

Índice

Codificación de texto	1
---------------------------------	---

En esta parte de la unidad veremos la forma de representar otras clases de información no numérica, como los textos y las imágenes.

Codificación de texto

Cuando escribimos texto en nuestra computadora, estamos almacenando temporalmente en la memoria una cierta secuencia de números que corresponden a los **caracteres**, o símbolos que tipeamos en nuestro teclado.

Estos caracteres tienen una **representación gráfica** en nuestro teclado, en la pantalla o en la impresora, pero mientras están en la memoria no pueden ser otra cosa que **bytes**, es decir, conjuntos de ocho dígitos binarios.

Para lograr almacenar caracteres de texto necesitamos adoptar una codificación, es decir, una tabla que asigne a cada carácter un patrón de bits fijo.

Esta codificación debe ser universal: para poder compartir información entre usuarios, o entre diferentes aplicaciones, se requiere algún estándar que sea comprendido y respetado por todos los usuarios y las aplicaciones.

Inicialmente se estableció con este fin el **código ASCII**, que durante algún tiempo fue una buena solución. El código ASCII relaciona cada secuencia de **siete bits** con un carácter (o **grafema**) específico de la **Tabla ASCII**.

Es decir que hay $2^7 = 128$ posibles caracteres codificados por el código ASCII.

Sin embargo, el código ASCII es insuficiente para muchas aplicaciones: no contempla las necesidades de diversos idiomas. Por ejemplo, nuestra letra Ñ no figura en la tabla ASCII. Tampoco las vocales acentuadas, ni con diéresis, como tampoco decenas de otros caracteres de varios idiomas europeos. Peor aún, con solamente 128 posibles patrones de bits, es imposible representar algunos idiomas orientales como el chino, que utilizan miles de ideogramas.

Por este motivo se estableció más tarde una familia de nuevos estándares, llamada Unicode. Uno de los estándares o esquemas de codificación definidos por Unicode, el más utilizado actualmente, se llama **UTF-8**. Este estándar mantiene la codificación que ya empleaba el código ASCII para su conjunto de caracteres, pero agrega códigos de dos, tres y cuatro bytes para otros símbolos. El resultado es que hoy, con UTF-8, se pueden representar todos los caracteres de cualquier idioma conocido.

Otro estándar utilizado, **ISO/IEC 8851**, codifica los caracteres de la mayoría de los idiomas de Europa occidental.

El código ASCII, los diferentes esquemas de Unicode, y el estándar ISO/IEC 8851, coinciden en la codificación de las letras del alfabeto inglés, que son comunes a

la mayoría de los idiomas occidentales, y en la codificación de símbolos usuales como los dígitos, símbolos matemáticos, y otros. Por este motivo son relativamente compatibles, aunque cuando el texto utiliza otros caracteres aparecen diferencias.