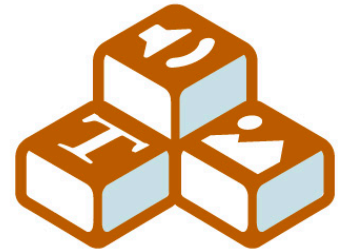


Información Digital

Representación y Codificación



Presentación del Módulo 2: Textos digitales.

Alberto Prieto

Departamento de Arquitectura y Tecnología de Computadores

Universidad de Granada

Una de las formas convencionales de intercambiar información entre los humanos es el **lenguaje escrito**; siendo esta misma forma la más tradicional y usual de comunicarnos con los sistemas digitales. El lenguaje escrito se basa en la composición de textos con ayuda de un alfabeto o conjunto de símbolos, que denominaremos **caracteres**. Consideramos distintos tipos o subconjuntos de caracteres, como las letras del abecedario, las 10 cifras numéricas, los signos de puntuación, los signos matemáticos, etc.

A cada carácter se le asocia un patrón de bits, realizándose esta transformación por medio de **códigos de entrada/salida**. Así, en un código determinado a cada carácter le puede corresponder un patrón de, por ejemplo, 8 bits. Obviamente cada carácter se codifica con un patrón distinto. Un código concreto de entrada/salida se define, sencillamente mediante una tabla que contiene todos los caracteres posibles y el patrón asociado a cada uno de ellos.

En principio podemos definir de forma arbitraria los códigos para textos, y de hecho así ocurría en las primeras décadas del desarrollo de los computadores, cuando cada fabricante de computadores establecía su propio código. Esto dificultaba el intercambio de datos entre distintos computadores cuando eran de constructores distintos. Afortunadamente este problema en gran parte ha desaparecido ya que se han definido por diversos organismos códigos estándar que son ampliamente reconocidos en el mundo digital.

En este tema, explicaremos con más detalle los conceptos anteriores, así como describiremos los códigos de entrada/salida normalizados más relevantes que se utilizan en la actualidad. También analizaremos el origen de algunos problemas que se presentan por la incompatibilidad entre distintos sistemas de codificación y que, por ejemplo, pueden provocar que algunos de los correos electrónicos que recibimos contengan algunas palabras ilegibles.

En resumen, los contenidos de este tema son los siguientes:

- RD2.1 Tipos de caracteres. (2:35)
- RD2.2 Codificación de caracteres. (2:35)
- RD2.3 Códigos ASCII. (7:19)
- RD2.4 Estándar UNICODE. (6:51)
- RD2.5 Codificación UTF-8. (6:25)
- RD2.6 Legibilidad de e-mails. (4:29)

