

Bits, Bytes y Datos y tipos de datos.

Cualquier computador, incluso el más complejo es en realidad sólo un gran volumen de bits bien organizados.

Es difícil definir el término **información**, ya que tiene muchos significados. De acuerdo con varias definiciones tradicionales, información quiere decir comunicación que tiene valor porque informa. Sin embargo, en el lenguaje de la teoría de la comunicación y la información, el término información puede aplicarse a casi cualquier cosa que pueda comunicarse, tenga o no valor. Según esta definición, que es la que usaremos, la información se presenta de muchas maneras.

Las palabras, los números y las imágenes que aparecen en esta página son símbolos que representan información. Si usted subraya esta frase, añadirá información a la página. Los sonidos y las imágenes que emite un televisor llevan información. (Recuerde que no toda la información tiene valor.)

La información es **digital** en el mundo de los computadores: se compone de unidades discretas (es decir, unidades que pueden contarse), por tanto puede subdividirse. En muchas situaciones, hay que reducir la información a unidades más sencillas para poder usarla de manera efectiva. Por ejemplo, un niño que intenta interpretar una palabra con la cual no está familiarizado, puede pronunciar cada letra en forma individual antes de atacar toda la palabra.

Un computador no entiende palabras, números, imágenes, notas musicales, ni siquiera las letras del alfabeto. Como un lector incipiente, un computador no puede procesar información sin dividirla en unidades más pequeñas. De hecho, los computadores sólo pueden digerir información que ha sido dividida en bits. Un **bit** (**binary digit**: - dígito binario) es la unidad de información más pequeña. Un bit sólo puede tener uno de dos valores: encendido o apagado. También se pueden considerar estos valores como sí o no, cero o uno, blanco o negro, o casi cualquier otra cosa que se le ocurra, que sea dicotómica por supuesto.

Si pensamos en las entrañas de un computador como una colección de microscópicos conmutadores de encendido/apagado, es fácil comprender por qué procesan la información bit a bit. Con cada conmutador se almacena una pequeña cantidad de información; por ejemplo, una señal para encender una luz o la respuesta a la pregunta de tipo "sí/no".

Un computador puede procesar grandes trozos de información tratando grupos de bits como unidades. Por ejemplo, una colección de ocho bits, a la que suele llamársele **byte**, puede representar 256 mensajes diferentes ($256 = 2^8$). Si consideramos a cada bit como una luz que puede estar encendida o apagada, podemos hacer que las diferentes combinaciones de luces representen mensajes distintos. (Los científicos de la computación por lo general hablan de ceros y unos, en lugar de encendido y apagado, pero el concepto no cambia.) El computador no sólo ve el número de luces encendidas, sino también su orden, de manera que 01 (apagado-encendido) es distinto de 10 (encendido-apagado).

Construcción con bits

¿Qué significa para el computador una combinación de bits como 01100110? No hay una respuesta única para esta pregunta; depende del contexto y el convenio. Una cadena de bits puede ser interpretada como un número, una letra del alfabeto o casi cualquier cosa.

- *Los bits como números.* Como los computadores se construyen a partir de dispositivos de conmutación que reducen toda la información a ceros y unos, pueden representar los números con el *sistema numérico binario* un sistema que denota todos los números con combinaciones de dos dígitos. Al igual que el sistema numérico decimal que usamos todos los días, el sistema binario

tiene reglas claras y consistentes para cada operación aritmética.

Representación decimal	Representación binaria	Representación decimal	Representación binaria
0	0	8	1000
1	1	9	1001
2	10	10	1010
3	11	11	1011
4	100	12	1100
5	101	13	1101
6	110	14	1110
7	111	15	1111

Para casi nadie es práctico leer números binarios, por lo que los computadores cuentan con software que convierte automáticamente los números decimales en números binarios y viceversa. Como resultado, el procesamiento de números binarios del computador es totalmente invisible para el usuario humano.

- Los bits como Códigos. Los computadores actuales trabajan casi tanto con texto como con números. Para que las palabras, frases y párrafos se ajusten a los circuitos exclusivamente binarios del computador, se han creado códigos que representan cada letra, dígito y carácter especial como una cadena única de bits.

El código más común, **ASCII** (abreviatura de American Standard Code for Information Interchange. Código estándar estadounidense para el intercambio de información, pronunciado usualmente como "asqui"), representa cada carácter como un código único de siete bits (más un octavo cuyo valor está determinado, por cuestiones técnicas, por los valores de los otros siete). Con los patrones de una cadena de siete bits se pueden obtener 128 códigos únicos, suficientes para asignar un código a cada una de las letras (mayúsculas y minúsculas), números y caracteres especiales usados en la comunicación escrita en inglés. Como el mundo cada vez se hace más pequeño y crecen nuestras necesidades de información, cada vez más usuarios de los computadores encuentran que los 128 caracteres del código ASCII no son suficientes, por lo cual se están elaborando nuevos esquemas de codificación. Para facilitar la computación multilingüe, es probable que los fabricantes cambien algún día del ASCII a un esquema de codificación más rico en información, como el conjunto de 65 000 caracteres de UniCode.

Un grupo de bits también puede representar colores, sonidos, mediciones cuantitativas del ambiente o casi cualquier otro tipo de información que pueda llegar a procesar un computador.

- Los bits como instrucciones en los programas. Hasta ahora hemos visto las formas en que se pueden usar los bits para representar datos, es decir, información de una fuente externa que será procesada por el computador. Sin embargo, hay otro tipo de información igual de importante para el computador: los programas que le indican qué hacer con los datos que le proporcionamos. El computador almacena los programas como colecciones de bits, lo mismo que los datos.

Los programas, al igual que los caracteres, se representan en notación binaria, utilizando códigos. Por ejemplo, el código 01101010 puede indicar al computador que sume dos números. Otros grupos de bits (instrucciones en el programa) contendrían códigos para indicar al computador dónde hallar esos números y dónde almacenar el resultado.

Bits, bytes y otros términos

Tratar de aprender sobre los computadores examinando su operación a nivel de bits es un poco

como tratar de aprender la apariencia o las acciones de la gente estudiando las células humanas, ello arroja mucha información, pero no es la manera más eficiente de averiguar lo que necesitamos saber. Por fortuna, casi todos podemos usar los computadores sin tener que pensar en términos de bits. No obstante, en el trabajo informático cotidiano sí se emplea algo de terminología relacionada con los bits. Específicamente, la mayoría de los usuarios de computadores necesitan conocer, aunque sea en forma somera, los términos siguientes:

- **Byte:** grupo de ocho bits. Si la mayor parte de su trabajo tiene que ver con palabras, puede pensar en un byte como un carácter de información.
- **Kb (kilobyte):** Técnicamente, 1K es 1024 bytes, ya que 1024 es 2^{10} , lo cual facilita la aritmética para los computadores binarios. Para quienes no pensamos en binario, 1000 es una buena aproximación)
- **Mb (megabyte):** aproximadamente un millón de bytes ($1024 \cdot 1024$)
- **Gb (gigabyte):** aproximadamente 1000 megabytes. Esta unidad de medición astronómica se aplica a los dispositivos de almacenamiento más grandes que pueden conseguirse en la actualidad.
- **Terabyte:** son 1.024 GigaBytes, cerca de un billón (un millón de millones) de bytes.

Los términos anteriores se usan para describir la capacidad de algunos componentes de los computadores. Por ejemplo, para describir un computador puede decirse que tiene 128K de caché, 32 Mb de memoria y que un disco duro tiene 8.4 GB de capacidad de almacenamiento.

Con estos mismos términos se cuantifica el tamaño de los archivos del computador. Un archivo es una colección organizada de información, como un artículo o un conjunto de nombres y direcciones, almacenada en una forma que puede leer el computador. El texto de esta guía, por ejemplo, se almacena en un archivo que ocupa 60K de espacio en disco.

Dato.

La computadora digital, tal y como se conoce hoy en día, fue ideada como un dispositivo para facilitar y acelerar las operaciones de cálculo complicadas y que consumen mucho tiempo (cálculo de integrales, solución de sistemas de ecuaciones, simulación entre otras). Hoy en día la capacidad de almacenar y acceder a grandes cantidades de información desempeña la parte principal y dominante de la computación (informática), mientras que la capacidad de contar o computar, es decir, calcular y realizar operaciones aritméticas se ha vuelto irrelevante.

Cualquiera sea el caso, la información representa una abstracción de una parte de la realidad. La información de que se dispone para procesar en la computadora consta de un conjunto de datos acerca del problema real, es decir, el conjunto que se considera relevante para el problema que se tiene, aquel conjunto del cual se cree pueden derivarse los resultados.

Los datos representan una abstracción de la realidad en el sentido de que ciertas características y propiedades de los objetos son ignorados por ser irrelevantes para el problema específico, es como una simplificación de hechos.

Al resolver un problema (con o sin computadora) se necesita elegir una abstracción de la realidad, o sea definir un conjunto de datos que representen la situación real a estudiar. Luego sigue una elección de la representación de esta información; esta elección es guiada por la herramienta que se usa para resolver el problema (los recursos que ofrece la computadora). En la mayoría de los casos estos dos hechos no son separables.

La elección de la representación de los datos es difícil y no se determina sólo por los recursos de que se dispone, sino que debe tomarse en cuenta las operaciones que se realizan con los datos. Se

sabe que las computadoras utilizan una representación interna basada en dígitos binarios (0 y 1), aunque esta representación sea inadecuada para los seres humanos debido al número grande de cifras comprendidas, pero es adecuada para los circuitos electrónicos del computador que trabaja con dos posibles estados que son presencia o ausencia de corrientes eléctricas, carga eléctrica o campos magnéticos.

Los datos deben:

- Representarse y almacenarse en cierta forma para accederlos posteriormente.
- Organizarse de manera adecuada para accederlos selectiva y eficientemente.
- Procesarse y presentarse de tal manera que puedan apoyar eficientemente al usuario.
- Protegerse y manejarse para que no pierdan su valor.

El tipo de datos.

En matemáticas se acostumbra clasificar las variables de acuerdo con ciertas características. Se hacen distinciones claras entre variables reales, complejas y lógicas, o bien entre variables que representan valores individuales, conjunto de valores o conjunto de conjuntos; o entre funciones, funcionales o conjunto de funciones. Nos apegaremos al hecho de que toda constante, variable, expresión o función es de cierto tipo.

El tipo asociado a un dato se hace explícito en una declaración de constante, variable, expresión o función, y que ésta declaración preceda a la utilización y aplicación de la misma. Evidentemente la cantidad de almacenamiento asignada a un dato tendrá que ser elegido dependiendo del tamaño del conjunto de valores que pueda tomar. De tal manera que el tipo de un dato determina:

- El conjunto de valores al cual pertenece una constante, o que puede tomar una variable o expresión, o que puede generar una función.
- La cantidad de memoria necesaria para su almacenamiento.
- El conjunto de operaciones válidas sobre los mismos. El tipo de un valor representado por una constante, variable o expresión puede derivarse de su forma o de su declaración sin que necesite ejecutarse un proceso de calculo. Cada operador o función aguarda la llegada de argumentos de un tipo fijo y da como resultado un tipo fijo. Si un operador admite argumentos de varios tipos, entonces el tipo de resultado se puede determinar de las reglas específicas del lenguaje.

Tipos de datos primitivos.

Se refiere a tipos de datos básicos o elementales, a partir de las cuales se forman estructuras de datos. Los tipos de datos primitivos considerados en los lenguajes de programación son: enteros, reales, lógicos y los caracteres.

- ◆ **El tipo entero (INTEGER)** comprende un subconjunto de números enteros cuyo tamaño puede variar entre los diversos sistemas de computación. Es un miembro del siguiente conjunto de números:

$$\{ \dots, -(n+1), -n, -(n-1), \dots, -2, -1, 0, 1, 2, \dots, (n-1), n, (n+1), \dots \}$$

Si una computadora utiliza n bits para representar un entero en complemento de 2, entonces los valores posibles de x están entre -2^{n-1} y $2^{n-1} - 1$. Las operaciones aritméticas sobre enteros se consideran exactas, es decir, el resultado es otro entero.

- ◆ **El tipo real (REAL)** representa un subconjunto de los números reales, denominado conjunto punto flotante. La aritmética de los números reales no producen resultados exactos, debido a

los problemas de redondeo ocasionados por operaciones con números finitos de cifras. Para este tipo de datos son válidas todas las operaciones aritméticas: suma, resta, multiplicación, división, potencia, etc.

- ◆ **El tipo lógico (BOOLEAN)** representa un datos que puede tener uno de dos estados posibles: verdadero y falso, si y no, activado y desactivado, etc. Estos valores se representan con las constantes: TRUE y FALSE. Las operaciones validas sobre los datos lógicos o booleano se denominan comparaciones lógicas, las cuales son: conjunción (AND), disyunción (ORA) y la negación (NO) lógica.
- ◆ **El tipo carácter (CHAR)** comprende un conjunto de caracteres imprimibles: las letras mayúsculas y ninusculas, los dígitos, el espacio en blanco, los caracteres de puntuación (+ - * / ! @ # \$ % ^& _ = ...) y un conjunto de caracteres gráficos.
- ◆ **Tipo cadena** es una secuencia finita de símbolos tomados del tipo primitivo carácter incluyendo la cadena nula o vacía. Por lo general el inicio y el final de una cadena lo delimitaremos con un apóstrofe ('). Cada cadena tiene un atributo llamado longitud, el cual es el número de caracteres en la misma.