

## Bài 4.

# BIỂU DIỄN VĂN BẢN, HÌNH ẢNH, ÂM THANH TRONG MÁY TÍNH

Học xong bài này, em sẽ:

- Biết được *bit* là gì.
- Biết được mỗi kí tự, mỗi văn bản được biểu diễn như thế nào trong máy tính.
- Biết được thế nào là số hoá dữ liệu.

### 1. Khái niệm bit



1

Câu lạc bộ “Xanh – Đỏ” của các chú Vẹt máy có quy ước như sau: Mỗi chú Vẹt thành viên phải đeo thẻ gồm một số ô tròn cạnh nhau, mỗi ô tròn có màu “xanh” hoặc màu “đỏ”.

Ban đầu câu lạc bộ chỉ có bốn thành viên, thẻ chỉ gồm hai ô tròn và bốn chú Vẹt khác nhau đeo bốn thẻ khác nhau (Hình 1):



Hình 1. Bốn chú Vẹt đeo thẻ khác nhau

- 1) Có thể tạo ra một thẻ mới (vẫn chỉ gồm hai ô tròn) cho thành viên thứ năm hay không? Chú ý rằng thẻ mới phải khác với bốn thẻ đã có.
- 2) Nếu mỗi thẻ gồm ba ô tròn thì có thể tạo được năm thẻ khác nhau cho năm chú Vẹt không?



Nếu thay màu xanh bằng kí hiệu “0” và màu đỏ bằng kí hiệu “1” thì mỗi thành viên câu lạc bộ có một thẻ số. Dù câu lạc bộ có bao nhiêu chú Vẹt thì vẫn có thể đảm bảo các thẻ là khác nhau, miễn là dãy kí hiệu “0”, “1” đủ dài. Điều này gợi ý hình thành khái niệm *bit*, một khái niệm cơ bản trong tin học.

■ *Bit*: đơn vị nhỏ nhất để biểu diễn và lưu trữ thông tin. Bit chỉ có thể nhận một trong hai trạng thái, kí hiệu là “0” và “1”.

## 2. Biểu diễn chữ cái và văn bản trong máy tính

Máy tính biểu diễn tất cả các chữ cái bằng dãy bit. Ví dụ, chữ cái A được biểu diễn bởi dãy 01000001, chữ cái B được biểu diễn bởi dãy 01000010.

Văn bản có thể chứa cả hình vẽ, hình ảnh. Văn bản chữ (không chứa hình vẽ, hình ảnh) không chỉ có các chữ cái mà còn nhiều phần tử khác nữa: các chữ số, dấu cách, dấu chính tả, dấu xuống dòng, kí hiệu khác,...

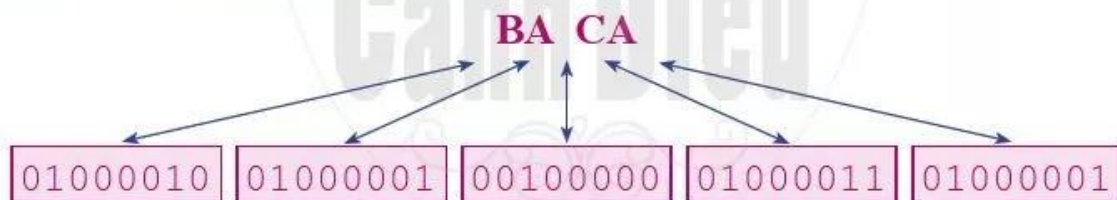
Các nhà tin học sử dụng thuật ngữ “*kí tự*” để bao gồm chung tất cả các phần tử có trong văn bản chữ nhằm tránh phải liệt kê chi tiết dài dòng. Có thể nói khi soạn thảo văn bản, cái mà em gõ nhập từ bàn phím chính là kí tự.

■ *Kí tự*: tên gọi chung cho chữ cái, chữ số, dấu cách, dấu chính tả, kí hiệu khác.

Giống như với các chữ cái, người ta cũng biểu diễn mỗi kí tự không là chữ cái bằng một dãy bit xác định; hai kí tự khác nhau thì biểu diễn bằng hai dãy bit khác nhau.

Nếu xếp lần lượt các dãy bit biểu diễn các kí tự theo trình tự từ trái sang phải, từ trên xuống dưới như khi đọc văn bản, ta sẽ được một biểu diễn văn bản bằng một dãy bit (Hình 2).

Trong máy tính, mỗi kí tự được biểu diễn bằng một dãy bit trong ứng xác định, mỗi văn bản được biểu diễn bằng một dãy bit.



Hình 2. Các dãy bit xếp thành “BA CA”

## 3. Số hoá văn bản, hình ảnh, âm thanh

Một văn bản trong máy tính là một dãy bit. *Số hoá văn bản là việc chuyển văn bản thành dãy bit.*

■ *Văn bản số*: kết quả số hoá một văn bản.

Trong máy tính, một hình ảnh cũng được biểu diễn bằng dãy liên tiếp các kí hiệu “0” hoặc “1”, tức là dãy bit.





2

Em có hình một cái bát màu đen trên nền trắng hình chữ nhật. Chia hình này thành nhiều ô nhỏ bằng một lưới ô vuông, ví dụ bằng lưới  $4 \times 8$ . Như vậy 32 ô vuông nhỏ, màu đen hoặc trắng sẽ tạo thành (gần đúng) hình cái bát.

Hãy thay ô màu đen bằng kí hiệu “1”, ô màu trắng bằng kí hiệu “0”, xếp liên tiếp theo trình tự từ trái sang phải, từ trên xuống dưới. Kết quả nhận được là gì?




*Số hoá hình ảnh là việc chuyển hình ảnh thành dãy bit. Kết quả số hoá một hình ảnh là “hình ảnh số”.*

Trong máy tính, một đoạn âm thanh cũng được biểu diễn bằng dãy các kí hiệu “0” hoặc “1” liên tiếp. Người ta chia đoạn âm thanh theo thời gian thành nhiều đoạn rất ngắn; đặt mỗi đoạn rất ngắn ấy tương ứng duy nhất với một nhóm bit xác định.

*Số hoá âm thanh là việc chuyển đoạn âm thanh thành dãy bit. Kết quả số hoá một đoạn âm thanh là “âm thanh số”.*

“Dữ liệu số hoá” hay ngắn gọn “dữ liệu số” là tên gọi chung cho dữ liệu đã được chuyển thành dãy bit để máy tính có thể xử lí.

*Chú ý: Dữ liệu số là nói tắt của dữ liệu số hoá, tránh hiểu lầm là dữ liệu số lượng.*

 **Số hoá dữ liệu:** chuyển dữ liệu thành dãy bit, tức là dãy các kí hiệu “0” hoặc “1” liên tiếp, để máy tính có thể xử lí.



Có thể dùng dãy bit để biểu diễn chữ cái. Em hãy đề xuất ba dãy bit khác nhau có cùng độ dài là bốn bit để biểu diễn ba chữ cái nào đó.



Một điệp viên hẹn gặp người cung cấp tin tại địa điểm bí mật. Để phòng ngừa mọi rủi ro, họ quy ước: Trước giờ gặp, người cung cấp tin nhìn lên cửa sổ căn phòng X; nếu ở cửa sổ có lọ hoa thì cuộc gặp diễn ra bình thường, còn không thì cuộc gặp bị huỷ. Có thể coi đây là cách truyền tin chỉ dùng một bit hay không? Vì sao?



Trong các câu sau, câu nào đúng?

- 1) Trong máy tính, mỗi kí tự là một dãy bit, hai kí tự khác nhau thì hai dãy bit biểu diễn chúng khác nhau.
- 2) Một bit là một dãy gồm nhiều số 0 và số 1 ghép lại để biểu diễn một chữ cái.
- 3) Số hoá dữ liệu là chuyển dữ liệu thành các số.
- 4) Ảnh số là kết quả số hoá dữ liệu ảnh.
- 5) Văn bản số là trang in gồm toàn các số.
- 6) Âm thanh số là kết quả số hoá dữ liệu âm thanh.

## TÓM TẮT BÀI HỌC

- ❖ Bit chỉ có thể nhận một trong hai trạng thái, kí hiệu là “0” và “1”.
- ❖ Tên gọi chung cho kí hiệu gõ nhập từ bàn phím khi soạn thảo văn bản là kí tự.
- ❖ Có thể biểu diễn mỗi kí tự bằng một dãy bit xác định.
- ❖ Có thể biểu diễn văn bản bằng dãy bit.
- ❖ Số hoá dữ liệu là việc chuyển dữ liệu thành dãy bit.

## Bài đọc thêm

### MÃ MORSE

Trước đây, khi chưa có điện thoại, để truyền tin đi xa người ta dùng mã Morse (đọc là “móc-xơ”). Mã Morse chỉ dùng hai tín hiệu “ngắn” và “dài” (còn gọi là “tạch” và “tè”) để biểu diễn các chữ cái, chữ số,... (Hình 3). Dãy các tín hiệu “tạch”, “tè” có thể gửi đi bằng cách nháy đèn pin, nháy cần ma-níp máy phát tin nhanh hay chậm hơn.

Nếu thay tín hiệu “ngắn” bằng kí hiệu “1”, tín hiệu “dài” bằng kí hiệu “0”, ta có một cách để biểu diễn các chữ cái, chữ số chỉ dùng hai kí hiệu “1” và “0”.

A	• —	U	• • —
B	— • •	V	• • • —
C	— • — •	W	— • • —
D	— • • •	X	— • — • —
E	•	Y	• • — • —
F	• • • •	Z	— • — • •
G	— • — •		
H	• • • •		
I	• •		
J	• — — —		
K	— • • —	1	— — — —
L	• — • •	2	• • — — —
M	— — •	3	• • • — —
N	— •	4	• • • • —
O	— — —	5	• • • • •
P	• — • —	6	— • • • •
Q	— • — • —	7	— • — • •
R	• — • •	8	— • — • • •
S	• • •	9	— • — • — •
T	— •	0	— — — — —

Hình 3. Mã Morse