



GT.0000026879

YÊN NGỌC CƯỜNG (Chủ biên)  
ThS. VŨ CHÍ QUANG

# Giáo trình **TIN HỌC** **CƠ SỞ**



GUYỄN  
LIÊU



NHÀ XUẤT BẢN THÔNG TIN VÀ TRUYỀN THÔNG

TS. NGUYỄN NGỌC CƯƠNG (Chủ biên)

ThS. VŨ CHÍ QUANG

Giáo trình  
**TIN HỌC**  
**CƠ SỞ**

TÁI BẢN LẦN 5

NHÀ XUẤT BẢN THÔNG TIN VÀ TRUYỀN THÔNG

Hà Nội, 2015

**Mã số: GD 06 HT 15**

## LỜI NÓI ĐẦU

Tin học ngày càng được ứng dụng rộng rãi và hiệu quả trong mọi lĩnh vực của Đời sống và Xã hội. Để cung cấp cho bạn đọc những kiến thức cơ bản về tin học, NXB Thông tin và Truyền thông tái bản lần thứ 5 cuốn “*Giáo trình Tin học cơ sở*” do TS. Nguyễn Ngọc Cương (Trường khoa Toán tin - Học viện An ninh Nhân dân) làm chủ biên. Cuốn sách đã được xuất bản năm 2010 và tái bản lần 1 năm 2011, tái bản lần 2 năm 2012, tái bản lần 3 năm 2013, tái bản lần 4 năm 2014.

*Tin học cơ sở* là môn học bắt buộc trong chương trình đào tạo đại học của Học viện An ninh Nhân dân. Mục tiêu của môn học là trang bị cho học viên những kiến thức cơ bản về Tin học, cấu trúc máy tính, hệ điều hành, mạng máy tính và Internet, trong đó chú trọng các kỹ năng về tin học văn phòng, hệ quản trị cơ sở dữ liệu.

Do đặc thù của Học viện là đào tạo các sỹ quan an ninh nên ngoài các kiến thức tin học chung như các trường đại học khác, chương trình môn học còn phải đáp ứng các kiến thức tin học ứng dụng trong công an như: chứng cứ điện tử và phương pháp phục hồi an toàn dữ liệu, kỹ thuật sử dụng Internet an toàn.

*Nội dung của cuốn sách gồm 4 phần chính:*

*Phần 1.* Thông tin và xử lý thông tin: trình bày các khái niệm cơ sở của tin học như dữ liệu, thông tin, hệ đếm, phần cứng và phần mềm máy tính.

*Phần 2.* Hệ điều hành Windows: giới thiệu cách sử dụng Microsoft Windows.

*Phần 3.* Tin học văn phòng: giới thiệu tương đối toàn diện về hệ soạn thảo Microsoft Word, bảng tính dữ liệu Microsoft Excel, hệ quản trị cơ sở dữ liệu Microsoft Access.

*Phần 4.* Phục hồi dữ liệu điện tử: trình bày khái niệm chứng cứ điện tử, qui trình thu thập, phục hồi dữ liệu điện tử và phương pháp phân tích. Trong phần này các tác giả cũng giới thiệu một số công cụ phục hồi dữ liệu điện tử trên đĩa từ máy tính, điện thoại di động và thiết bị có bộ nhớ kỹ thuật số.

Hy vọng cuốn sách sẽ thực sự hữu ích đối với các Học viên Học viện An ninh Nhân dân, Học viện Cảnh sát, còn là tài liệu tham khảo cần thiết cho các cán bộ giảng dạy, sinh viên các trường đại học cũng như các kỹ sư, kỹ thuật viên chuyên ngành Công nghệ Thông tin.

Nhà xuất bản xin trân trọng giới thiệu cùng bạn đọc và rất mong nhận được ý kiến đóng góp của quý vị. Mọi ý kiến đóng góp xin gửi về *Nhà xuất bản Thông tin và Truyền thông* - Số 9, ngõ 90, phố Ngụy Như Kon Tum, quận Thanh Xuân, Hà Nội. Điện thoại: (04) 35772143; Fax: (04) 35579858; e-mail: [nxb.tttt@mic.gov.vn](mailto:nxb.tttt@mic.gov.vn) hoặc gửi trực tiếp cho tác giả theo địa chỉ: [cuongnnhvan@yahoo.com](mailto:cuongnnhvan@yahoo.com).

*Trân trọng cảm ơn./.*

**NHÀ XUẤT BẢN  
THÔNG TIN VÀ TRUYỀN THÔNG**

# THÔNG TIN VÀ XỬ LÝ THÔNG TIN

---

---

## 1.1. THÔNG TIN VÀ XỬ LÝ THÔNG TIN

### 1.1.1. Khái niệm về thông tin (Informations)

Đối với chúng ta, thông tin là một khái niệm rất quen thuộc và được sử dụng hàng ngày. Thông tin giúp con người hiểu tốt hơn về đối tượng mà mình quan tâm, nhằm đưa ra các giải pháp giải quyết vấn đề một cách tối ưu trong những điều kiện có thể. Trong tin học, thông tin được định nghĩa đơn giản như sau: *Thông tin là tất cả những gì đem lại hiểu biết, nhận thức cho con người về đối tượng mình quan tâm.*

Bản thân thông tin không phải là một thực thể vật chất, nghĩa là không mang năng lượng nội tại, nhưng khi được tham gia vào các quá trình hoạt động của con người, nó lại thể hiện khả năng vật chất của mình, tức là góp phần làm gia tăng năng lượng, mà với sự hỗ trợ của công nghệ thông tin và kỹ thuật truyền thông hiện đại, độ gia tăng đó ngày càng trở nên đáng kể, thậm chí còn vượt xa dự kiến của con người.

Thông tin mang lại cho chúng ta sự hiểu biết, làm tăng kiến thức (do vật mang tin đem lại) về một sự vật, một hiện tượng hay một quá trình nào đó. Vật mang tin có thể có nhiều dạng thức khác nhau song lượng kiến thức mà nó hàm chứa lại là không đổi. Giá trị của lượng thông tin nhận được còn phụ thuộc trực tiếp vào đối tượng nhận tin. Cụ thể là trước cùng một tin, không phải đối tượng nhận tin nào cũng thu được một lượng thông tin như nhau; kết quả còn phụ thuộc vào nhiều yếu tố liên quan đến tin đó.

Thông tin luôn tồn tại khách quan. Thông tin có thể tạo ra, phát sinh, truyền đi, lưu trữ, chọn lọc. Thông tin cũng có thể bị méo mó, sai lệch do nhiễu hay do người xuyên tạc.

Thông tin không đồng nhất với tín hiệu. Cùng một thông tin có thể được thể hiện bằng các tín hiệu khác nhau. Ví dụ: Tiếng đại bác có thể là tín hiệu chiến tranh, nhưng cũng có thể là tín hiệu chào mừng một nguyên thủ quốc gia.

### 1.1.2. Dữ liệu (data)

Dữ liệu sau khi tập hợp lại và xử lý sẽ cho ta thông tin. Nói cách khác dữ liệu là nguồn gốc, là vật mang thông tin, là vật liệu sản xuất ra tin. Thông tin chứa đựng ý nghĩa còn dữ liệu là các dữ kiện không có cấu trúc và không có ý nghĩa rõ ràng nếu không được tổ chức và xử lý. Cùng một thông tin có thể được biểu diễn bằng những dữ liệu khác nhau. Dữ liệu có nguồn gốc tự nhiên không có tính quy ước còn dữ liệu do con người tự đặt ra để mã hóa thông tin thì phải có quy ước về cách biểu diễn thông tin.

*Ví dụ:* cùng biểu diễn một đơn vị nhưng trong hệ thập phân số một được biểu diễn bởi 1, còn trong hệ đếm La-mã thì lại dùng ký hiệu I.

Mỗi dữ liệu có thể được thể hiện bằng những tín hiệu vật lý khác nhau. Cũng là gậy đầu, đối với nhiều nước trên thế giới là tín hiệu thể hiện sự đồng tình, nhưng ngược lại đối với người Hy Lạp gậy đầu để biểu lộ sự bất đồng.

Mỗi tín hiệu có thể dùng để thể hiện các thông tin khác nhau. Nâng ly rượu, chúc mừng ngày vui của lứa đôi nhưng cũng có thể để giã từ, gợi nhớ mối tình một thuở.

Dữ liệu có nhiều dạng khác nhau, có thể là:

- Tín hiệu vật lý (physical signal): tín hiệu điện, tín hiệu sóng điện-từ, tín hiệu ánh sáng, tín hiệu âm thanh, nhiệt độ, áp suất...
- Các số liệu (number) là dữ liệu bằng số mà ta đã quen với tên gọi số liệu. Đó là các số liệu trong các biểu bảng tính toán, thống kê, tài chính...

- Các ký hiệu (symbol) như các chữ viết (character) và các ký hiệu khắc trên tre, nứa, đá, trên bia, trên vách núi...
- Văn bản, chữ viết (text, character): sách báo, truyện, thông báo, thông tư, công văn...
- Âm thanh (sound): tiếng nói, âm nhạc, tiếng ồn
- Hình ảnh (image): phim ảnh, tivi, camera, tranh vẽ
- Đồ họa (graphic)
- ....

Với sự bùng nổ của thông tin, núi dữ liệu ngày càng khổng lồ. Người ta ước tính: lượng thông tin trên thế giới cứ sau 20 tháng lại tăng gấp đôi. Yếu tố thành công trong mọi lĩnh vực hoạt động ngày nay là việc biết sử dụng thông tin một cách có hiệu quả. Điều đó có nghĩa là từ các dữ liệu sẵn có, phải tìm ra được những thông tin tiềm ẩn có giá trị mà trước đó chưa được phát hiện, tìm ra những xu hướng phát triển và những yếu tố tác động lên chúng. Thực hiện các công việc đó chính là thực hiện quá trình phát hiện tri thức trong cơ sở dữ liệu (*Knowledge Discovery in Database - KDD*) mà trong đó kỹ thuật cho phép ta lấy được các tri thức chính là kỹ thuật khai phá dữ liệu (*Datamining*).

### 1.1.3. Độ đo thông tin

Thông tin có liên quan chặt chẽ đến khái niệm về độ bất định. Mỗi đối tượng chưa xác định hoàn toàn đều có một độ bất định nào đó. Tính bất định này chưa cho biết một cách đích xác và đầy đủ về đối tượng đó. Độ bất định sẽ giảm đi khi nhận thêm thông tin và có liên quan chặt chẽ đến khái niệm xác suất - độ đo khả năng có thể xảy ra của sự kiện (biến cố). Nếu một biến cố không bao giờ xảy ra, xác suất của nó có giá trị bằng 0. Nếu có một biến cố chắc chắn xảy ra, xác suất của nó bằng 1. Đại lượng xác suất có giá trị đoạn  $[0,1]$ .

Người ta có thể định lượng tin tức bằng cách đo độ bất định của hành vi, trạng thái. Xác suất xuất hiện của tin càng thấp thì lượng tin càng cao. Năm 1948, C. Shannon đưa ra công thức sau để xác định độ bất định của sự kiện (tính lượng tin) gọi là Entropi:



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

Trong đó  $P_i$  là xác suất xuất hiện sự kiện  $i$  của hệ, hệ có  $n$  khả năng khác nhau. Ví dụ: khi gieo xấp ngửa đồng xu, lượng tin xuất hiện mặt ngửa sẽ là:

$$H = -\log_2(1/2) = 1$$

Tin tức chỉ xuất hiện khi tối thiểu có hai trạng thái, hai khả năng. Do vậy trường hợp hai trạng thái (xấp, ngửa của đồng xu) đã cho ta đơn vị đo thông tin  $H = 1$  bit. Lượng thông tin chứa trong một bit là vừa đủ để nhận biết một trong hai trạng thái có xác suất xuất hiện như nhau. Tại mỗi thời điểm trong một bit chỉ lưu giữ được hoặc là chữ số 0 hoặc là chữ số 1. Từ bit là từ viết tắt của Binary Digit (chữ số nhị phân). Trong tin học ta thường dùng một số đơn vị bội của bit như bảng 1.1.

Bảng 1.1

Tên gọi	Viết tắt	Giá trị
Byte	B	8 bit
Kilobyte	KB	1024 bytes = $2^{10}$ B
Megabyte	MB	1024 KB = $2^{10}$ KB
Gigabyte	GB	1024MB = $2^{10}$ MB

#### 1.1.4. Mã hóa thông tin

Trong xử lý thông tin tự động, dạng mã quan trọng được dùng là dạng mã nhị phân. Thông tin được mã hóa trên bảng chữ cái gồm hai ký hiệu là chữ số 0 và chữ số 1. Ví dụ, với bảng chữ mã chuẩn của Mỹ ASCII (*American Standard Code for Information Interchange*), mỗi ký tự (chữ cái, chữ số, dấu, ký hiệu đặc biệt) tương ứng với một mã 8 bit (dãy liên tiếp 8 chữ số 0 và 1).

Toàn bộ bảng mã ASCII có 128 ký hiệu được mã hóa, trong đó có:

- 26 chữ cái La-tinh in thường (a, b, c,..., x, y, z)
- 26 chữ cái La-tinh in hoa (A, B, C,...,X, Y, Z)

- 10 chữ số thập phân
- 10 ký tự toán học thông dụng (+, -, \*, /, =, >, <)
- Các dấu chính tả (?, ..., [, ], {, }, %, #, &, \$, ;, :; ...)
- Một số ký hiệu điều khiển, ví dụ CR dùng để điều khiển máy in chuyển đến cột đầu tiên của dòng in.

Bảng 1.2

Ký tự	Mã ASCII	Ký tự	Mã ASCII
0	0011 0000	A	0100 0001
2	0011 0010	B	0100 0010
3	0011 0011	Y	0101 1001
4	0011 0100	Z	0101 1010
5	0011 0101	...	...
6	0011 0110	a	0110 0001
7	0011 0111	b	0110 0010

Bảng mã ASCII mở rộng có 256 ký tự được mã hóa. Mỗi mã của ký tự là một từ 8 bit. Mỗi mã ký tự được lưu trữ trong một byte. Mỗi nước có một bảng mã riêng. Thường các bảng mã đó bao gồm ASCII và phần mở rộng dùng để cài đặt các ký tự riêng của mỗi nước (bảng 1.2).

*Nhận xét:*

- Chữ cái hoa và chữ thường chênh nhau một khoảng là 32. Mã chữ cái 'A' là 65 thì mã chữ cái 'a' là  $97 = 65 + 32$ .
- Cụ thể là mã nhị phân cho chữ cái:
- 'A' sẽ là 40H = 0100 0000
- 'a' sẽ là 61H = 0110 0001
- Còn chữ số '0' sẽ có mã là 30H = 0011 0000
- Các mã từ 0 đến 31 được dùng để mã hóa các ký tự điều khiển. Mã số 127 cũng là mã điều khiển, không biểu diễn ký tự được.
- Ví dụ: Mã 7, BELL: tạo ra tiếng chuông