



BÀI GIẢNG TIN HỌC CƠ SỞ



BÀI I. THÔNG TIN VÀ XỬ LÝ THÔNG TIN

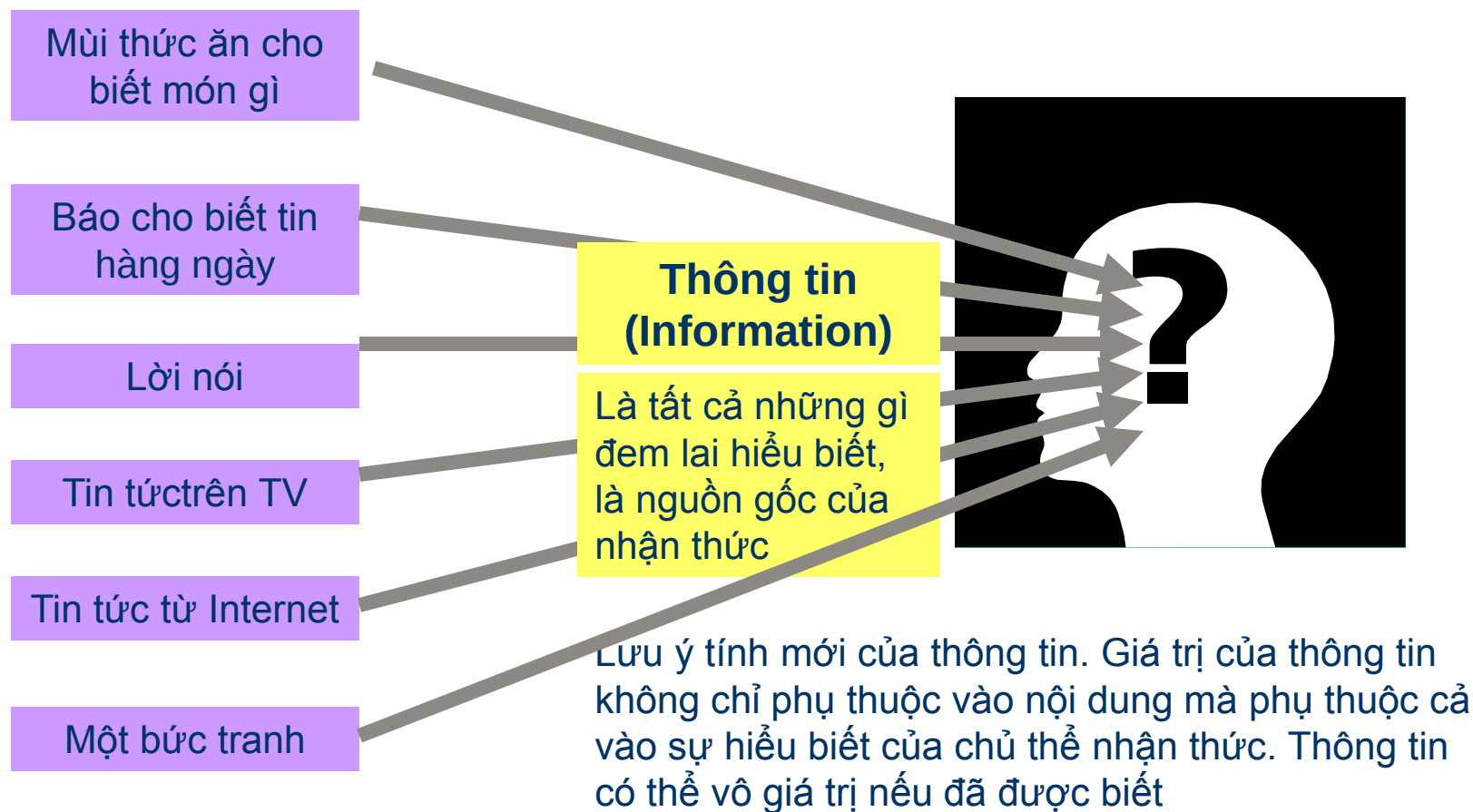


Giảng viên: ĐÀO KIẾN QUỐC
Mobile 098.91.93.980
Email: dkquoc@vnu.edu.vn

NỘI DUNG

- Khái niệm thông tin và các đặc trưng của thông tin
- Mã hoá và lượng hoá thông tin
- Xử lý thông tin bằng máy
- Khái niệm về tin học và công nghệ thông tin
- Câu hỏi và bài tập

THÔNG TIN LÀ GÌ



ĐẶC TRƯNG CỦA THÔNG TIN

Nơi chứa
Giá mang (support)

Giấy, băng từ, đĩa CD..

Hình thức vật lý
Tín hiệu (Signal)

Âm thanh, hình ảnh, mùi,
vị, nhiệt độ...

Tri thức(Knowledge)
có tính khái quát hơn
thông tin. Nó chỉ những
nhận thức có được từ
nhiều thông tin trong
một lĩnh vực nào đó
có tính hướng mục đích.
Tri thức là mục đích
của nhận thức

Dữ liệu là hình thức thể
hiện trong mục đích xử lý
lưu trữ và truyền tin

Ý nghĩa mà thông tin
chuyển tải
Ngữ nghĩa (semantic)

Đặc tính liên tục hay rời rạc – miền giá trị thể hiện
của nó là liên tục hay rời rạc (kể ra được)

MÃ HOÁ

- Mã hoá có mục đích biểu diễn các đối tượng mà vẫn phân biệt được đối tượng khác nhau. Đối với thông tin rời rạc luôn có thể mã hoá được
- Các mã hoá thường dùng là mã hoá trên một tập hợp hữu hạn các kí hiệu (symbol) mà ta gọi là bảng chữ (alphabet). Một từ (word) là một chuỗi hữu hạn các kí hiệu. Để mã hoá, mỗi đối tượng được gán một từ khác nhau. Tính chất này đảm bảo khi biết mã có thể tìm được đối tượng một cách duy nhất.
- Ví dụ đánh số báo danh các thí sinh của một kỳ thi: bảng chữ là tập các chữ số, mỗi thí sinh được mã hoá bằng một số nhiều chữ số. Đặt tên người không phải là một phép mã hoá vì tính không đơn trị của phép đặt tên
- Mã hoá là con đường làm dữ liệu

MÃ HOÁ NHỊ PHÂN

- Nếu bộ chữ chỉ có hai ký hiệu thì phép mã hoá trên đó gọi là mã hoá nhị phân. Ví dụ mã Moorse với hai ký hiệu chấm ■ và vạch ____ là mã nhị phân được biết sớm nhất
- Trong tin học sử dụng bảng chữ nhị phân với hai kí hiệu là $\{0,1\}$
- Nếu sử dụng mã nhị phân có không quá k kí hiệu thì có thể biểu diễn 2^k đối tượng khác nhau. Ví dụ với $k = 3$ có thể có 3 mã 000 001 010 011 100 101 110 111. Ngược lại nếu có n đối tượng thì phải dùng không quá $\lceil \log_2 n \rceil + 1$ ký hiệu để có đủ mã phân biệt các đối tượng
- Mỗi chữ số nhị phân trong một hệ thống mã nhị phân mang một lượng tin nào đó về đối tượng và được lấy làm đơn vị đo lượng tin. Đơn vị đo lượng tin là bit có nguồn gốc từ Binary Digi**T** cũng có nghĩa là “chữ số nhị phân”

CÁC ĐƠN VỊ ĐO LƯỢNG TIN

Đơn vị	Viết tắt	Lượng tin
bít	b	
byte	B	8 bít
Kilô byte	KB	2^{10} B = 1024 B
Mega byte	MB	2^{10} KB
Giga byte	GB	2^{10} MB
Tera byte	TB	2^{10} GB

ĐƠN VỊ ĐO LƯỢNG THÔNG TIN

- Biết thông tin có nghĩa là loại trừ sự “mù mờ”. Có một lượng tin có nghĩa là giảm “độ bất định” hay độ “mù mờ”
- Trong một không gian các đối tượng, độ mù mờ sẽ tăng nếu số đối tượng nhiều và khả năng xác định của một đối tượng là nhỏ. Khả năng xác định đối tượng đo bằng xác suất - tỉ lệ xuất hiện của đối tượng. Xác suất của một đối tượng là một con số nằm trong khoảng $[0, 1]$. Một đối tượng chắc chắn xuất hiện sẽ có xác suất là 1 – khi đó nó hoàn toàn xác định. Một đối tượng không bao giờ xuất hiện có xác suất 0 – nó không xác định.
- Theo Shannon, độ bất định của một không gian các đối tượng có thể đo được. Nếu không gian đối tượng có n đối tượng $A_1, A_2, \dots, A_n$ phân biệt có xác suất lần lượt là $p_1, p_2, \dots, p_n$ ($0 \leq p_i \leq 1, \sum p_i = 1$) và b là một cơ số (số dương nào đó) nào đó thì độ bất định (entropy) được tính bằng

$$H = - \sum p_i \log_b p_i$$

VÍ DỤ

TÍNH ĐỘ BẤT ĐỊNH VÀ LƯỢNG TIN

- Khoa CNTT có 128 cán bộ. Thông tin “một cán bộ nghiên cứu khoa học” rất mù mờ và có độ bất định là:

$$\begin{aligned} H1 &= - \sum p_i \log_b p_i \\ &= - \sum (1/128) \times (\log_2 1/128) \\ &= -128 \times (1/128) \times (-7) = 7 \end{aligned}$$

- Khi biết thêm tin: cán bộ này làm đang nghiên cứu “quy trình thiết kế phần mềm theo mẫu – pattern design” thì ta biết cán bộ này ở bộ môn công nghệ phần mềm (chỉ có 8 cán bộ). Độ bất định khi này là:

$$H2 = - \sum (1/8) \times (\log_2 1/8) = 3$$

- Độ xác định đã tăng nhờ biết thông tin trên. Lượng tin nhận được chính là lượng giảm của độ bất định

$$E = H1 - H2 = 4 \text{ (bit)}$$

LÀM RÕ THÊM ĐƠN VỊ “BIT”

- Định nghĩa chính xác: bit là lượng tin cần thiết để xác định một đối tượng trong không gian có hai trạng thái đồng xác suất ($1/2$). Entropy ban đầu là $H1 = - \sum (1/2) \times (\log_2 1/2) = 1$. Entropy khi hoàn toàn xác định (còn 1 trạng thái với xác suất 1) $H2 = 0$.
- Trong trường hợp có 128 cán bộ, không gian cán bộ cần được mã hoá bởi 7 chữ số nhị phân ($128 = 2^7$)
- Trong trường hợp có 8 cán bộ, không gian cán bộ cần được mã hoá bởi 3 chữ số nhị phân ($8 = 2^3$)
- Vì thế số chữ số nhị phân cần thiết để mã hoá có thể coi là độ bất định của không gian đối tượng và độ đo của lượng tin chính là lượng bit loại trừ được. Chính vì lý do này mà đơn vị đo tin lấy là bit với ý nghĩa gốc là chữ số nhị phân.

XỬ LÝ THÔNG TIN

- Xử lý thông tin là tìm ra những thể hiện mới của thông tin phù hợp với mục đích sử dụng.
- Xử lý thông tin không làm tăng lượng tin mà chỉ hướng hiểu biết vào những khía cạnh có lợi trong hoạt động thực tiễn. Mục đích của xử lý thông tin là tri thức.

XỬ LÝ THÔNG TIN BẰNG MÁY

- Thông tin vào máy cần được mã hoá để máy có thể hiểu được
- Kịch bản xử lý phải được cung cấp trước. Máy không tự hiểu được phải làm gì và làm như thế nào
- Kết quả máy tạo ra cũng là mã (nhị phân)

001101001100100
10010010010101
1101001100101

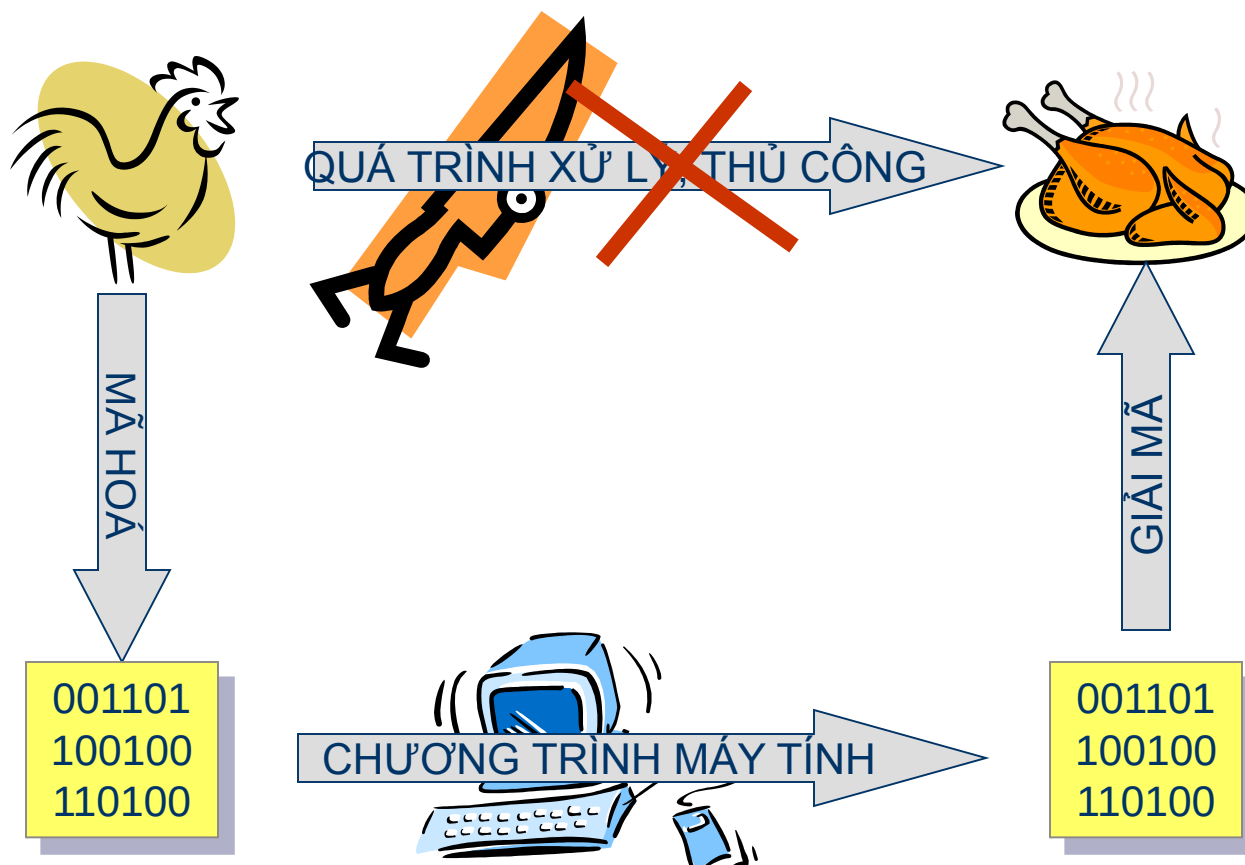
DỮ LIỆU
KỊCH BẢN



001101001100100
100100100010101
110100110010101

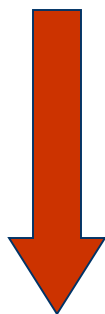
KẾT QUẢ

XỬ LÝ THÔNG TIN BẰNG MÁY



TIN HỌC (INFORMATICS)

Khoa học xử lý thông tin tự động, mà công cụ ngày nay là MTĐT



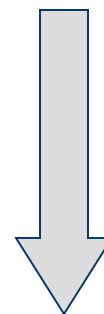
Khía cạnh phương pháp thể hiện qua phần mềm (software)

Các giải pháp tính toán có hiệu quả, kinh tế, phương pháp luận về làm phần mềm



Khía cạnh thiết bị (hardware)

Các công nghệ chế tạo máy tính và các thiết bị có hiệu năng cao, giá thành giảm, các hệ thống tích hợp



Trong tương lai, có thể có các máy tính tự động theo nguyên lý sinh học hay lượng tử

CÔNG NGHỆ THÔNG TIN

Computer Science
Khoa học máy tính

Computer
Máy tính

Information Processing
Xử lý thông tin

Computer
Điện toán

Informatics
Tin học

Computing
Tính toán bằng máy tính

Nghị quyết 49-CP: "Công nghệ Thông tin là tập hợp các phương pháp khoa học, các phương tiện và công cụ kỹ thuật hiện đại - chủ yếu là kỹ thuật máy tính và viễn thông - nhằm tổ chức và khai thác và sử dụng có hiệu quả nguồn tài nguyên thông tin rất phong phú và tiềm tàng trong mọi lĩnh vực hoạt động của con người và xã hội... Công nghệ thông tin được phát triển trên nền tảng phát triển của các công nghệ Tin học-Điện tử- Viễn thông và Tự động hoá".

Luật Công nghệ Thông tin định nghĩa CNTT như sau:

Công nghệ thông tin là tập hợp các phương pháp khoa học, công nghệ và công cụ kỹ thuật hiện đại để sản xuất, truyền đưa, thu thập, xử lý, lưu trữ và trao đổi thông tin số - thông tin được tạo lập bằng phương pháp dùng tín hiệu số.

TỔNG KẾT

- Thông tin: tất cả những gì mang lại hiểu biết, thông tin là nguồn gốc của nhận thức. Thông tin thể hiện qua các hình thức vật lý là tín hiệu
- Thông tin có thể được mã hoá, được biểu diễn theo mục đích sử dụng. Thông thường với mục đích xử lý bằng máy nó có biểu diễn nhị phân.
- Dữ liệu là hình thức biểu diễn của thông tin, có ý nghĩa phụ thuộc vào hoàn cảnh sử dụng. Mã hoá là con đường làm dữ liệu
- Xử lý thông tin có mục đích phát hiện những thể hiện của thông tin hướng vào các hoạt động thực tiễn. Xử lý thông tin không làm tăng lượng tin. Mục đích của xử lý thông tin là tri thức.
- Công nghệ thông tin là tập hợp các phương pháp khoa học, công nghệ và công cụ kỹ thuật hiện đại để sản xuất, truyền đưa, thu thập, xử lý, lưu trữ và trao đổi thông tin số - thông tin được tạo lập bằng phương pháp dùng tín hiệu số.

CÂU HỎI VÀ BÀI TẬP

1. Hãy làm rõ mối liên hệ giữa các khái niệm thông tin, tin hiệu, dữ liệu ?
2. Tìm một ví dụ minh họa có thông tin nghĩa là giảm độ bất định.
3. Một lớp có 48 sinh viên trong đó có 36 nam và 12 nữ. Trong một cuộc thi học sinh giỏi tin học của trường một sinh viên của lớp được giải nhất. Người ta muốn biết người đó là ai. Sau đó người ta được thông báo thêm, người đoạt giải cũng đã từng nhận giải nhì trong một cuộc thi cắm hoa của nữ sinh tổ chức nhân ngày 8/3. Tính lượng tin nhận được trong thông báo trên.
4. Đơn vị đo tin là bit. Nhưng bit chính lại là chữ viết tắt của cụm từ chữ số nhị phân "Binary Digit". Hãy lý giải mối liên hệ giữa hai điều này.
5. Tại sao nói xử lý thông tin không làm tăng lượng tin
6. Hãy nêu vai trò của thông tin trong cuộc sống .

CẢM ƠN ĐÃ THEO DÕI



HẾT BÀI 1. HỎI VÀ ĐÁP

