
THÔNG TIN VÀ XỬ LÝ THÔNG TIN

ThS. LÊ ANH TUẤN

--- oOo ---

	trang ...
I. THÔNG TIN	2
1. Khái niệm về thông tin, phân loại thông tin	2
2. Đơn vị đo thông tin	2
3. Mã hóa thông tin rời rạc	3
II. XỬ LÝ THÔNG TIN	4
1. Sơ đồ tổng quát của một quá trình xử lý thông tin	4
2. Xử lý thông tin bằng máy tính điện tử	5
III. TIN HỌC	5
1. Các lĩnh vực nghiên cứu của tin học	5
2. Ứng dụng của tin học	5
3. Máy tính điện tử và lịch sử phát triển	6

THÔNG TIN VÀ XỬ LÝ THÔNG TIN

ThS. LÊ ANH TUẤN

--- oOo ---

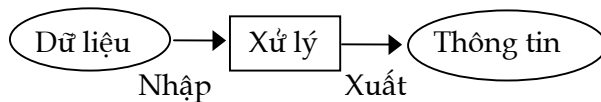
I. THÔNG TIN

1. Khái niệm về thông tin, phân loại thông tin

Dữ liệu (data) là các sự kiện không có cấu trúc, không có ý nghĩa rõ ràng, cho đến khi chúng được tổ chức theo một tiến trình tính toán nào đó.

Thông tin (Information) là một khái niệm trừu tượng được thể hiện qua các thông báo, các biểu hiện ... đem lại một nhận thức chủ quan cho một đối tượng nhận tin. Thông tin là dữ liệu đã được xử lý xong, mang ý nghĩa rõ ràng. Thông tin cũng có thể bị diễn đạt sai lệch, xuyên tạc do tác động cố ý hay vô ý của con người hay sinh vật khác.

Một *hệ thống thông tin* (information system) là một tiến trình ghi nhận dữ liệu, xử lý nó và cung cấp tạo nên dữ liệu mới có ý nghĩa thông tin, liên quan một phần đến một tổ chức, để trợ giúp các hoạt động liên quan đến tổ chức.



2. Đơn vị đo thông tin

Đơn vị dùng để đo thông tin gọi là **bit**. Một bit tương ứng với một chỉ thị hoặc một thông báo nào đó về 1 sự kiện có trong 2 trạng thái có số đo khả năng xuất hiện đồng thời là Tắt(Off) / Mở(On) hay Đúng(True) / Sai(False).

Ví dụ 1. Một mạch đèn có 2 trạng thái là:

- Tắt (Off) khi mạch điện qua công tắc là hở
- Mở (On) khi mạch điện qua công tắc là đóng

Số học nhị phân sử dụng hai số hạng 0 và 1. Vì khả năng sử dụng hai số 0 và 1 là như nhau nên một chỉ thị chỉ gồm 1 chữ số nhị phân có thể xem như là chứa đơn vị thông tin nhỏ nhất.

Bit là chữ viết tắt của **BI**nary digi**T**. Trong tin học, người ta thường sử dụng các đơn vị đo thông tin lớn hơn sau:

Bảng 1. Bảng đơn vị đo thông tin

Tên gọi	Ký hiệu	Giá trị
Byte	B	8 bit
KiloByte	KB	2^{10} B = 1024 Bytes
MegaByte	MB	2^{20} B
GigaByte	GB	2^{30} B
TetraByte	TB	2^{40} B

Năm 1948, nhà bác học Shannon đã đưa ra công thức sau để tính lượng thông tin, bằng cách đo khả năng xuất hiện các sự kiện trong một thông báo, ký hiệu là H và gọi là *Entropi* :

$$H = - \sum_{i=1}^n p_i \cdot \log_2(p_i)$$

trong đó : p_i là xác suất xuất hiện sự kiện thứ i , với $0 \leq p_i \leq 1$ và $\sum_{i=1}^n p_i = 1$

n là số sự kiện lớn nhất có khả năng xuất hiện

Ví dụ 2. Gieo ngẫu nhiên 1 đồng xu, có thể xảy ra 1 trong 2 sự kiện trong mỗi lần gieo là khả năng xuất hiện *mặt sấp* hoặc *mặt ngửa*. Vì đồng xu xem như đồng chất, nên xác suất xuất hiện của mỗi sự kiện là $p = 1/2$. Vậy lượng tin sẽ là :

$$H = - \sum_{i=1}^n p_i \cdot \log_2(p_i) = - \sum_{i=1}^2 \frac{1}{2} \log_2\left(\frac{1}{2}\right) = - \log_2\left(\frac{1}{2}\right) = 1$$

Ví dụ 3. Tương tự như trên nếu ta gieo 1 con súc sắc, thì $n = 6$ và $p = 1/6$

Lượng tin H sẽ là :

$$H = - \sum_{i=1}^n p_i \cdot \log_2(p_i) = - \sum_{i=1}^6 \frac{1}{6} \log_2\left(\frac{1}{6}\right) = - \log_2\left(\frac{1}{6}\right) = 2,5849$$

Ví dụ 4. Một trạm khí tượng X bằng việc phân tích nhiều số liệu khác nhau đã đưa ra dự đoán khả năng có mưa rơi một vùng nào đó vào ngày mai là 75%.

Suy ra khả năng không có mưa sẽ là $100\% - 75\% = 25\%$. Như vậy thông báo trên có lượng tin theo Shannon là :

$$H = - \sum_{i=1}^n p_i \cdot \log_2(p_i) = -(0.75 \cdot \log_2 0.75 + 0.25 \cdot \log_2 0.25) \approx 0.8$$

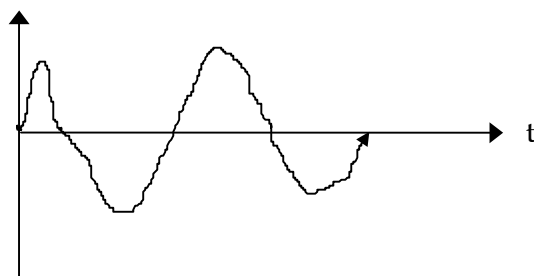
So sánh 3 ví dụ trên, ta thấy *xác suất xuất hiện một tin càng thấp thì lượng tin càng cao* vì mức độ bất ngờ của nó càng lớn và ngược lại. Tuy nhiên, không phải lúc nào ta cũng xác định được xác suất xuất hiện sự kiện nên việc áp dụng công thức Shannon bị hạn chế.

3. Mã hóa thông tin rời rạc

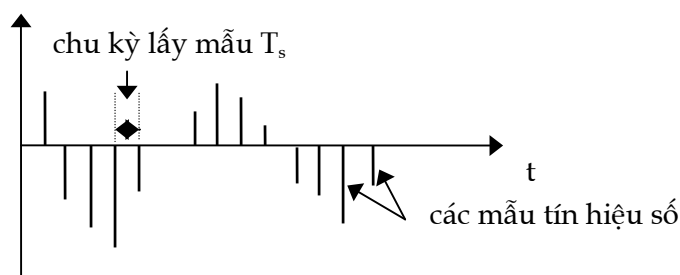
Tất cả các thông tin ở dạng văn bản (text), chữ (character), số (number), ký hiệu (symbol), đồ họa (graphic), hình ảnh (image) hoặc âm thanh (sound) ... đều được gọi là các tín hiệu (signals). Tín hiệu có thể là liên tục hay rời rạc. Máy tính tương tự (Analog Computer) là máy tính chuyên dụng xử lý một số các tín hiệu liên tục như tín hiệu điện, âm thanh... Trong khi đó, hầu hết các dữ liệu mà chúng ta có được thường ở dạng các tín hiệu rời rạc để diễn tả các tín hiệu liên tục qua các số đo hữu hạn. Khi đưa các tín hiệu này vào máy tính, chúng được mã hóa theo các tín hiệu số (digital signal) nhằm giúp máy tính có thể hiểu được thông tin đưa vào. Đây là cơ sở thực tiễn của *nguyên lý mã hoá thông tin rời rạc*. Nguyên lý này tập trung các điểm chủ yếu sau :

- Tín hiệu liên tục có thể xem như một chuỗi xấp xỉ các tín hiệu rời rạc với chu kỳ lấy mẫu nhỏ ở mức độ chấp nhận được (hình 1).

- Tín hiệu rời rạc có thể được đặc trưng qua các bộ ký hiệu hữu hạn (chữ cái, chữ số, dấu, ...) gọi là phép mã hóa (encode) (hình 2). Mọi phép mã hóa đều có thể xây dựng trên bộ ký hiệu các chữ số, đặc biệt chỉ cần bộ ký hiệu gồm 2 chữ số là 0 và 1. Ngược với phép mã hoá gọi là phép giải mã (decode).



Hình 1. Tín hiệu tương tự



Hình 2. Tín hiệu số

Tín hiệu rời rạc là tín hiệu có trục thời gian bị rời rạc hoá với chu kỳ lấy mẫu là $T_s = 1/F_s$, với F_s là tần số lấy mẫu. Tiếng nói con người thường có tần số $F_s = 10$ kHz. Một ví dụ về thông tin rời rạc là hình trên phim khi được chiếu lên màn ảnh là các ảnh rời rạc xuất hiện với tốc độ 25 ảnh/giây. Mắt người không phân biệt sự rời rạc này nên có cảm tưởng hình ảnh là liên tục.

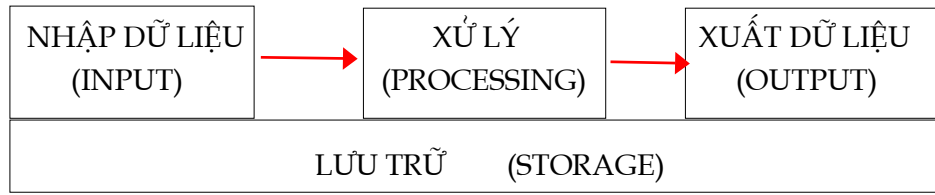
Mã hoá thông tin rời rạc là một khái niệm rất căn bản trong kỹ thuật máy tính.

II. XỬ LÝ THÔNG TIN

1. Sơ đồ tổng quát của một quá trình xử lý thông tin

Mọi quá trình xử lý thông tin bằng máy tính hay bằng con người đều được thực hiện theo một qui trình sau :

Dữ liệu (data) được nhập ở đầu vào (input). Máy tính hay con người sẽ thực hiện quá trình xử lý nào đó để nhận được thông tin ở đầu ra (output). Quá trình nhập dữ liệu, xử lý và xuất thông tin đều có thể được lưu trữ (hình 3).



Hình 3. Mô hình tổng quát quá trình xử lý thông tin

2. Xử lý thông tin bằng máy tính điện tử

Thông tin là kết quả bao gồm nhiều quá trình xử lý các dữ liệu và thông tin có thể trở thành dữ liệu mới để theo một quá trình xử lý khác tạo ra thông tin mới hơn theo ý đồ của con người.

Con người có nhiều cách để có dữ liệu và thông tin. Người ta có thể lưu trữ thông tin qua tranh vẽ, giấy, sách báo, hình ảnh trong phim, băng từ, ... Trong thời đại hiện nay, khi lượng thông tin đến với chúng ta càng lúc càng nhiều thì con người có thể dùng một công cụ hỗ trợ cho việc lưu trữ, chọn lọc và xử lý lại thông tin gọi là *máy tính điện tử* (computer). Máy tính điện tử giúp con người tiết kiệm rất nhiều thời gian, công sức và tăng độ chính xác cao trong việc tự động hoá một phần hay toàn phần của quá trình xử lý dữ liệu hay thông tin.

III. TIN HỌC

1. Các lĩnh vực nghiên cứu của tin học

Tin học (Informatics) được định nghĩa là ngành khoa học nghiên cứu các phương pháp, công nghệ và kỹ thuật xử lý thông tin tự động. Công cụ chủ yếu của tin học là máy tính điện tử và các thiết bị truyền tin khác. Việc nghiên cứu chính của tin học nhằm vào 2 kỹ thuật phát triển song song :

- *Kỹ thuật phần cứng* (hardware engineering): nghiên cứu chế tạo các thiết bị, linh kiện điện tử, công nghệ vật liệu mới... hỗ trợ cho máy tính và mạng máy tính đầy mạnh khả năng xử lý toán học và truyền thông tin.

- *Kỹ thuật phần mềm* (software engineering): nghiên cứu phát triển các hệ điều hành, ngôn ngữ lập trình cho các bài toán khoa học kỹ thuật, mô phỏng, điều khiển tự động, tổ chức dữ liệu và quản lý hệ thống thông tin.

2. Ứng dụng của tin học

Tin học hiện đang được ứng dụng rộng rãi trong tất cả các ngành nghề khác nhau của xã hội từ khoa học kỹ thuật, y học, kinh tế, công nghệ sản xuất đến khoa học xã hội, nghệ thuật,... như:

- Tự động hóa văn phòng
- Thống kê
- Công nghệ thiết kế
- Y học
- Nông nghiệp
- Quản trị kinh doanh
- An ninh, quốc phòng
- Giáo dục
- Công nghệ in
- Nghệ thuật, giải trí, v.v....

3. Máy tính điện tử và lịch sử phát triển

Do nhu cầu cần tăng độ chính xác và giảm thời gian tính toán, con người đã quan tâm chế tạo các công cụ tính toán từ xưa: bàn tính tay của người Trung quốc, máy cộng cơ học của nhà toán học Pháp Blaise Pascal (1623 - 1662), máy tính cơ học có thể cộng trừ nhân chia của nhà toán học Đức Gottfried Wilhelm von Leibniz (1646 - 1716), máy sai phân để tính các đa thức toán học, máy phân giải điều khiển bằng phiếu đục lỗ của Charles Babbage (1792 - 1871) ...

Tuy nhiên, máy tính điện tử thực sự hình thành bắt đầu vào thập niên 1950 và đến nay đã trải qua 5 thế hệ được phân loại theo sự tiến bộ về công nghệ điện tử và vi điện tử cũng như các cải tiến về nguyên lý, tính năng và loại hình của nó.

* *Thế hệ 1 (1950 - 1958)*: máy tính sử dụng các bóng đèn điện tử chân không, mạch riêng rẽ, vào số liệu bằng phiếu đục lỗ, điều khiển bằng tay. Máy có kích thước rất lớn, tiêu thụ năng lượng nhiều, tốc độ tính chậm khoảng 300 - 3.000 phép tính mỗi giây. Loại máy tính điển hình thế hệ 1 như EDVAC (Mỹ) hay BESM (Liên xô cũ), ...

* *Thế hệ 2 (1958 - 1964)*: máy tính dùng bộ xử lý bằng đèn bán dẫn, mạch in. Máy đã có chương trình dịch như Cobol, Fortran và hệ điều hành đơn giản. Kích thước máy còn lớn, tốc độ tính khoảng 10.000 đến 100.000 phép/s. Điển hình như loại IBM-1070 (Mỹ) hay MINSK (Liên xô cũ), ...

* *Thế hệ 3 (1965 - 1974)*: máy tính được gắn các bộ xử lý bằng vi mạch điện tử cỡ nhỏ có thể có được tốc độ tính khoảng 100.000 đến 1 triệu phép/s. Máy đã có các hệ điều hành đa chương trình, nhiều người dùng đồng thời hoặc theo kiểu chia thời gian. Kết quả từ máy tính có thể in ra trực tiếp ở máy in. Điển hình như loại IBM 360 (Mỹ) hay EC (Liên Xô cũ), ...

* *Thế hệ 4 (1974 đến nay)*: máy tính bắt đầu có các vi mạch đa xử lý có tốc độ tính hàng chục triệu đến hàng tỷ phép/giây. Giai đoạn này hình thành 2 loại máy tính chính : máy tính cá nhân để bàn (Personal Computer - PC) hoặc xách tay (Laptop hoặc Notebook computer) và các loại máy tính chuyên nghiệp thực hiện đa chương trình, đa vi xử lý ... hình thành các hệ thống mạng máy tính (Computer Networks), và các ứng dụng phong phú đa phương tiện.

* *Thế hệ 5 (1990 - nay)*: bắt đầu có các nghiên cứu tạo ra các máy tính mô phỏng các hoạt động của não bộ và hành vi con người, có trí khôn nhân tạo với khả năng tự suy diễn phát triển các tình huống nhận được và những hệ quản lý kiến thức cơ sở để giải quyết các bài toán đa dạng.

=====