasar a binario 527

5	-	2	-	7
↓		↓		↓
101		010		111

Conversión binario-octal

Pasar a binario 010101100

010	-	101	-	100
↓		↓		↓
2		5		4

Otras conversiones

Conversión hexadecimal-octal

Pasar 1AB0C

Convertir a binario

1	A	B	0	C
↓	↓	↓	↓	↓
0001	1010	1011	0000	1100

Convertir a octal

011	010	101	100	001	100
↓	↓	↓	↓	↓	↓
3	2	5	4	1	4

Conversión octal-hexadecimal

Pasar 3710

Convertir a binario

3	7	1	0
↓	↓	↓	↓
011	111	001	000

Convertir a hexadecimal

111	1100	1000
↓	↓	↓
7	C	8

¿Qué es un carácter?

Un **carácter** es una unidad de información abstracta. Algunos ejemplos de caracteres son los números y letras del alfabeto, los signos de puntuación, los emojis, y los caracteres de control como un salto de línea.

Un **juego de caracteres** es una colección de caracteres usada para escribir en un lenguaje particular. Algunos idiomas usan uno, otros usan varios (por ejemplo, kanji y kana del japonés), y otros usan variantes de un mismo juego de caracteres (por ejemplo, todos los lenguajes europeos usan variantes del alfabeto romano).

Un **glifo** es una representación visual de un carácter. En un ordenador, los glifos asociados a caracteres suelen estar almacenados en **fuentes**, también llamados tipos de letra. Cada fuente contiene un estilo homogéneo de glifos para un alfabeto. Este es un ejemplo de varios glifos que representan un mismo carácter:



Tipos de caracteres

A grandes rasgos hay tres sistemas de escritura

Alfabético. Un alfabeto es el conjunto de letras usadas para escribir un lenguaje. Cada una representa un fonema o se usa junto a otras para representar un fonema. La mayoría de alfabetos en uso se basan en el alfabeto romano.

Silábico. Contiene caracteres que representan cada una de las sílabas de un lenguaje. Una sílaba es una combinación de vocal o vocal y consonante. Es apropiado para lenguajes con pocas sílabas, como el japonés, que contiene un centenar. El inglés en cambio, contiene miles de combinaciones de conjuntos de vocales y consonantes.

Ideográfico. Contiene caracteres que representan ideas. Por ejemplo, el chino, o el egipcio antiguo.

Estas categorías no son rígidas. Por ejemplo, el español es alfabético, pero también usa un ideograma como el símbolo de euro, que representa la estabilidad de Europa (es una e de Europa con dos líneas indican estabilidad).

ASCII

El único tipo de dato que un ordenador es capaz de almacenar y manipular es el número binario. El truco para representar texto es asignar un número a cada letra del alfabeto, e implementar un sistema de renderizado que visualiza el carácter correspondiente a cada número.

Llamamos código al sistema de normas para convertir información simple en información codificada. Una codificación de caracteres por ejemplo, es un código que convierte caracteres a números.

La codificación de caracteres más popular se llama ASCII (American Standard Code for Information Interchange). ASCII asigna números decimales del 0 al 127 a las letras del alfabeto inglés, los números, y algunos signos de puntuación y de control. Los signos de control son indicadores para procesar información, por ejemplo, el salto de línea. La letra "A" en ASCII tiene asignado el número 65 (0x41 en hexadecimal), y las letras consecutivas del alfabeto tienen asignados números consecutivos:

ASCII

letra	binario	decimal
A	1000001	65
B	1000010	66
C	1000011	67
D	1000100	68
etc.		

Low Ascii															
000:	013:~	026:~	039:'	052:4	065:a	078:H	091:[104:h	117:u						
001:0	014:~	027:~	040:(053:5	066:B	079:O	092:\	105:l	118:v						
002:0	015:~	028:~	041:)	054:6	067:C	080:P	093:]	106:j	119:w						
003:~	016:~	029:~	042:*	055:7	068:D	081:Q	094:~	107:k	120:x						
004:~	017:~	030:~	043:~	056:8	069:E	082:R	095:~	108:l	121:y						
005:~	018:~	031:~	044:~	057:9	070:F	083:S	096:~	109:n	122:z						
006:~	019:~	032:~	045:~	058:~	071:G	084:T	097:a	110:n	123:~						
007:~	020:~	033:~	046:~	059:~	072:H	085:U	098:b	111:o	124:l						
008:~	021:~	034:~	047:~	060:<	073:I	086:V	099:c	112:p	125:~						
009:~	022:~	035:~	048:~	061:=	074:J	087:W	100:d	113:q	126:~						
010:~	023:~	036:~	049:~	062:>	075:K	088:X	101:e	114:r	127:~						
011:~	024:~	037:~	050:~	063:~	076:L	089:Y	102:f	115:s							
012:~	025:~	038:~	051:~	064:~	077:M	090:Z	103:g	116:t							

ASCII

Según este código, la representación de las letras “abc” en binario es la siguiente: 1000001
1000010 1000011

El documento que estas leyendo en este momento tiene una representación en binario en tu ordenador en forma de ceros y unos. A nivel físico se representa en la memoria mediante la presencia o ausencia de voltaje, y en el disco mediante magnetismo.

Cuando ASCII se publicó en 1964, las máquinas de entonces transferían 8 bits a la vez. Sin embargo ASCII usó solo 7, y reservó el octavo bit para detectar errores de transmisión. Esos 7 bits permiten expresar $2^7 = 128$ valores, que ASCII usó para codificar 33 señales de control (usadas entre máquinas), y 95 números, símbolos de puntuación, y letras del alfabeto inglés.

Variantes del ASCII

En los '80 aparecieron variantes de ASCII que usaban el octavo bit para representar caracteres adicionales, lo que permitió representar caracteres adicionales propios de alfabetos no ingleses, como por ejemplo el español. Debido a que 8 bits no son suficientes para representar todos los alfabetos del mundo, continuaron apareciendo variantes ASCII de 8 bits incompatibles entre sí. Estas variantes se llaman a veces **ASCII extendido**, pero no forman parte del estándar ANSI.

Hay varios conjuntos de "ASCII extendido", cada uno de los cuales contiene codificaciones para muchos lenguajes:

Windows code pages, usado en aplicaciones gráficas Windows.

OEM code pages, usando en aplicaciones de consola Windows.

ISO-8859 es un estándar ISO para codificación en 8 bits. Tiene 16 partes. La primera se llama ISO-8859-1, también conocida como Latin-1, que cubre la mayoría de lenguajes de Europa occidental.

En ASCII solo es posible trabajar con un alfabeto a la vez. ASCII tampoco es válido para representar alfabetos asiáticos, porque contienen miles de caracteres. Estas deficiencias, junto a la popularidad de la informática, y la potencia creciente de los ordenadores, hicieron deseable y posible la creación de un estándar más exhaustivo.

Unicode

Unicode es una codificación de caracteres que asigna un número a cada uno de los caracteres de prácticamente todos los alfabetos existentes, incluyendo las lenguas muertas como el egipcio antiguo y otros. El estándar Unicode lo publica y mantiene el consorcio Unicode, una entidad sin ánimo de lucro formada por empresas e individuos.

El estándar Unicode también especifica algoritmos sobre cómo manejar texto, que probablemente solo necesites si realizas tareas especializadas. Por ejemplo:

- Dividir palabras y líneas.

- Ordenar texto.

- Formatear números, fechas, y horas.

- Mostrar texto que fluye de derecha a izquierda.

- Mostrar texto cuya forma escrita se combina y reordena.

- Tratar problemas de seguridad relativos a caracteres parecidos.

Unidades y puntos de código

Unicode comenzó como un sistema de codificación de 16 bits, aunque desde Unicode 2.0 (publicado en 1996), es una codificación de 21 bits.

Un punto de código es el número con el que se identifica un carácter en el estándar Unicode. El punto de código se escribe con el formato U+xxxx donde las xxxx son de cuatro a seis dígitos en sistema de numeración hexadecimal.

En Unicode 7.0 el rango válido de puntos de código va de 0 a 10FFFF₁₆. El hexadecimal se usa por conveniencia en lugar del binario, porque es más fácil recordar. Por ejemplo, Unicode asigna el número 65 a la letra a latina mayúscula. El punto de código correspondiente es U+0041 porque 65_{decimal} = 0x41_{hexadecimal}.

Unicode es compatible con la codificación ISO-8859-1, porque los 256 primeros caracteres de Unicode coinciden con los caracteres de ISO-8859-1. Esto hace que la mayoría de texto en uso requiera solo un byte por carácter.

El punto de código se representa con grupos de 8, 16, o 32 bits dependiendo respectivamente, de si el tipo de codificación es UTF-8, UTF-16, o UTF-32.

Cada uno de estos grupos se llama unidad de código.

Una unidad de código es el mínimo grupo de bits necesario para representar una unidad de texto codificado.

En UTF-8 es 8 bit, en UTF-16 es 16 bit, y en UTF-32 es 32 bit. Por ejemplo, U+0041 requiere 2 bytes, es decir, dos unidades de código en UTF-8, y una unidad de código en UTF-16.