

Bài 2: THÔNG TIN VÀ DỮ LIỆU

I. Khái niệm thông tin và dữ liệu:

1. Thông tin: là sự hiểu biết của con người về một thực thể nào đó, có thể thu nhập, lưu trữ, xử lý được

Ví dụ: Bạn Hoa 16 tuổi, nặng 50 kg, học giỏi, chăm ngoan, ... đó là thông tin về Hoa.

2. Dữ liệu : là thông tin đã được đưa vào máy tính.

II. Đơn vị đo thông tin:

- Đơn vị đo của thông tin là bit (binary digit).
- Trong tin học, thuật ngữ bit thường dùng để chỉ phần nhỏ nhất của bộ nhớ máy tính để lưu trữ một trong hai ký hiệu, được sử dụng biểu diễn thông tin trong máy tính, là 0 và 1.

Ngoài ra, người ta còn dùng các đơn vị cơ bản khác để đo thông tin:

Tên gọi	Ký hiệu	Giá trị
Byte	B	8 bit
KiloByte	KB	1024Byte
MegaByte	MB	1024KB
GigaByte	GB	1024MB
TetraByte	TB	1024GB
PetaByte	PB	1024TB

III. Các dạng thông tin:

- Có thể phân loại thông tin thành loại số (số nguyên, số thực, ...) và phi số (văn bản, hình ảnh, ...).
- Một số dạng thông tin phi số:
 - Dạng văn bản: báo chí, sách, vở ...
 - Dạng hình ảnh: bức tranh vẽ, ảnh chụp, băng hình, ...
 - Dạng âm thanh: tiếng nói, tiếng chim hót, ...

IV. Mã hoá thông tin trong máy tính:

- Muốn máy tính xử lý được, thông tin phải được biến đổi thành một dãy bit. Cách biến đổi như vậy gọi là một cách mã hoá thông tin.
- Để mã hóa thông tin dạng văn bản, ta chỉ cần mã hóa các ký tự, thường dùng bảng mã ASCII hoặc bảng mã UNICODE.

- Bảng mã ASCII gồm 256 kí tự, dùng 8 bit để mã hóa.
- Bảng mã UNICODE gồm 2^{16} kí tự, dùng nhiều hơn 8 bit để mã hóa.

Ví dụ: Mã hóa thông tin dãy bóng đèn tắt (T) hoặc sáng (S):

TSSTSTTS $\rightarrow$ 01101001

Ví dụ: Mã hóa kí tự A:

Mã ASCII thập phân: 65

Mã ASCII nhị phân là: 01000001

(Xem Phụ lục 1. BỘ MÃ ASCII CƠ SỞ sách giáo khoa Tin học 10 trang 169)

V. Biểu diễn thông tin trong máy tính:

1. Thông tin loại số:

a) Hệ đếm: Là tập hợp các kí hiệu và qui tắc sử dụng tập kí hiệu đó để biểu diễn và xác định giá trị các số.

– Có hệ đếm phụ thuộc vị trí (Hệ đếm La Mã: kí hiệu: I = 1, V = 5, X = 10, L = 50, C = 100, D = 500, M = 1000) và hệ đếm không phụ thuộc vị trí (hệ thập phân, hệ nhị phân,...)

b) Biểu diễn số trong các hệ đếm:

• **Hệ thập phân:** Kí hiệu: 0, 1, 2, ..., 9.

- Giá trị của mỗi chữ số phụ thuộc vào vị trí của nó trong biểu diễn.

- Qui tắc: Mỗi đơn vị ở 1 hàng bất kì có giá trị bằng 10 đơn vị của hàng kế cận ở bên phải.

Ví dụ: $125_{10} = 1 \times 10^2 + 2 \times 10^1 + 5 \times 10^0$

• **Hệ nhị phân:** (cơ số 2) chỉ dùng 2 kí hiệu là chữ số 0 và 1.

Ví dụ: $1011_2 = 1 \cdot 2^3 + 0 \cdot 2^2 + 1 \cdot 2^1 + 1 \cdot 2^0 = 11_{10}$.

• **Hệ 16:** (hệ Hexa): sử dụng các kí hiệu: 0, 1, ..., 9, A, B, C, D, E, F trong đó A, B, C, D, E, F có các giá trị tương ứng là 10, 11, 12, 13, 14, 15 trong hệ thập phân.

Ví dụ: $2AC_{16} = 2 \cdot 16^2 + 10 \cdot 16^1 + 12 \cdot 16^0 = 684$

c) Biểu diễn số nguyên:

Số nguyên có thể có dấu hoặc không có dấu. Ta có thể chọn 1 byte (= 8 bit), 2 byte, 3 byte... để biểu diễn.

d) Biểu diễn số thực: số thực dấu phẩy tĩnh và số thực dấu phẩy động.

Số thực dấu phẩy động có dạng: $\pm M \times 10^{\pm k}$

Trong đó: $0.1 \leq M < 1$ là phần định trị; k là số nguyên không âm, gọi là phần bậc.

Ví dụ: $13456.25 = 0.1345625 \times 10^5$

2. Thông tin loại phi số:

- Văn bản: Máy tính có thể dùng một dãy byte, mỗi byte biểu diễn một kí tự theo thứ tự từ trái sang phải.

Ví dụ: dãy 3 byte: **01010100** **01001001** **01001110**
 T **I** **N**

biểu diễn xâu: “TIN”

- Các dạng khác: Để xử lí âm thanh, hình ảnh ta cũng phải mã hóa chúng thành các dãy bit.

• Nguyên lý mã hoá nhị phân:

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

BÀI TẬP TỰ LUYỆN

A. Câu hỏi tự luận:

1. Hãy nêu một vài ví dụ về thông tin. Với mỗi thông tin đó, hãy cho biết dạng của nó.
2. Hệ đếm cơ số 16 sử dụng các kí hiệu nào?
3. Phát biểu “Ngôn ngữ máy tính là ngôn ngữ nhị phân (chỉ dùng kí tự 0 và 1)” là đúng hay sai? Hãy giải thích.

B. Câu hỏi trắc nghiệm:

Câu 1: Một quyển sách A gồm 200 trang nếu lưu trữ trên đĩa chiếm khoảng 5MB. Hỏi 1 đĩa cứng 40GB thì có thể chứa khoảng bao nhiêu cuốn sách có dung lượng thông tin xấp xỉ cuốn sách A?

- A. 8000
- B. 8129
- C. 8291
- D. 8192

Câu 2: Chọn câu đúng trong các câu sau:

- A. 1MB = 1024KB
- B. 1B = 1024 Bit
- C. 1KB = 1024MB
- D. 1Bit = 1024B

Câu 3: Thông tin khi đưa vào máy tính, chúng đều được biến đổi thành dạng chung đó là:

- A. Hình ảnh
- B. Văn bản
- C. Dây bit
- D. Âm thanh

Câu 4: Phát biểu nào sau đây là phù hợp nhất về khái niệm bit?

- A. Đơn vị đo khối lượng kiến thức
- B. Chính chữ số 1
- C. Đơn vị đo lượng thông tin
- D. Một số có 1 chữ số

Câu 5: Tại sao phải mã hoá thông tin?

- A. Để thay đổi lượng thông tin
- B. Làm cho thông tin phù hợp với dữ liệu trong máy
- C. Để chuyển thông tin về dạng câu lệnh của ngôn ngữ máy
- D. Tất cả đều đúng

Câu 6: Đơn vị đo lượng thông tin cơ sở là:

- A. Byte
- B. Bit
- C. GB
- D. GHz

Câu 7: Mã hoá thông tin là quá trình:

- A. Đưa thông tin vào máy tính
- B. Chuyển thông tin về bit nhị phân
- C. Nhận dạng thông tin
- D. Chuyển dãy hệ nhị phân về hệ đếm khác

Câu 8: Thông tin là gì?

- A. Các văn bản và số liệu
- B. Hiểu biết của con người về một thực thể, sự vật, khái niệm, hiện tượng nào đó
- C. Văn bản, hình ảnh, âm thanh
- D. Hình ảnh, âm thanh