

# Bài 1

## THÔNG TIN, VÀ XỬ LÝ THÔNG TIN

### NỘI DUNG

- Khái niệm thông tin và các đặc tính 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



# THÔNG TIN LÀ GÌ ?

## Thông tin (Information)

Mùi thức ăn cho biết món gì

Đèn đỏ cho biết không được đi

Tiếng sấm báo hiệu một cơn giông

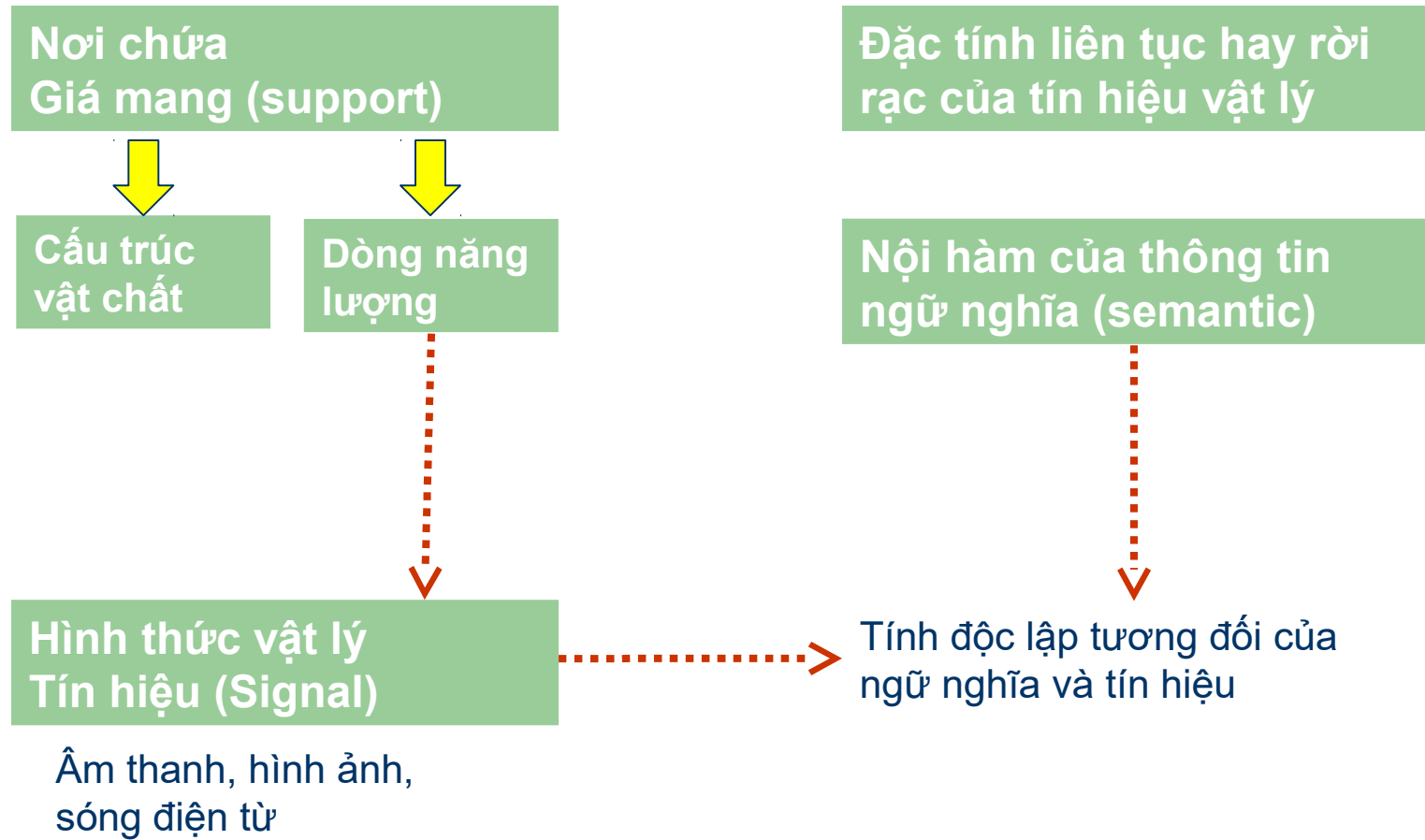
Tin từ TV, Internet cho biết cuộc sống hàng ngày

Bức xạ tàn dư, cho biết lịch sử hình thành vũ trụ



Là tất cả những gì đem lại hiểu biết về thế giới xung quanh,

# ĐẶC TÍNH CỦA THÔNG TIN



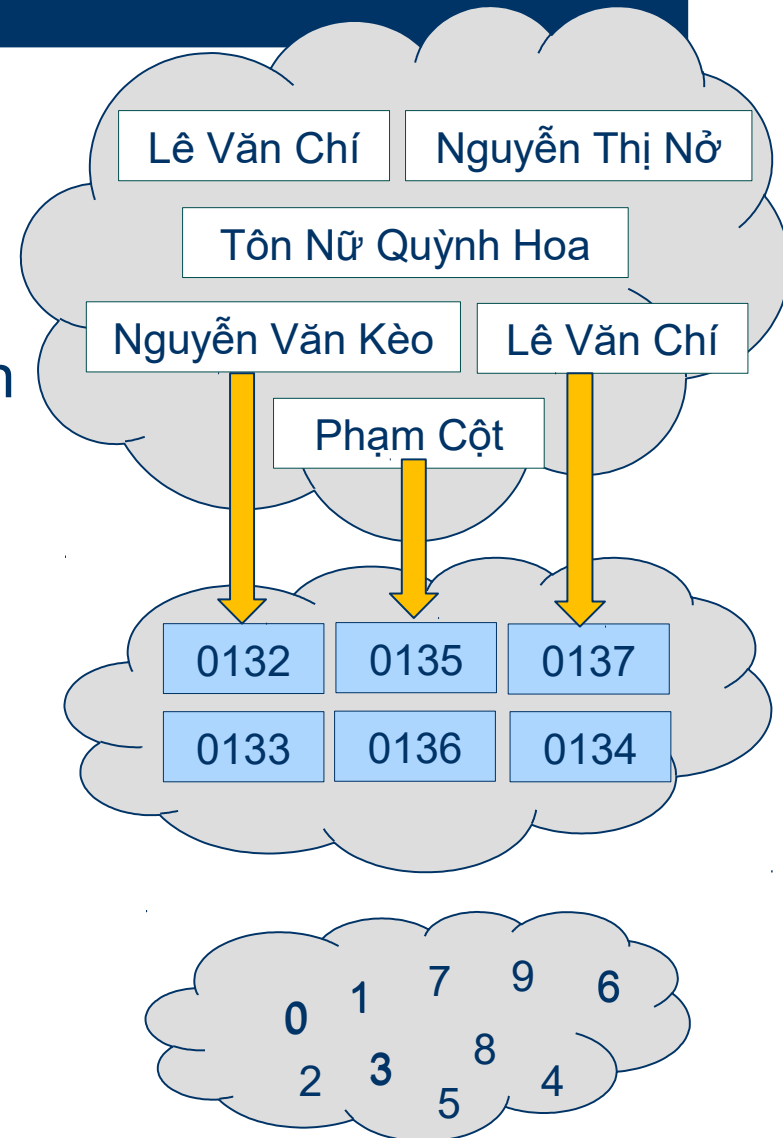
# DỮ LIỆU VÀ TRI THỨC

- Dữ liệu (data) là hình thức thể hiện (biểu diễn) thông tin
- Thuật ngữ
  - Dữ liệu
  - Data (datum)
  - Données
  - Данных

= những cái được cho
- Trong mục đích xử lý, lưu trữ và truyền tin, dữ liệu là đối tượng của máy tính
- Tri thức/ Hiểu biết (knowledge) có được từ một hoặc nhiều thông tin sau khi xử lý (nhận thức)
- Tri thức là mục đích của nhận thức, hướng dẫn hoạt động thực tiễn
- Tri thức bao giờ cũng gắn với chủ thể nhận thức
- Tri thức trở thành thông tin khi nó tiếp tục cung cấp hiểu biết cho một chủ thể nhận thức

# MÃ HOÁ

- Mã hoá có mục đích biểu diễn và phân biệt các đối tượng
- Cách mã hoá
  - Dùng một tập gọi là **bảng chữ** (alphabet) gồm hữu hạn các **kí hiệu** (symbol) mà ta gọi là **chữ cái**
  - Tạo ra tập hợp các **mã** (code) gồm các từ (word). Mỗi **từ** là một chuỗi hữu hạn các chữ cái.
  - Gán cho mỗi đối tượng một mã **khác nhau**. Tính khác nhau này đảm bảo mã xác định được duy nhất một đối tượng.



# MÃ HOÁ

- Mã hoá có mục đích biểu diễn và phân biệt các đối tượng
- Cách mã hoá
  - Dùng một tập gọi là **bảng chữ** (alphabet) gồm hữu hạn các **kí hiệu** (symbol) mà ta gọi là **chữ cái**
  - Tạo ra tập hợp các **mã** (code) gồm các từ (word). Mỗi **từ** là một chuỗi hữu hạn các chữ cái.
  - Gán cho mỗi đối tượng một mã **khác nhau**. Tính khác nhau này đảm bảo mã xác định được duy nhất một đối tượng.
- Mã hoá thực chất là cách làm dữ liệu
- Thông tin rời rạc luôn có thể mã hoá được
- Thông tin liên tục cần được xấp xỉ bởi thông tin rời rạc

# MÃ HOÁ NHỊ PHÂN

- Là mã hóa mà bảng chữ chỉ có hai ký hiệu
  - Mã Moore 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 các chữ số nhị phân {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 8 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.
  - Với 64 đối tượng ta cần mã 6 chữ số nhị phân
  - Với 8 đối tượng ta cần mã 3 chữ số nhị phân

# ĐƠN VỊ ĐO THÔNG TIN

- Đơn vị đo lượng tin là **bit**, có nguồn gốc từ “**B**inary **DigiT**” có nghĩa là “chữ số nhị phân”. Trong mã hóa nhị phân, mỗi chữ số nhị phân sẽ mang một lượng tin nào đó và được lấy làm đơn vị đo lượng tin.
- Khái niệm đối ngẫu của “có thông tin” là độ bất định hay độ “mù mờ” của thông tin.
- Có thông tin tức là giảm được sự “mù mờ”. Đo được độ mù mờ tức là đo được lượng thông tin nhờ tính được mức chênh lệch của độ mù mờ sau khi có thêm thông tin.

# TÍNH ĐỘ BẤT ĐỊNH

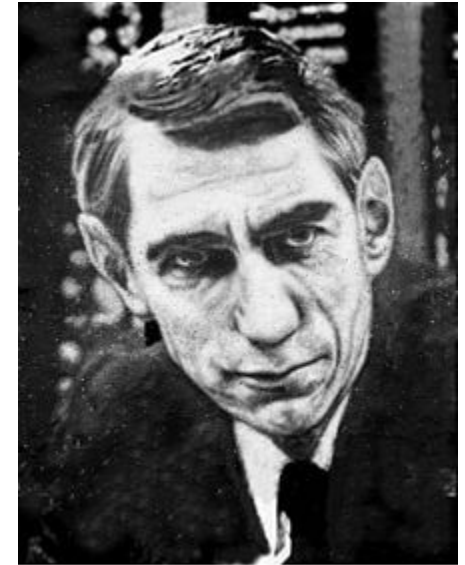
- Độ bất định (entropy) có liên quan đến khái niệm xác suất (probability).
- Xác suất của một sự kiện là độ đo cho khả năng xảy ra của sự kiện
- Xác suất có thể tính bằng tỉ số  $p=m/n$  trong đó  $m$  là số lần sự kiện xảy ra trong tổng số  $n$  lần quan trắc.
  - VD: Xác suất xuất hiện mỗi mặt của một con xúc xắc cân đối và đồng chất là  $1/6$ , xác suất xuất hiện mặt ngửa khi tung một đồng xu là  $1/2$
- Một không gian càng có nhiều sự kiện với xác suất nhỏ (phân tán) thì độ bất định càng lớn (khó đoán trước)

# TÍNH ĐỘ BẤT ĐỊNH

- Claude E. Shannon đã đưa ra một số giả thuyết hợp lý về độ bất định
- Nếu không gian có  $n$  sự kiện (hay 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 nào đó thì hàm số sau thỏa mãn các giả thuyết của Shannon

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

- $H$  được gọi là **entropy** làm độ đo cho độ mù mờ thông tin của không gian các sự kiện





# VÍ DỤ TÍNH LƯỢNG TIN TỪ ENTROPY

- Khoa Công nghệ Thông tin có 64 cán bộ trong đó có Bộ môn Công nghệ phần mềm có 8 cán bộ.
- Thông tin “một cán bộ nghiên cứu khoa học” rất mù mờ và có độ “mù mờ” là:

$$\begin{aligned} H1 &= - \sum p_i \log_b p_i = - \sum (1/64) \times (\log_2 1/64) \\ &= - 64 \times (1/64) \times (-6) = 6 \text{ (đơn vị)} \end{aligned}$$

Nhận xét: entropy của không gian đối tượng trùng với độ dài của mã nhị phân cần thiết để mã hóa tập đối tượng

- Khi biết thêm tin: cán bộ này làm đang nghiên cứu “quy trình thiết kế phần mềm” thì ta biết cán bộ này ở bộ môn công nghệ phần mềm. Độ “mù mờ” khi đó là:

$$H2 = - \sum (1/8) \times (\log_2 1/8) = 3 \text{ (đơn vị)}$$

- Độ bất định giảm được chính là lượng tin của thông tin trên:

# ĐƠN VỊ ĐO THÔNG TIN

- Xét trường hợp không gian chỉ có hai sự kiện  
$$H1 = - \sum (1/2) \times (\log_2 1/2) = 1$$
- Một thông tin xác định hoàn toàn 1 trong 2 sự kiện sẽ đưa không gian về 1 sự kiện duy nhất với xác suất 1. Khi đó  $H2 = 1 \times \log_2 1 = 0$ . Lượng tin của thông tin chính là  $H1 - H2 = 1$ .
- Đơn vị đo thông tin là lượng tin để xác định 1 trong 2 trạng thái có xác suất như nhau. Hệ quả: **Mỗi chữ số nhị phân mang lượng tin 1 bit.**
- Thông tin làm giảm độ bất định cũng trùng hợp với việc giảm số chữ số nhị phân trong mã hóa và độ đo của lượng tin chính là lượng bit loại trừ được.



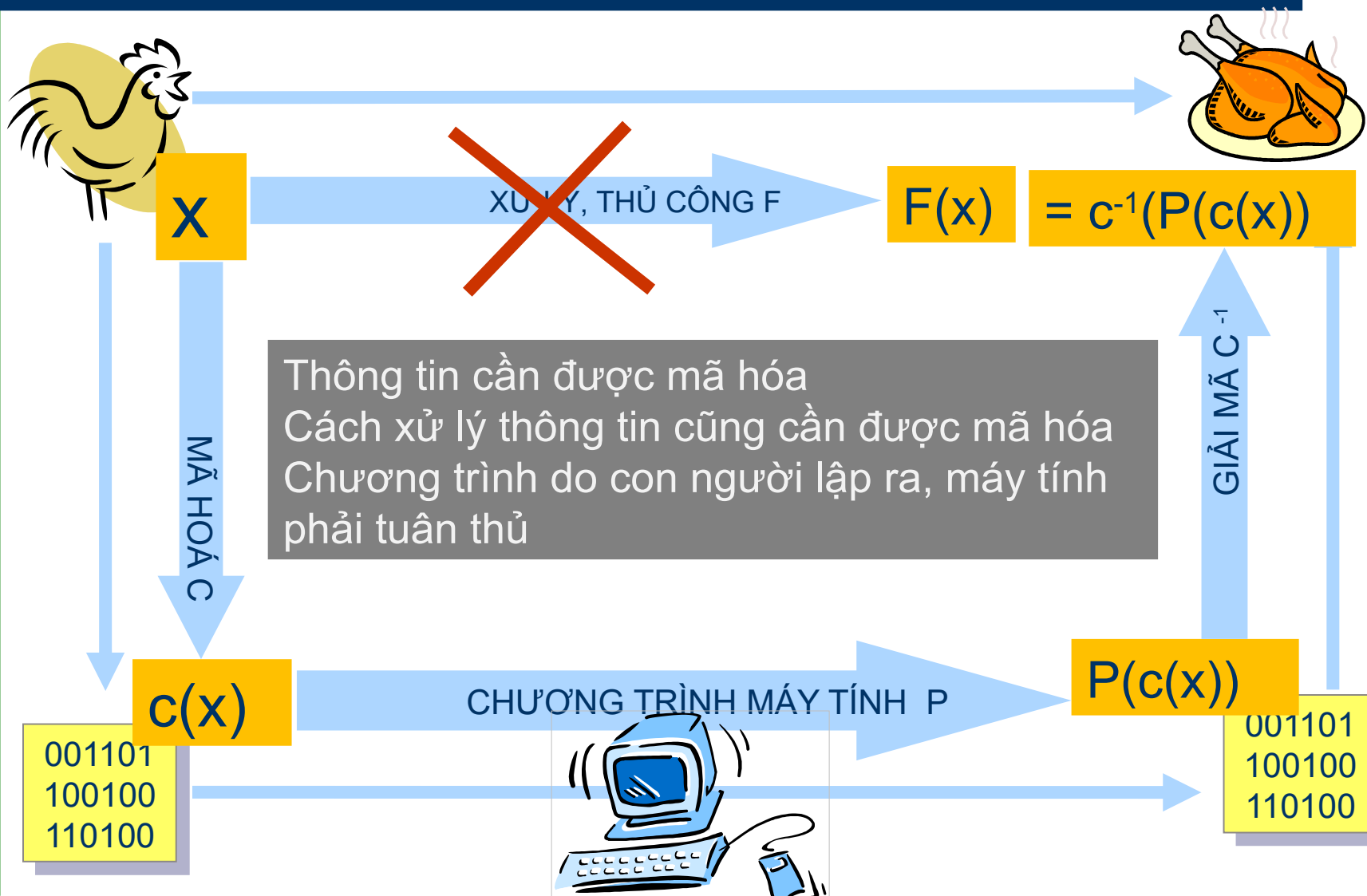
# CÁC ĐƠN VỊ ĐO LƯỢNG TIN

Đơn vị thông tin	Viết tắt	Lượng tin
bít	b	1 bit
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
Peta byte	PB	$2^{10}$ TB

# 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.
  - VD: Về tính bất định, việc cho biết  $b$  và  $c$  của phương trình  $x^2 + b.x + c = 0$  cũng không khác gì cho hai nghiệm  $x_1$  và  $x_2$  (dù có giải hay không)
  - Tìm nghiệm có thể có một mục đích sử dụng khác

# XỬ LÝ THÔNG TIN BẰNG MÁY TÍNH





# TIN HỌC và CÔNG NGHỆ THÔNG TIN

Computer Science

Khoa học máy tính

Điện toán

Computer

Máy tính điện tử

Máy điện toán

Information Processing

Xử lý thông tin

Computing

Tính toán bằng máy tính

Data Processing

Xử lý dữ liệu

Informatics

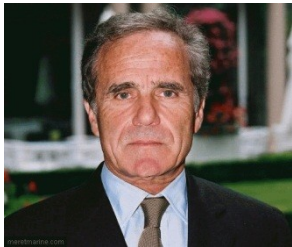
Tin học



VAIP = Vietnam Associate  
for Information Processing

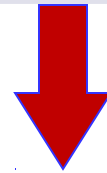
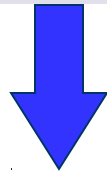
# TIN HỌC - ĐỊNH NGHĨA CỦA Philippe Dreyfus (1962) được công nhận bởi Viện hàn lâm khoa học Pháp

Information + Automatique > Informatique > Informatics



Science du traitement rationnel, notamment par machines automatiques, de l'information considérée comme le support des connaissances humaines et des communications dans les domaines technique, économique et social

**Khoa học xử lý thông tin bằng công cụ tự động**

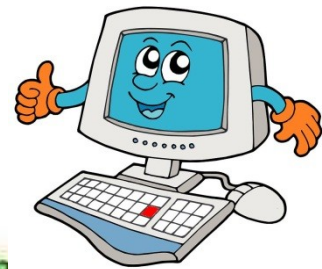


**Khía cạnh phương pháp thể hiện qua phần mềm (software)**

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



Atom B



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

# CÔNG NGHỆ THÔNG TIN

**Nghị quyết 49-CP (1996):** "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".



**Định nghĩa trong luật Công nghệ Thông tin (2005):** 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

**IT** = Information Technology

**và ICT** = Information and Communication Technology



# THÔNG TIN VÀ XỬ LÝ THÔNG TIN

**CẢM ƠN ĐÃ THEO DÕI BÀI GIẢNG**



# 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 .