

## Bài 2: THÔNG TIN VÀ DỮ LIỆU

### 1. Khái niệm thông tin và dữ liệu

- Thông tin là những hiểu biết có thể có được về một thực thể nào đó.
- Trong tin học, dữ liệu là thông tin đã được đưa vào máy tính.

### 2. Đơn vị đo lường thông tin

- Bit là đơn vị nhỏ nhất để đo lường thông tin, sử dụng hai ký hiệu là 0 và 1 để biểu diễn thông tin trong máy tính.
- Các đơn vị khác để đo thông tin:

1 Byte (1B) = 8 Bit  
1 KB (Kilobyte) = 1024B  
1 MB (Megabyte) = 1024KB  
1 GB (Gigabyte) = 1024MB  
1 TB (Têgabyte) = 1024GB  
1 PB (Pêtabyte) = 1024TB

### 3. Các dạng thông tin

Bao gồm: loại số( số nguyên, số thực,..) và loại phi số( văn bản, hình ảnh, âm thanh,..)

a) Dạng văn bản: báo, sách, vở, ...



hình 3. Tờ báo – thông tin dạng văn bản

b) Dạng hình ảnh: tranh vẽ, ảnh chụp, bản đồ,...



**hình 4. biển báo - thông tin dạng hình ảnh**

c) Dạng âm thanh: tiếng nói con người, tiếng sóng biển, tiếng đàn,...

#### **4. Mã hóa thông tin trong máy tính**

- Để máy tính xử lý được, thông tin cần phải được biến đổi thành dãy bit. Cách biến đổi như vậy gọi là mã hóa thông tin.

- Để mã hóa văn bản dùng mã ASCII (8 bit) gồm 256 ký tự được đánh số từ 0-255, số hiệu này được gọi là mã ASCII thập phân của ký tự.

- Bộ mã Unicode (16 bit) có thể mã hóa 65536 ký tự khác nhau, cho phép thể hiện trong máy tính văn bản của tất cả các ngôn ngữ trên thế giới.

#### **5. Biểu diễn thông tin trong máy tính**

a. Thông tin loại số

##### **• Hệ đếm**

Sử dụng các quy tắc và tập kí hiệu để biểu diễn và xác định các số.

- Hệ đếm La Mã:

+ Không phụ thuộc vào vị trí.

+ Tập kí hiệu: I = 1; V = 5; X = 10; L = 50; C = 100; D = 500; M = 1000.

### • Các hệ đếm dùng trong tin học

- Hệ đếm thập phân( hệ cơ số 10): là hệ dùng các số 0, 1,...,9 để biểu diễn.

$$\text{VD: } 536,4 = 5 \times 10^2 + 3 \times 10^1 + 6 \times 10^0 + 4 \times 10^{-1}$$

- Hệ nhị phân: là hệ chỉ dùng 2 số 0 và 1 để biểu diễn.

$$\text{Vd: } 110_2 = 1 \times 2^2 + 1 \times 2^1 + 0 \times 2^0 = 6_{10}$$

- Hệ cơ số mười sáu:

+ Sử dụng các kí hiệu: 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, A, B, C, D, E, F trong đó A, B, C, D, E, F có giá trị tương ứng là 10, 11, 12, 13, 14, 15 trong hệ thập phân.

$$\text{Vd: } A01_{16} = 10 \times 16^2 + 0 \times 16^1 + 1 \times 16^0 = 2561_{10}$$

### b. Thông tin loại phi số

• Văn bản: để biểu diễn một xâu ký tự máy tính có thể dùng một dãy byte, mỗi byte biểu diễn một ký tự theo thứ tự từ trái sang phải.

VD: biểu diễn xâu ký tự “TIN”: 01010100 01001001 01001110

• Các dạng khác: (hình ảnh, âm thanh,...) ta cũng phải mã hóa chúng thành dãy bit.

### • Nguyên lí mã hóa nhị phân

Thông tin có nhiều dạng khác nhau như số, văn bản, hình ảnh, âm thanh,... Khi đưa vào máy tính, chúng đều biến đổi thành dạng chung - dãy bit. Dãy bit đó là mã nhị phân của thông tin mà nó biểu diễn.