HỌC VIỆN NÔNG NGHIỆP VIỆT NAM
KHOA CÔNG NGHỆ THÔNG TIN

CHỦ BIÊN: PHẠM QUANG DŨNG

GIÁO TRÌNH
TIN HỌC ĐẠI CƯƠNG

2015

LỜI NÓI ĐẦU

Cuốn giáo trình Tin học đại cương này được viết bởi các giảng viên Khoa Công nghệ thông tin, Học viện Nông nghiệp Việt Nam. Giáo trình bao gồm những kiến thức cơ bản và những kiến thức cập nhật của một số khía cạnh chủ yếu trong lĩnh vực Khoa học máy tính và Công nghệ thông tin như phần cứng, phần mềm, hệ điều hành, mạng máy tính và Internet, cơ sở dữ liệu, thuật toán, ngôn ngữ lập trình, an toàn thông tin. Mỗi chương được viết bởi giảng viên có kinh nghiệm giảng dạy nhiều năm các học phần chuyên về nội dung tương ứng.

Cuốn giáo trình này được dùng để làm tài liệu giảng dạy và học tập chính cho học phần Tin học đại cương gồm 2 tín chỉ trong chương trình đào tạo của các ngành không chuyên Tin học thuộc Học viện Nông nghiệp Việt Nam. Tất nhiên giáo trình cũng có thể được dùng làm tài liệu tham khảo cho nhiều đối tượng bạn đọc khác muốn mở mang kiến thức.

Giáo trình gồm 7 chương, mỗi chương là một mảng kiến thức tương đối độc lập với các chương khác. Bạn đọc có thể lựa chọn đọc những phần phù hợp với nhu cầu mà hầu như không gặp trở ngại về sự đòi hỏi kiến thức từ các phần trước.

Chương 1: Giới thiệu chung, do TS. Phạm Quang Dũng và ThS. Trần Thị Thu Huyền viết. Chương này giới thiệu một số kiến thức nền tảng chung gồm: (1) Các khái niệm về dữ liệu, thông tin, tin học, công nghệ thông tin; (2) Các hệ thống số dùng trong máy tính; (3) Mã hóa và biểu diễn thông tin trong máy tính; và (4) Các ứng dụng của công nghệ thông tin.

Chương 2: Cấu trúc máy tính, do TS. Phạm Quang Dũng biên soạn. Chương 2 đề cập đến: (1) Chức năng, sơ đồ tổ chức, nguyên lý hoạt động của máy tính; (2) Cấu trúc và chức năng của các thành phần cơ bản của máy tính, thuộc 4 khối chức năng gồm: bộ xử lý trung tâm, bộ nhớ, thiết bị ngoại vi và liên kết hệ thống.

Chương 3: Phần mềm máy tính và hệ điều hành, được viết bởi ThS. Phạm Thủy Vân. Qua chương này bạn đọc có thể hiểu được các vấn đề về: (1) Khái niệm và phân loại phần mềm máy tính, quy trình phát triển phần mềm; (2) Khái niệm, lịch sử phát triển hệ điều hành, một số hệ điều hành thông dụng cho máy tính và các thiết bị di động.

Chương 4: Mạng máy tính và Internet, do ThS. Phan Thị Thu Hồng đảm nhiệm. Chương này giới thiệu đến bạn đọc các kiến thức cơ bản về mạng máy tính và Internet bao gồm: (1) Khái niệm, các thành phần cơ bản, các mô hình kết nối và giao thức mạng, phân loại mạng máy tính; (2) Các khái niệm, các dịch vụ phổ biến như world wide web, tìm kiếm, thư điện tử, lưu trữ đám mây.

Chương 5: Cơ sở dữ liệu, do ThS. Hoàng Thị Hà biên soạn. Trong chương này bạn đọc có thể nắm được những kiến thức về: (1) Cơ sở dữ liệu: khái niệm, lợi ích các mức thể hiện của cơ sở dữ liệu, hệ cơ sở dữ liệu; (2) Khái niệm và chức năng của hệ quản trị cơ sở dữ liệu; và (3) Các câu lệnh của ngôn ngữ truy vấn có cấu trúc (SQL).

Chương 6: Thuật toán và ngôn ngữ lập trình, được viết bởi ThS. Lê Thị Nhung. Chương này giới thiệu với bạn đọc: (1) Khái niệm, các tính chất, các cách diễn đạt thuật toán, thiết kế thuật toán và đánh giá độ phức tạp; (2) Khái niệm về ngôn ngữ lập trình, trình biên dịch và trình thông dịch.

Chương 7: Các vấn đề xã hội của công nghệ thông tin; do ThS. Nguyễn Văn Hoàng đảm nhiệm. Qua chương này bạn đọc có thể thu nhận được các kiến thức về: (1) An toàn thông tin như các tài nguyên có thể bị xâm phạm, các hình thức tấn công để lấy cắp hay phá hoại thông tin; (2) Một số điều trong Bộ luật hình sự về tội phạm trong lĩnh vực tin học; và (3) Vấn đề sở hữu trí tuệ nói chung và sở hữu trí tuệ trong công nghệ thông tin nói riêng.

Do đây là lần đầu tiên viết giáo trình Tin học đại cương theo hướng cải tiến nên chắc chắn chúng tôi không thể tránh khỏi những thiếu sót. Nhóm tác giả mong nhận được những ý kiến góp ý từ các bạn đọc để lần tái bản sau giáo trình sẽ tốt hơn. Mọi ý kiến đóng góp xin gửi về địa chỉ pqdung@vnua.edu.vn.

Chúng tôi xin chân thành cảm ơn!

Hà Nội, tháng 1 năm 2015

Nhóm tác giả

MỤC LỤC

DANH MỤC CHỮ VIẾT TẮT

Chữ viết tắt	Từ tiếng Anh	Nghĩa tiếng Việt
AAC	Advanced Audio Coding	Mã hóa âm thanh tiên tiến
ADSL	Asymmetric Digital Subscriber Line	Đường dây thuê bao số bất đối xứng
ALU	Arithmetic and Logic Unit	Đơn vị toán học và logic, nằm trong CPU
AMD	Advanced Micro Devices	Các vi thiết bị tiên tiến, cũng là tên của một tập đoàn phát triển các thiết bị loại này như bộ vi xử lý máy tính và các công nghệ liên quan
ARM	Advanced RISC Machine	Một loại cấu trúc vi xử lý 32-bit kiểu RISC
ASCII	American Standard Code for Information Interchange	Bảng mã chuẩn của Mỹ dùng để trao đổi thông tin
ATA	Advanced Technology Attachment	Cách gọi ngắn gọn của Paralell ATA, là một chuẩn giao tiếp kết nối giữa máy tính và các ổ đĩa cứng, ổ đĩa quang trong máy tính
ATI	Array Technology Inc.	Tên tập đoàn được thành lập năm 1985, chuyên sản xuất cạc đồ họa cho máy tính cá nhân. Năm 2006, ATI được tập đoàn AMD mua lại.
BD	Bluray Disk	Một dạng đĩa quang cho phép ghi/phát lại hình ảnh/âm thanh với chất lượng cao
BIOS	Basic Input/Output System	Hệ thống vào/ra cơ bản
CCFL	Cold-Cathode Fluorescent Lamp	Đèn huỳnh quang catốt lạnh
CD	Compact Disc	Đĩa CD (gọn nhẹ), một dạng đĩa quang
CIDR	Classless Inter-Domain Routing	Lược đồ địa chỉ mới của Internet
CISC	Complex Instruction Set Computer [Architecture]	Kiến trúc tập lệnh phức tạp
CMOS	Complementary Metal–Oxide–Semiconductor,	Công nghệ bán dẫn kim loại bù, dùng để chế tạo các vi mạch tích hợp
CMYK		Hệ màu gồm 4 màu cơ sở: Cyan, Magenta, Yellow, Black
CNTT		Công nghệ thông tin
CPU	Central Processing Unit	Bộ xử lý trung tâm
CSDL		Cơ sở dữ liệu
CU	Control Unit	Đơn vị điều khiển, nằm trong CPU
DBMS	Database Management System	Hệ quản trị cơ sở dữ liệu

Chữ viết tắt	Từ tiếng Anh	Nghĩa tiếng Việt
DDR-SDRAM	Double Data Rate SDRAM,	SDRAM có tốc độ truyền dữ liệu gấp đôi
DIB	Device-Independent Bitmap	Dạng biểu diễn ảnh Bitmap mà không phụ thuộc vào thiết bị lưu trữ
DNS	Domain Name System,	Hệ thống tên miền Internet, cho phép thiết lập tương ứng giữa địa chỉ IP và tên miền
DRAM	Dynamic RAM	RAM động, được dùng để chế tạo bộ nhớ chính
DVD	Digital Video Disc, hoặc Digital Versatile Disc	Đĩa video số, hoặc đĩa đa năng số, một dạng đĩa quang
EEPROM	Electrically Erasable Programmable ROM	Bộ nhớ chỉ đọc có thể lập trình và xóa bằng điện
ENIAC	Electronic Numerical Intergator and Computer	Máy tính và bộ tích hợp số điện tử, tên của chiếc máy tính điện tử đầu tiên trên thế giới (năm 1946)
EOF	End Of File	Tên của ký tự đặc biệt đánh dấu kết thúc tệp tin
FSB	Front Side Bus	Tên bus bên trong CPU
GPL	General Public License	Một điều kiện áp dụng cho việc sử dụng phần mềm nguồn mở
HD (1)	High Definition	Độ rõ nét cao
HD (2)	Hard Disk	Đĩa cứng, ngầm hiểu là dùng công nghệ từ tính. Một dạng ổ cứng khác dùng công nghệ flash là SSD – ổ cứng thể rắn.
HQTCSDL		Hệ quản trị cơ sở dữ liệu
HTML	Hypertext Markup Language	Ngôn ngữ đánh dấu siêu văn bản
IDE (1)	Integrated Drive Electronics	Một chuẩn kết nối khác giữa máy tính với các loại ổ đĩa nêu trên
IDE (2)	Integrated Development Environment	Môi trường phát triển tích hợp
IP	Internet Protocol	Giao thức mạng Internet
ISP	Internet Service Provider	Nhà cung cấp dịch vụ Internet
KHKT		Khoa học kỹ thuật
LAN/ MAN/ WAN/ VAN	Local/ Metropolitan/ Wide/ Vast Area Network	Mạng cục bộ / đô thị / diện rộng / toàn cầu
LCD	Liquid Crystal Display	Màn hình tinh thể lỏng
LED	Light-Emitting Diode	Điốt phát sáng
LPT	Line Printer Terminal	Loại cổng song song để kết nối với máy in theo dòng

Chữ viết tắt	Từ tiếng Anh	Nghĩa tiếng Việt
MIPS	Million Instructions per Second	Số triệu lệnh trên một giây
MPEG	Moving Picture Experts Group	Một định dạng tệp video
MS-DOS	Microsoft Disk Operating System	Hệ điều hành hướng đĩa của hãng Microsoft
NAT	Network Address Translation	Biên dịch địa chỉ mạng
NCP	Network Control Protocol	Giao thức điều khiển mạng
NIC	Network Interface Card	Cạc giao diện mạng
ODBMS	Object BDMS	Hệ quản trị cơ sở dữ liệu hướng đối tượng
OS	Operating System	Hệ điều hành
PC	Personal Computer	Máy tính cá nhân
RAM	Random Access Memory	Bộ nhớ truy nhập ngẫu nhiên
RDBMS	Relational BDMS	Hệ quản trị cơ sở dữ liệu quan hệ
RGB		Hệ màu gồm 3 màu cơ sở: Red, Green, Blue
RISC	Reduced Instruction Set Computer [Architecture],	Kiến trúc tập lệnh tập lệnh rút gọn
ROM	Read Only Memory	Bộ nhớ chỉ cho phép đọc
SATA	Serial ATA	Một chuẩn gắn kết nối tiếp, để nối ổ đĩa cứng hoặc ổ đĩa quang với bo mạch chủ
SDRAM	Synchronous DRAM	RAM động làm việc được đồng bộ bởi xung đồng hồ
SQL	Structured Query Language	Ngôn ngữ truy vấn có cấu trúc
SRAM	Static RAM	RAM tĩnh, được dùng để chế tạo bộ nhớ cache
SSD	Solid State Drive	Ổ cứng thể rắn, dùng công nghệ flash
TCP/IP	Transmission Control Protocol/ Internet Protocol	Giao thức điều khiển truyền dữ liệu/giao thức Internet
USB	Universal Serial Bus	Bus nối tiếp đa năng, là một chuẩn kết nối các thiết bị ngoại vi với máy tính
VGA	Video Graphics Array	Tên cạc đồ họa video hoặc cổng kết nối máy tính với màn hình hoặc máy chiếu (projector)
VXL		Vi xử lý
XML	eXtensible Markup Language,	Ngôn ngữ đánh dấu có thể mở rộng, có khả năng mô tả nhiều loại dữ liệu khác nhau, với mục đích chính là đơn giản hóa việc chia sẻ dữ liệu giữa các hệ thống khác nhau, đặc biệt là các hệ thống được kết nối với Internet

Chương 1

GIỚI THIỆU CHUNG

Chương 1 giới thiệu những kiến thức cơ bản và nền tảng nhất của Tin học. Mục 1.1 nêu những khái niệm về dữ liệu, thông tin, tin học và công nghệ thông tin. Mục 1.2 trình bày về biểu diễn dữ liệu trong máy tính, các hệ thống số và chuyển đổi giữa các hệ cơ số. Các mã hóa một số dạng dữ liệu thông dụng sẽ được trình bày trong mục 1.3. Cuối cùng, mục 1.4 sẽ giới thiệu những ứng dụng của công nghệ thông tin trong các lĩnh vực đời sống.

1.1. MỘT SỐ KHÁI NIỆM CƠ BẢN

Đối với chúng ta, quá trình hình thành trí tuệ bắt đầu từ việc thu nhận và xử lý dữ liệu rời rạc để có thông tin, rồi kiểm nghiệm thông tin để có thể vận dụng vào mục đích cụ thể nào đó ta gọi là tri thức. Trí tuệ là khả năng sử dụng tri thức một cách khôn ngoan nhằm đạt được mục đích. Một ví dụ cụ thể về quá trình trên như sau:

- Dữ liệu: mưa, nắng, râm, cao, thấp, vừa, bay, chuồn chuồn, trời.
- Thông tin: khi thấy chuồn chuồn bay thấp thường thấy một lúc sau trời mưa, bay cao vừa phải thì trời râm mát, còn khi bay cao thì trời nắng.
- Tri thức: chuồn chuồn bay thấp trời mưa, bay cao trời nắng, bay vừa trời râm.
- Trí tuệ: khi thấy chuồn chuồn bay thấp thì ta cất quần áo đang phơi. Ví dụ đỉnh cao lợi dụng thời tiết thời Tam Quốc như Chu Du lợi dụng gió đánh hỏa công trận Xích Bích, Gia Cát Lượng lợi dụng mưa tuyết phá trận xe thiết xa của rợ Khương.

Trong lĩnh vực công nghệ thông tin, máy tính trợ giúp con người chủ yếu ở khâu đầu tiên, từ dữ liệu đến thông tin. Phần này sẽ giới thiệu tới bạn đọc các khái niệm về dữ liệu, thông tin, tin học và công nghệ thông tin.

1.1.1. Dữ liệu

Dữ liệu (Data) là những con số hoặc dữ kiện thuần túy, rời rạc do quan sát hoặc đo đếm được, không có ngữ cảnh hay diễn giải. Dữ liệu sau khi được tổ chức lại và xử lý sẽ cho ra thông tin.

Ví dụ: Với một quyển sách thì chữ, hình ảnh là dữ liệu còn nội dung của quyển sách là thông tin. Để biết được nội dung thì phải đọc sách. Việc đọc sách chính là xử lý dữ liệu.

Trong thực tế dữ liệu có thể là:

- Văn bản: Sách, báo, truyện, công văn...

- Các loại số liệu: Số liệu thống kê về nhân sự, thời tiết, kho tàng...

- Âm thanh, hình ảnh: Tiếng nói, âm nhạc, phim ảnh, tranh vẽ...

1.1.2. Thông tin

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.

Tương tự như dữ liệu, thông tin có thể tồn tại dưới nhiều hình thức khác nhau như âm thanh, hình ảnh, ký tự..., có thể được nén, giải nén, mã hóa, giải mã và được truyền tải qua các môi trường vật lý khác nhau như ánh sáng, sóng âm, sóng điện từ.

Ví dụ: Khi chúng ta nói chuyện trực tiếp với nhau, thông tin được thể hiện dưới dạng âm thanh và được truyền tải qua môi trường sóng âm. Còn khi chúng ta trao đổi với nhau qua điện thoại, thông tin được biểu diễn dưới dạng âm thanh nhưng được truyền tải qua môi trường sóng điện từ.

Tuy nhiên, giữa dữ liệu và thông tin không phải lúc nào cũng đồng nhất với nhau.

Ví dụ: Một số kí hiệu trong hệ đếm La Mã mang ý nghĩa thông tin là số nhưng trong hệ thống chữ La-tinh lại mang ý nghĩa là chữ cái.

Cùng một dữ liệu nhưng tùy thuộc vào đối tượng tiếp nhận dữ liệu khác nhau lại có thể cho ra khối lượng và chất lượng thông tin khác nhau.

Ví dụ: Cùng một tài liệu có sinh viên hiểu được 100% nhưng cũng có sinh viên chỉ hiểu được 50%, có sinh viên có thể phát triển liên hệ được với các vấn đề khác nhưng có sinh viên lại không...

Hình thức vật lý của thông tin được gọi là tín hiệu. Giữa thông tin và tín hiệu không phải lúc nào cũng đồng nhất với nhau. Cùng một thông tin có thể được biểu diễn bởi nhiều tín hiệu khác nhau và ngược lại cùng một tín hiệu có thể biểu diễn nhiều dạng thông tin khác nhau.

Ví dụ: Cùng một thông tin về chỉ dẫn giao thông nhưng có thể được biểu diễn dưới các tín hiệu khác nhau như là đèn tín hiệu giao thông hay là chỉ dẫn của cảnh sát giao thông... Hay cùng tín hiệu là gật đầu trong từng trường hợp khác nhau lại biểu diễn thông tin khác nhau tùy thuộc vào câu hỏi...

1.1.3. Tin học

Thông tin nằm trong dữ liệu, xử lý thông tin bao gồm nhiều quá trình xử lý dữ liệu để rút ra thông tin hữu ích phục vụ con người. Khi xã hội càng phát triển thì khối lượng thông tin, dữ liệu ngày càng nhiều và con người không thể xử lý thông tin một cách thủ công được mà cần tới sự hỗ trợ của máy móc để xử lý thông tin một cách tự động. Trước yêu cầu đó của con người, một ngành khoa học mới đã ra đời, đó là Tin học.

Tin học (Informatics) là một ngành khoa học chuyên nghiên cứu các phương pháp, công nghệ và các kỹ thuật xử lý thông tin một cách tự động.

Hay nói một cách khác: Tin học là một ngành khoa học chuyên nghiên cứu về khả năng lưu trữ, truyền tải và xử lý thông tin. Điều này đã được thể hiện rõ qua quá trình hình thành và phát triển của ngành tin học. Trước đây, những thiết bị lưu trữ chưa đa dạng có dung lượng nhỏ với dung lượng tính theo Megabyte, Gigabyte được thay thế dần bởi các thiết bị lưu trữ đa dạng, gọn nhẹ hơn nhưng có dung lượng lưu trữ lớn hơn rất nhiều, tính theo Terabyte; Đường truyền thông tin có tốc độ thấp dần được thay thế bởi đường truyền tốc độ cao, không dây; Những bộ vi xử lý có tốc độ thấp dần được thay thế bởi những bộ vi xử lý tốc độ cao, bộ đa xử lý...

Sản phẩm mà tin học phát minh ra để giúp con người xử lý thông tin tự động là máy vi tính hay máy tính (computer).

Từ "tin học" đã được dịch từ từ *informatique* trong tiếng Pháp. Từ *informatics* trong tiếng Anh cũng bắt nguồn từ từ tiếng Pháp này, nhưng theo thời gian *informatics* đã mang nghĩa khác dần với nghĩa ban đầu. Ngày nay, thuật ngữ tiếng Anh tương đương với *informatique* là *computer science*, nghĩa là "khoa học máy tính". Thuật ngữ này được sử dụng rất rộng rãi trên thế giới và đa số các trường đại học ở nước ngoài sử dụng cụm từ *Computer Science* để đặt tên cho khoa chuyên môn.

Khoa học máy tính là ngành nghiên cứu các cơ sở lý thuyết về thông tin và tính toán cùng sự thực hiện và ứng dụng của chúng trong các hệ thống máy tính. Khoa học máy tính gồm nhiều ngành hẹp; một số ngành tập trung vào các ứng dụng thực tiễn cụ thể chẳng hạn như đồ họa máy tính, trong khi một số ngành khác lại tập trung nghiên cứu đến tính chất cơ bản của các bài toán tính toán như lý thuyết độ phức tạp tính toán. Ngoài ra còn có những ngành khác nghiên cứu các vấn đề trong việc thực thi các phương pháp tính toán. Ví dụ, ngành lý thuyết ngôn ngữ lập trình nghiên cứu những phương thức mô tả cách tính toán khác nhau, trong khi ngành lập trình nghiên cứu cách sử dụng các ngôn ngữ lập trình và các hệ thống phức tạp và ngành tương tác người-máy tập trung vào những thách thức trong việc làm cho máy tính và công việc tính toán hữu ích và dễ sử dụng đối với mọi người dùng. (*trích từ wikipedia*)

1.1.4. Công nghệ thông tin

Thuật ngữ *Công nghệ thông tin (Information Technology)* mang ý nghĩa về khía cạnh kỹ thuật, công nghệ hơn là khía cạnh khoa học. Ở Việt Nam, thuật ngữ này được sử dụng rộng rãi, dễ gây nhầm lẫn là "Công nghệ thông tin" mang nghĩa rộng hơn "Khoa học máy tính".

Luật Công nghệ thông tin do Quốc hội nước Cộng hòa XHCN Việt Nam ban hành ngày 29/6/2006 có đưa ra khái niệm: "*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*".

Ở đây, thông tin số là thông tin được tạo lập bằng phương pháp dùng tín hiệu số. Trước đây, khi muốn gửi thư người ta thường dùng phương pháp chuyển thư tay nhưng giờ đây đã được thay thế bằng thư điện tử. Với khối lượng lớn công văn giấy tờ trong các cơ quan, xí nghiệp, trường học... việc lưu trữ số trở nên đơn giản và gọn nhẹ hơn.

1.2. CÁC HỆ THỐNG SỐ VÀ CÁC PHÉP TOÁN DÙNG TRONG MÁY TÍNH

Trong máy tính có sử dụng 3 hệ thống số là hệ cơ số 10 (gọi tắt là *hệ 10*), hệ cơ số 2 và hệ cơ số 16. Trong đó, hệ 2 là hệ cốt lõi được bộ vi xử lý sử dụng để tính toán, xử lý (ta sẽ xét kỹ hơn ở mục 1.3). Trong mục này sẽ giới thiệu các nội dung về 3 hệ thống số và chuyển đổi giữa chúng, các phép toán số học trên hệ 2 và các phép toán logic.

1.2.1. Các hệ thống số

a. Hệ cơ số 10 (Hệ thập phân - Decimal Numeral System)

Hệ 10 là hệ đếm được sử dụng để đếm và tính toán trong đời sống hàng ngày. Hệ 10 sử dụng 10 ký hiệu số 0, 1, 2, 3, 4, 5, 6, 7, 8, 9 để biểu diễn các số. Các chương trình máy tính thường cho phép người dùng nhập vào các số hệ 10, xuất kết quả ở hệ 10, nhưng quá trình tính toán trung gian ở bên trong là các số hệ 2.

Khi làm việc với nhiều hệ thống số khác nhau, để phân biệt một số viết trong hệ cơ số này với một số viết trong hệ cơ số khác người ta thường viết kèm theo chỉ số có giá trị bằng cơ số của hệ đếm.

Ví dụ: 2092_{10}; $789,12_{10}$; $12A_{16}$; 10110_2

Một số hệ 10 có thể biểu diễn ở dạng khai triển theo cơ số 10. Ví dụ:

$$8623,56_{10} = 8\times10^3 + 6\times10^2 + 2\times10^1 + 3\times10^0 + 5\times10^{-1} + 6\times10^{-2}$$

Trong đó:

- 8, 6, 2, 3, 5, 6 là các chữ số thành phần của số 8623,56

- 10 là cơ số của hệ
- Số mũ tương ứng với vị trí của chữ số thành phần: bằng 0 với chữ số phần nguyên nhỏ nhất, tăng dần về phía trái (1, 2, 3...), giảm dần về phía phải (-1, -2...).

Từ hệ cơ số 10 ta tổng quát hóa cho hệ cơ số *a* ($a \geq 2$). Số hệ *a* có các chữ số là b_i, $N_a = b_n b_{n-1} \ldots b_1 b_0, b_{-1} b_{-2} \ldots b_{-m}$ có thể biểu diễn theo cơ số a như sau:

$$N_a = b_n \times a^n + b_{n-1} \times a^{n-1} + \ldots + b_1 \times a^1 + b_0 \times a^0 + b_{-1} \times a^{-1} + b_{-2} \times a^{-2} + \ldots + b_{-m} \times a^{-m} \qquad (1.1)$$

Giá trị của tổng ở vế phải trong công thức (1.1) được gọi là giá trị của số N_a. Công thức (1.1) sẽ được sử dụng để chuyển đổi số hệ a sang hệ 10.

b. Hệ cơ số 2 (Hệ nhị phân - Binary Numeral System)

- Hệ 2 hay hệ nhị phân chỉ sử dụng 2 ký hiệu số là 0 và 1 để biểu diễn các số. Đây là hệ cơ số cơ sở của máy tính. Máy tính chỉ lưu trữ và xử lý các dữ liệu ở dạng số nhị phân.

- Có thể biểu diễn một số trong hệ 2 ra thành tổng các hệ số nhân theo quy tắc nêu ở phần trên:

Ví dụ: $10011_2 = 1 \times 2^4 + 0 \times 2^3 + 0 \times 2^2 + 1 \times 2^1 + 1 \times 2^0 = 19_{10}$

c. Hệ cơ số 16 (Hexadecimal Numeral System)

- Hệ 16 sử dụng 16 ký hiệu để biểu diễn các số: 10 ký hiệu số từ 0, 1..., 9 để biểu diễn các giá trị từ 0 đến 9 và 6 ký hiệu chữ A, B, C, D, E, F để biểu diễn các giá trị từ 10 đến 15.

- Hệ 16 được dùng để đánh địa chỉ các ô nhớ, địa chỉ vật lý của các máy tính trong mạng (địa chỉ MAC), địa chỉ của các cổng vào-ra trong máy tính. Các địa chỉ này hiển thị cho người dùng ở dạng số hệ 16 mà không phải hệ 2 vì lý do làm cho địa chỉ ngắn gọn hơn, dễ nhớ, dễ sử dụng hơn.

- Có thể biểu diễn một số trong hệ 16 ra thành tổng các hệ số nhân với lũy thừa của cơ số.

Ví dụ: $12A_{16} = 1 \times 16^2 + 2 \times 16^1 + A \times 16^0 = 298_{10}$

1.2.2. Chuyển đổi giữa các hệ cơ số

a. Chuyển từ hệ a sang hệ 10

Quy tắc: Muốn chuyển một số hệ *a* (2 hoặc 16) sang hệ 10 ta đem triển khai số trong hệ *a* ra thành tổng các hệ số nhân với lũy thừa của cơ số, khi đó ta sẽ được biểu thức trong hệ 10. Tính giá trị của biểu thức đó ta sẽ được số tương ứng trong hệ 10.

Ví dụ:

$$110101_2 = 1 \times 2^5 + 1 \times 2^4 + 0 \times 2^3 + 1 \times 2^2 + 0 \times 2^1 + 1 \times 2^0 = 32 + 16 + 4 + 1 = 53_{10}$$

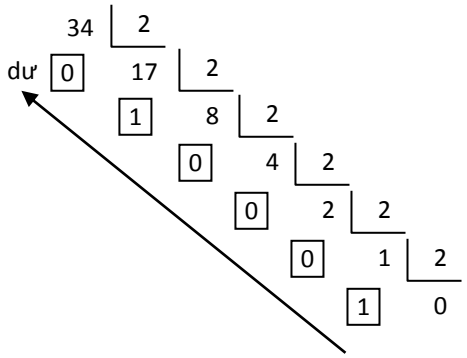
$$10F_{16} = 1 \times 16^2 + 0 \times 16_1 + F \times 16^0 = 256 + 15 \times 16^0 = 256 + 15 = 271_{10}$$

b. Chuyển từ hệ 10 sang hệ a

Ta chỉ xét trường hợp chuyển số nguyên hệ 10 sang hệ a.

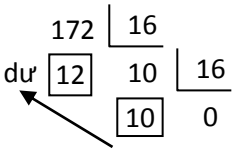
* Quy tắc: Đem số hệ 10 chia nguyên liên tiếp cho cơ số *a* cho tới khi thương bằng 0 thì dừng lại, với mỗi phép chia ta nhận được một số dư. Lấy các số dư của phép chia theo thứ tự ngược lại ta được số trong hệ *a* (số dư của phép chia cuối cùng là chữ số có trọng số lớn nhất, chữ số đầu tiên bên trái).

Ví dụ 1: Đổi số 34_{10} sang hệ 2, ta thực hiện các phép chia như sau:



Kết quả là $34_{10} = 100010_2$

Ví dụ 2: Đổi số 172_{10} sang hệ 16



Kết quả là $172_{10} = AC_{16}$ (10 ứng với A, 12 ứng với C)

c. Chuyển từ hệ 2 sang hệ 16

Bảng 1.1. Mười sáu số đầu tiên trong 3 hệ cơ số

Hệ 10	Hệ 2	Hệ 16
0	0000	0
1	0001	1
2	0010	2
3	0011	3
4	0100	4
5	0101	5
6	0110	6
7	0111	7
8	1000	8
9	1001	9
10	1010	A
11	1011	B
12	1100	C
13	1101	D
14	1110	E
15	1111	F

Nhận xét: Từ bảng trên ta thấy khi hệ 16 dùng đến chữ số lớn nhất thì hệ 2 phải dùng đến 4 chữ số. Vì $16 = 2^4$, mỗi chữ số hệ 16 sẽ tương đương với 4 chữ số hệ 2.

Quy tắc: Để chuyển một số từ hệ 2 sang hệ 16 ta nhóm thành các nhóm 4 chữ số hệ 2 từ phải qua trái, sau đó chuyển từng nhóm 4 chữ số hệ 2 thành các chữ số hệ 16.

Ví dụ: $110\ 1100\ 1011_2 = ?_{16}$

$0110\ \ 1100\ \ 1011 => 6CB_{16}$

d. Chuyển từ hệ 16 sang hệ 2

Quy tắc: chuyển từng chữ số hệ 16 thành 4 chữ số hệ 2.

Ví dụ: $9C0A_{16} = ?_2 = 1001\ 1100\ 0000\ 1010_2$

1.2.3. Các phép toán số học trên hệ 2

a. Phép cộng

Bảng cộng hai bit:

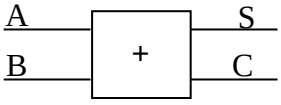
A	B	Tổng (Sum, S)	Số nhớ (Carry, C)
0	0	0	0
0	1	1	0
1	0	1	0
1	1	0	1

Cách thức thực hiện phép cộng hai số hệ 2 thủ công: Thực hiện cộng như trong hệ 10, cộng từng cột bit từ phải qua trái, có nhớ sang cột bit cao hơn.

Ví dụ:

C			1	1	
A	0	1	0	1	1
B	1	0	0	1	1
A+B	1	1	1	1	0

Trong máy tính, phép cộng hai bit được thực hiện bằng mạch cộng như sau:



b. Phép trừ

Trong kỹ thuật máy tính, để tận dụng các mạch cộng đã có sẵn người ta thực hiện phép trừ thông qua phép cộng và phép lấy số đối: cộng số bị trừ với số đối của số trừ.

$$A - B = A + (-B)$$

Vấn đề đặt ra là phải có cách biểu diễn số âm trong hệ 2 trong máy tính để các phép tính toán vẫn cho kết quả đúng. Cách biểu diễn số nguyên và số thực (dương, âm) sẽ được trình bày ở mục 1.3.

c. Phép nhân và phép chia

Trong máy tính, phép nhân và phép chia được thực hiện qua phép cộng, phép trừ và phép dịch bit. Ở mức đại cương, chúng ta không xét sâu hơn hai phép toán này ở đây.

1.2.4. Các phép toán logic

Ngoài việc hỗ trợ các phép toán số học, máy tính cũng phải hỗ trợ các phép toán logic như NOT, AND, OR vì sự cần thiết của chúng. Ta hay phải làm việc với các phép toán logic trong bảng tính Excel hay trong khi lập trình. Ví dụ với ngôn ngữ lập trình Pascal, để kiểm tra một biến *x* có nằm trong khoảng [7,8) hay không, nếu đúng thì viết dòng "Điểm B":

```
If (x>=7) and (x<8) then
    Writeln("Diem B");
```

Giá trị logic biểu diễn một trong hai trạng thái đối lập là đúng/sai, có/không. Trong đại số logic giá trị đúng được gọi là TRUE, giá trị sai được gọi là FALSE. Các phép toán logic tác động trên các giá trị logic TRUE, FALSE gồm có:

a. Phép toán NOT (phủ định hay đảo)

Bảng chân lý:

X	NOT X
FALSE	TRUE
TRUE	FALSE

b. Phép toán AND (và)

Bảng chân lý:

X	Y	X AND Y
FALSE	FALSE	FALSE
FALSE	TRUE	FALSE
TRUE	FALSE	FALSE
TRUE	TRUE	TRUE

Nhận xét: Phép toán AND chỉ cho kết quả "đúng" khi cả hai toán hạng đều "đúng".

c. Toán tử OR (hoặc)

Bảng chân lý:

X	Y	X OR Y
FALSE	FALSE	FALSE
FALSE	TRUE	TRUE
TRUE	FALSE	TRUE
TRUE	TRUE	TRUE

Nhận xét: Phép toán OR chỉ cho kết quả "sai" khi cả hai toán hạng đều "sai".

d. Toán tử XOR (eXclusive OR, hoặc loại trừ)

Bảng chân lý:

X	Y	X XOR Y
FALSE	FALSE	FALSE
FALSE	TRUE	TRUE
TRUE	FALSE	TRUE
TRUE	TRUE	FALSE

Nhận xét: Phép toán XOR cho kết quả "đúng" khi hai toán hạng khác nhau, cho kết quả "sai" khi hai toán hạng giống nhau.

e. Biểu thức logic và thứ tự ưu tiên các phép toán

- Biểu thức logic là sự kết hợp các giá trị logic bằng các phép toán logic để tạo ra một giá trị logic mới. Mỗi biểu thức logic có kết quả là một giá trị hoặc đúng (TRUE) hoặc sai (FALSE).

- Nếu trong biểu thức logic có chứa nhiều phép toán logic thì các phép toán logic được thực hiện theo thứ tự ưu tiên sau: NOT → AND → OR, XOR (OR và XOR cùng mức ưu tiên). Các phép toán cùng mức ưu tiên được thực hiện từ trái qua phải.

Ví dụ:

(Các SV có hộ khẩu Hà Nội) AND NOT (Các SV dân tộc Kinh)

Trong ví dụ này cần ưu tiên thực hiện phép NOT trước rồi mới thực hiện phép AND.

1.3. BIỂU DIỄN VÀ MÃ HÓA THÔNG TIN

1.3.1. Biểu diễn thông tin trong máy tính và các đơn vị thông tin

a. Biểu diễn thông tin trong máy tính

Như đã đề cập ở mục 1.1, dữ liệu thực có thể tồn tại ở nhiều dạng khác nhau: dạng số (số nguyên, số thực) hoặc phi số (chữ viết, âm thanh, hình ảnh). Tuy nhiên, máy tính chỉ xử lý được dữ liệu ở dạng số nhị phân (hệ cơ số 2). Do vậy, dữ liệu thực bất kể ở dạng nào muốn đưa vào máy tính để lưu trữ, xử lý, hay truyền tải cần phải được mã hóa (số hóa thành số nhị phân). Sau khi xử lý, dữ liệu kết quả sẽ được khôi phục lại ở dạng dễ hiểu, dễ cảm nhận đối với con người.

Sở dĩ trong máy tính chỉ dùng được số nhị phân để biểu diễn thông tin là vì các linh kiện và vật liệu điện tử dùng để chế tạo bộ nhớ trong của máy tính (flip-flop, tụ điện) chỉ có hai trạng thái đối lập (ví dụ như *có điện* hoặc *không có điện*), tương ứng được biểu diễn là 1 và 0.

b. Các đơn vị thông tin

Trong khoa học máy tính, mỗi chữ số nhị phân được gọi là một bit (viết tắt của từ tiếng Anh là BInary digiT). Mỗi ô nhớ trong máy tính lưu trữ được một bit. Bản thân mỗi ô nhớ đó còn được gọi là một bit. Các bit được đánh số thứ tự bắt đầu từ 0.

Một nhóm 8 bit bắt đầu từ bit thứ 8i ($i \geq 0$, nguyên) được gọi là một byte. Các byte được đánh địa chỉ bắt đầu từ 0. Byte 0 gồm các bit từ 0 đến 7, byte 1 gồm các bit từ 8 đến 15... Một đơn vị nữa cũng hay được sử dụng là word (từ nhớ). Một word gồm 2, 4, hay 8 byte tùy thuộc vào bộ vi xử lý (CPU) cụ thể có thể xử lý mỗi lần được bao nhiêu byte.

Các đơn vị bội của byte hay được sử dụng và cách quy đổi như sau:

1 Kilobyte (1 KB) = 2^{10} byte = 1024 byte

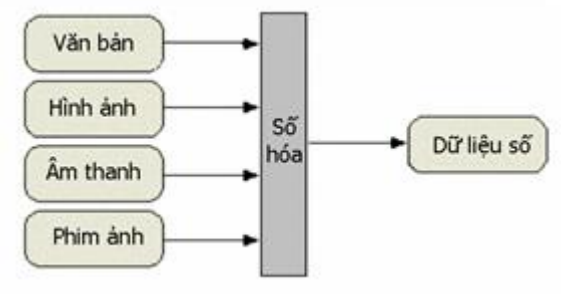
1 Megabyte (1 MB) = 2^{10} KB = 2^{20} byte = 1.048.576 byte

1 Gigabyte (1 GB) = 2^{10} MB = 2^{20} KB = 2^{30} byte

1 Terabyte (1 TB) = 2^{10} GB = 2^{20} MB = 2^{30} KB = 2^{40} byte

1.3.2. Khái niệm về mã hóa

Mã hóa thông tin trong máy tính thực chất là số hóa dữ liệu thành các chuỗi số nhị phân (hình 1.1) theo những quy ước chung để các máy tính có thể lưu trữ, xử lý và trao đổi thông tin với nhau.



Hình 1.1. Sơ đồ số hóa dữ liệu

Trong máy tính người ta dùng các số nhị phân có độ dài (số bit) cố định để biểu diễn thông tin. Các số nhị phân này được gọi là từ mã. Với độ dài từ mã là *n*, ta có thể biểu diễn được 2^n thông tin khác nhau.

Ví dụ:

- Nếu dùng 1 byte (8 bit) để biểu diễn các số nguyên không dấu thì ta có thể biểu diễn được $2^8 = 256$ số có giá trị từ 0 đến 255 như sau:

Từ mã	**Số nguyên**
0000 0000	0
0000 0001	1
0000 0010	2
…	….
1111 1111	255

- Nếu dùng 1 byte để biểu diễn các ký tự (chữ cái, chữ số thập phân, các dấu chấm câu, các ký hiệu phép toán...) thì có thể biểu diễn được $2^8 = 256$ ký tự khác nhau.

Cách mã hóa các loại dữ liệu được tuân theo những chuẩn chung để các máy tính có thể "hiểu" được nhau khi trao đổi, xử lý thông tin.

- Các ký tự: mã hóa theo bảng mã ASCII hoặc Unicode
- Các số nguyên: mã hóa theo một số chuẩn quy ước
- Các số thực: mã hóa theo số dấu phẩy động
- Dữ liệu ảnh, âm thanh, phim: mã hóa rời rạc thành các ma trận số thực biểu diễn cường độ sáng, tần số âm.

Câu hỏi đặt ra là làm sao máy tính phân biệt được chuỗi số nhị phân nào ứng với dữ liệu dạng số, chuỗi nào ứng với dạng ký tự... Ví dụ với một chuỗi 8 bit 0100 0001, khi nào thì máy tính cần hiểu đó là biểu diễn của số 65_{10}, khi nào cần hiểu là biểu diễn của ký tự 'A'? Để thực hiện được điều này, bằng cách nào đó các chương trình máy tính hoặc người sử dụng phải khai

báo kiểu và cấu trúc dữ liệu của các thành phần trong chương trình để hệ điều hành ghi nhớ vào các vùng nhớ thích hợp có địa chỉ và kích thước xác định. Ví dụ, với file ảnh thì các thông tin đó được chương trình tạo ảnh số lưu ở đầu file. Với các ngôn ngữ lập trình, người lập trình sẽ khai báo các hằng, biến qua các câu lệnh. Ta lấy ví dụ trong ngôn ngữ Pascal như sau:

```
Var   ch: char;          {biến ch có kiểu ký tự}
      st: string;        {biến st có kiểu chuỗi ký tự}
      i: byte;           {biến i có kiểu số nguyên không dấu 8 bit}
      j: shortint;       {biến j có kiểu số nguyên có dấu 8 bit}
      k1: word;          {biến k1 có kiểu số nguyên không dấu 16 bit}
      k2: integer;       {biến k2 có kiểu số nguyên có dấu 16 bit}
      m: longint;        {biến m có kiểu số nguyên có dấu 32 bit}
      r: real;           {biến r có kiểu số thực 6 byte}
      t: double;         {biến t có kiểu số thực 8 byte}
```

Các mục con tiếp theo sẽ trình bày việc mã hóa các loại dữ liệu thông dụng gồm: ký tự, số nguyên, số thực, ảnh tĩnh, âm thanh và phim ảnh.

1.3.3. Mã hóa tập ký tự

Về nguyên tắc, mỗi quốc gia đều có thể tự thiết kế một bảng mã riêng để biểu diễn các ký tự của nước mình. Nếu làm như vậy thì các máy tính và thậm chí các bộ phận của cùng một máy tính sẽ không hiểu nhau khi kết nối với nhau. Bởi vậy, các nước phải quy định dùng chung một bảng mã để biểu diễn ký tự, bảng mã này được gọi là bảng mã chuẩn. Trong thực tế có nhiều bảng mã chuẩn nhưng được sử dụng phổ biến nhất trên máy tính hiện nay là bảng mã ASCII (America Standard Code for Information Interchange) và bảng mã Unicode.

Bảng mã ASCII có 256 từ mã 8 bit, biểu diễn 256 ký tự khác nhau. Bảng mã ASCII được chia thành phần tiêu chuẩn (gồm các từ mã có giá trị trong hệ 10 từ 0 đến 127) và phần mở rộng (có mã từ 128 đến 255). Các máy tính trên thế giới có bảng mã ASCII tiêu chuẩn giống nhau, phần mở rộng có thể khác nhau vì được dùng để biểu diễn các ký tự của riêng từng nước.

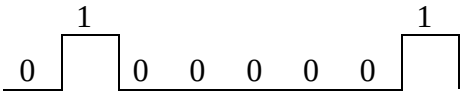
Bảng 1.2. Bảng mã ASCII tiêu chuẩn

Các bit cao

Các bit thấp (poids faibles)	000	001	010	011	100	101	110	111
0000	NUL	DLE	SP	0	@	P	\	p
0001	SOH	DC1	!	1	A	Q	a	q
0010	STX	DC2	"	2	B	R	b	r
0011	ETX	DC3	#	3	C	S	c	s
0100	EOT	DC4	$	4	D	T	d	t
0101	ENQ	NAK	%	5	E	U	e	u
0110	ACK	SYN	&	6	F	V	f	v
0111	BEL	ETB	'	7	G	W	g	w
1000	BS	CAN	(	8	H	X	h	x
1001	HT	EM	)	9	I	Y	i	y
1010	LF	SUB	*	:	J	Z	j	z
1011	VT	ESC	+	;	K	[	k	{
1100	FF	FS	'	<	L	Ç	l	ù
1101	CR	GS	-	=	M	]	m	}
1110	SO	RS	.	>	N	↑	n	≈
1111	SI	US	/	?	O	<--	o	DEL

code ASCII

Ví dụ: ký tự ‘A’ được mã hóa thành 0100 0001 (= 65_{10}). Khi ta ấn Shift+A trên bàn phím, một xung điện truyền đến bộ xử lý máy tính có dạng tương ứng là:



Tại các chu kỳ có mức điện thế thấp, ô nhớ tương ứng sẽ được ghi bit 0; các ô nhớ tương ứng với các chu kỳ có mức điện thế cao sẽ được ghi bit 1. Máy tính xử lý chuỗi nhị phân đó rồi "vẽ" lên màn hình ký tự 'A'.

Do bảng mã ASCII mở rộng của các nước trên thế giới khác nhau nên khi gửi một văn bản từ nước này sang nước khác thì văn bản không hiển thị đúng. Bởi vậy, cả thế giới lại thống nhất dùng chung một bảng mã trong đó biểu diễn được tất cả ký tự của các nước, bảng mã này được gọi là bảng mã Unicode. Bảng mã Unicode có 65536 (= 2^{16}) từ mã 16 bit. 128 từ mã đầu tiên của Unicode mã hóa giống với ASCII.

1.3.4. Mã hóa số nguyên và số thực

Khi ta nhập vào các số hệ 10, máy tính cũng phải tính toán ở hệ 2. Số nguyên và số thực được biểu diễn trong máy tính theo các chuẩn khác nhau.

a. Số nguyên

Máy tính có thể dùng 8 bit, 16 bit hoặc 32 bit để biểu diễn một số nguyên, càng dùng nhiều bit thì biểu diễn được số nguyên càng lớn. Với 32 bit, máy tính có thể biểu diễn được các số nguyên trong đoạn [-2.147.483.648, 2.147.483.647]. Muốn biểu diễn, tính toán được với số nguyên vượt ra ngoài khoảng này, ta phải dùng cách biểu diễn của số thực.

Có 2 loại số nguyên biểu diễn trong máy tính:

- Số nguyên không dấu: các số nguyên lớn hơn hoặc bằng 0.
- Số nguyên có dấu: gồm số 0 và các số nguyên âm, số nguyên dương.

Sau đây, ta xét cách mã hóa với số nguyên không dấu và có dấu 8 bit. Số nguyên 16 bit và 32 bit được suy ra tương tự.

∗ ***Số nguyên không dấu 8 bit***

Với số không dấu, máy tính dùng cả 8 bit để biểu diễn độ lớn, có thể biểu diễn được 2^8 = 256 số nguyên. Dải biểu diễn là 0000 0000 → 1111 1111 (hay 0 → 255_{10}).

Số mã hóa	**Số nguyên**
0000 0000	0
0000 0001	1
0000 0010	2
…	…
1111 1111	255

Cách biểu diễn:

+ Đổi số hệ 10 sang hệ 2.

+ Thêm vào bên trái số nhị phân các bit 0 cho đủ 8 bit.

Với số 34_{10} trong một ví dụ ở trên, ta đã tính được $34_{10} = 100010_2$. Vậy số 34_{10} được biểu diễn trong máy tính bởi 8 bit như sau: 0010 0010.

∗ ***Số nguyên có dấu 8 bit***

Với số nguyên có dấu, máy tính dùng bit đầu tiên để biểu diễn dấu, 7 bit còn lại biểu diễn độ lớn. Bit dấu bằng 0 thể hiện số dương, bằng 1 thể hiện số âm. Dải biểu diễn của số nguyên có dấu 8 bit là 1000 0000 → 0111 1111 (hay $-128_{10} \rightarrow +127_{10}$).

Bit dấu							

Với nửa dương, dải biểu diễn từ 0000 0000 → 0111 1111

Số mã hóa	**Số nguyên**
0000 0000	0
0000 0001	1
0000 0010	2
…	…
0111 1111	127

Với nửa âm, dải biểu diễn từ 1000 0000 → 1111 1111

Số mã hóa	**Số nguyên**
1000 0000	-128
1000 0001	-127
1000 0010	-126
…	…
1111 1111	-1

Mỗi số nguyên âm được biểu diễn bởi số bù 2 của biểu diễn số nguyên dương tương ứng chứ không đơn giản chỉ thay mỗi bit đầu tiên từ 0 thành 1. Cách tìm biểu diễn 8 bit của một số nguyên âm được máy tính thực hiện qua các bước:

- Bước 1: Tìm biểu diễn 8 bit của số nguyên dương tương ứng.
- Bước 2: Tìm số bù 1 của số vừa tìm được bằng cách đảo tất cả các bit (dùng toán tử NOT).
- Bước 3: Tìm số bù 2 bằng cách lấy số bù 1 cộng thêm 1.

Ví dụ: Để tìm biểu diễn của số nguyên -34_{10}:

- Bước 1: Tìm được biểu diễn 8 bit của $+34_{10}$ là 0010 0010
- Bước 2: Tìm được số bù 1 là 1101 1101
- Bước 3: Cộng 1 vào số bù 1

```
Số bù 1:    1101 1101
          +         1
            ---------
Số bù 2:    1101 1110
```

Vậy biểu diễn 8 bit của số -34_{10} trong máy tính là 1101 1110.

Lưu ý là cách lấy bù 2 có tính chất 2 chiều. Tìm số bù 2 của biểu diễn số nguyên âm ta cũng được biểu diễn của số nguyên dương tương ứng.

Cách biểu diễn này giúp cho việc tính toán trong máy tính cho những kết quả chính xác. Ví dụ: Phép cộng 2 số đối nhau phải cho kết quả bằng 0. Ta sẽ kiểm nghiệm với phép cộng $+34_{10}$ với -34_{10} được thực hiện trong máy tính.

$+34_{10}$:	0 0 1 0 0 0 1 0
-34_{10}:	+ 1 1 0 1 1 1 1 0
Tổng:	0 0 0 0 0 0 0 0

Kết quả lấy 8 bit là 0000 0000, bằng 0_{10}. Số nhớ tràn ra ngoài 8 bit không được tính vào tổng cuối.

b. Số thực

Các số thực được biểu diễn bởi số dấu phẩy động (*floating point number*) theo chuẩn IEEE 754. Một số thực có thể được viết theo nhiều cách với vị trí dấu phẩy khác nhau (nên gọi là "dấu phẩy động"). Ví dụ với số $580,04_{10}$:

$$580,94_{10} = 5,8094 \times 10^2 = 58,094 \times 10^1 = 5809,4 \times 10^{-1} = 58094,0 \times 10^{-2}$$

Để biểu diễn số phẩy động, số phải được phân tích dưới dạng mũ:

$$X = \pm m_x \times 10^{\pm Px}$$

Trong đó: m_x là phần định trị.

$\pm Px$ là phần mũ.

Ta thấy, một số có thể được phân tích ra thành nhiều số dạng mũ khác nhau, tuy nhiên nếu ràng buộc cho phần định trị là một số trong khoảng 1 và 10^{-1} thì việc phân tích luôn luôn là duy nhất và được gọi là dạng chuẩn.

Bit dấu	Phần mũ	Phần định trị

Chuẩn IEEE 754 sử dụng các dạng 32 bit, 44 bit, 64 bit và 80 bit. Ví dụ với dạng 32 bit, bit đầu tiên biểu diễn dấu, 8 bit kế tiếp cho phần mũ, 23 bit còn lại cho phần định trị; dải biểu diễn (xấp xỉ) từ $-10^{-38} \rightarrow 10^{38}$.

1.3.5. Mã hóa dữ liệu logic

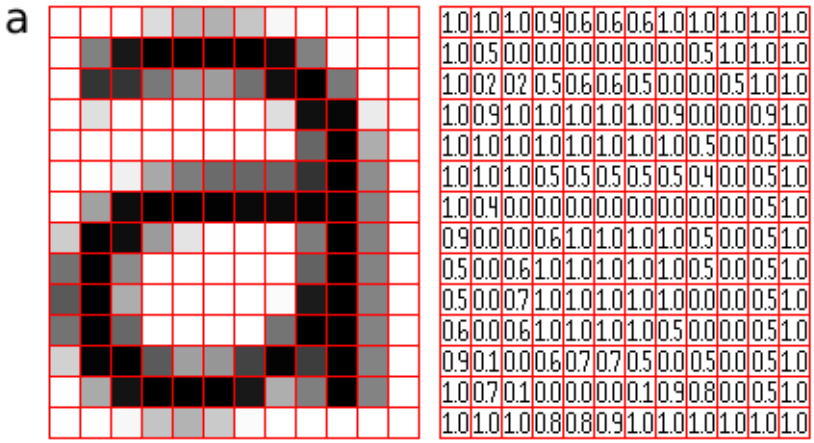
Dữ liệu logic là loại dữ liệu chỉ nhận một trong hai giá trị là TRUE/FALSE, hoặc 1/0. Do đó, việc mã hóa dữ liệu logic người ta thường chỉ dùng 1 byte.

1.3.6. Mã hóa hình ảnh tĩnh

Kích thước của các hình ảnh là đáng kể, vì thế người ta cần có phương pháp mã hóa để giảm kích thước của các ảnh. Có rất nhiều kiểu mã hóa ảnh trong đó ảnh bitmap và ảnh vector là hai kiểu thông dụng nhất.

Ảnh bitmap

Ảnh bitmap dùng lưới các điểm ảnh (pixel) để biểu thị hình ảnh. Mỗi điểm ảnh được gán một vị trí và gán giá trị mầu cụ thể. Do đó, ảnh bitmap là ảnh được tạo bởi ma trận các điểm ảnh. Một ảnh theo chuẩn VGA với độ phân giải 640x480 có nghĩa là một ma trận gồm 480 đường ngang và mỗi đường gồm 640 điểm ảnh.



Hình 1.2. Ví dụ dạng số hóa của một ảnh bitmap

Một điểm ảnh được mã hóa tùy thuộc vào chất lượng của ảnh:

- Ảnh đen trắng: Sử dụng một bit để mã hóa một điểm: giá trị 0 cho điểm ảnh màu đen và 1 cho điểm ảnh màu trắng.
- Ảnh 256 mức xám: Mỗi điểm được thể hiện bằng một byte (8 bit).
- Ảnh màu: Thường sử dụng hệ màu RGB, gồm phối trộn của 3 màu đỏ (Red), xanh lá (Green) và xanh dương (Blue) theo tỷ lệ khác nhau để tạo ra hàng triệu màu. Vì thế một màu bất kỳ có thể được biểu biễn bởi biểu thức:

$$x = aR + bG + cB$$

Trong đó a, b, c là các lượng của các màu cơ bản. Thông thường một ảnh đẹp sẽ có lượng màu với giá trị từ 0 đến 255. Và như thế, một ảnh màu thuộc loại này được thể hiện bằng 3 ma trận tương ứng cho 3 loại màu cơ bản. Mỗi phần tử của mảng có giá trị của 8 bit. Chính vì thế cần có 24 bit để mã hóa cho một điểm ảnh màu.

Ảnh vector

Ảnh Vector được tạo bởi các đoạn thẳng và đường cong được định nghĩa bằng các đối tượng toán học gọi là Vector. Hình Vector mô tả hình ảnh dựa trên các thuộc tính hình học của hình ảnh đó.

Hình 1.3. Ảnh vector

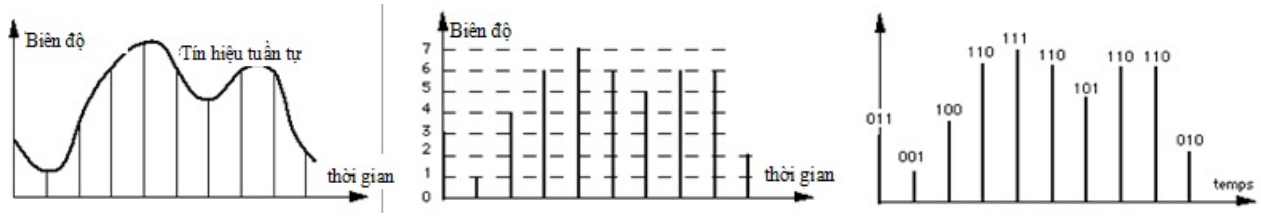
Với ảnh vector, kiểu này chỉ phù hợp với các ảnh có thành phần là các điểm rời rạc, các đường hoặc hình thể hiện bằng các đường biên (bản đồ, bản vẽ kỹ thuật...). Cách lưu trữ là lưu thông tin về các thành phần của ảnh. Ví dụ: lưu tọa độ các đầu mút đối với đoạn thẳng, lưu tọa độ tâm và bán kính đối với hình tròn... Với cách lưu thông tin như trên, các ảnh vector thường gọn gàng và linh hoạt trong việc phóng to thu nhỏ.

1.3.7. Mã hóa âm thanh và phim ảnh

Các tín hiệu âm thanh và phim ảnh là những tín hiệu dạng tương tự (Analog), tần số (cao độ) và thời gian (trường độ) đều là các số nên mã hóa được.

Việc số hóa các tín hiệu này có thể thực hiện qua 3 bước: lấy mẫu, lượng tử và mã hóa.

- Lấy mẫu: Từ tín hiệu liên tục ban đầu có thể rời rạc hóa để thu được tập các số đo biên độ theo thời gian (hình 1.4a).
- Lượng hóa: Lấy tương ứng các giá trị của biên độ với giá trị thang đo (hình 1.4b).
- Mã hóa: Mỗi một giá trị sau đó được mã hóa thành các giá trị nhị phân và đặt vào trong các tệp tin. Ví dụ: 011001100110111110101110110010… (hình 1.4c)



a) b) c)

Hình 1.4. Số hóa tín hiệu tương tự

Như vậy, âm thanh hay phim ảnh chúng ta nghe hay xem từ máy tính thực chất là chuỗi những âm thanh, hình ảnh rời rạc nhau, nhưng các khoảng rời rạc đó là quá nhỏ khiến chúng ta vẫn có cảm giác những âm thanh, hình ảnh đó là liên tục. Khoảng cách (tần số) lấy mẫu càng nhỏ thì âm thanh, hình ảnh càng "liên tục", tuy nhiên khi đó kích thước dữ liệu cần lưu trữ, xử lý sẽ tăng lên đáng kể.

Ngoài tần số lấy mẫu, dung lượng tệp tin dạng này còn phụ thuộc vào dải giá trị biên độ. Ví dụ trong hình 1.4, các giá trị biên độ được lượng hóa thành các giá trị nguyên trong đoạn [0,7], gồm 8 giá trị. Tập giá trị này chỉ cần 3 bit để mã hóa (vì $8 = 2^3$). Nhưng nếu tập giá trị biên độ chỉ cần mở rộng thành [0,8], ta sẽ phải cần đến 4 bit để mã hóa. Như vậy, dung lượng lưu trữ cho mỗi mẫu tăng thêm 1 bit. Với số lượng hàng triệu mẫu cho mỗi file âm thanh thì dung lượng tăng thêm là rất đáng lưu ý.

1.4. ỨNG DỤNG CỦA CÔNG NGHỆ THÔNG TIN

Công nghệ thông tin được ứng dụng trong tất cả các lĩnh vực hoạt động của con người mà cần xử lý thông tin tự động.

1.4.1. Các bài toán khoa học kỹ thuật

Đối với các bài toán khoa học kỹ thuật, thuật toán thường phức tạp. Ví dụ nhóm các bài toán dự báo (tài chính, thời tiết...), nhóm các bài toán thiết kế (robot, tên lửa, công trình thủy điện...), nhóm các bài toán thăm dò (khoáng sản, thị trường, vũ trụ...)... Để giải các bài toán đó đòi hỏi phải thực hiện một khối lượng các phép toán khổng lồ lên đến hàng trăm triệu, thậm chí hàng tỉ phép toán. Việc sử dụng tính toán dựa trên sức người là điều không tưởng trong trường hợp này, việc đó không những tốn kém về thời gian công sức mà kết quả cũng như độ tin cậy không được đánh giá cao. Với sự ra đời của máy tính điện tử đã góp phần giải quyết được các bài toán khoa học kỹ thuật một cách tối ưu và triệt để. Trong giai đoạn đầu, những năm 1950-1970, do máy tính còn ít và giá thành máy tính khá đắt nên chúng chưa được ứng dụng rộng rãi trong cuộc sống mà mới chỉ được dùng cho mục đích khoa học kỹ thuật.

Ngày nay, đối với rất nhiều bài toán khoa học kỹ thuật, người ta muốn kết quả đưa ra không phải dưới dạng số liệu mà còn thể hiện minh hoạ cho lời giải. Vì thế kể từ khi máy tính được trang bị những màn hình có khả năng thể hiện đồ hoạ thì xử lý hình học là một trong những vấn đề được quan tâm nhiều đối với các bài toán khoa học kỹ thuật. Với những máy tính như vậy người ta có thể làm việc theo kiểu tương tác với các sự kiện đang mô phỏng trên máy tính như sửa chữa các bản thiết kế, điều khiển một nhóm đối tượng phức tạp thông qua các hình ảnh mô phỏng trên màn hình.

Cùng với sự phát triển của mình, CNTT đã được áp dụng trên hầu hết các mặt của đời sống xã hội. Tuy nhiên các bài toán khoa học kỹ thuật vẫn có chỗ đứng nhất định, là cơ sở, là nền tảng cho sự phát triển các lĩnh vực khác và sự ra đời siêu máy tính đã trở thành công cụ hữu ích hỗ trợ giải quyết các bài toán khoa học kỹ thuật.

1.4.2. Các bài toán quản lý

Trước đây, khi CNTT còn chưa phổ biến, các hoạt động văn thư và hành chính trong các cơ quan chủ yếu là các thao tác thủ công và phụ thuộc rất nhiều vào sự góp mặt của con người. Các thao tác thủ công đó rất đơn giản trong quy trình và dễ thực hiện các thao tác. Do đó, với các nhân viên có chút kinh nghiệm và hiểu biết về nghiệp vụ đều có thể thực hiện được. Tuy nhiên, nhược điểm của nó lại quá lớn: tốn thời gian, công sức, không cập nhật thường xuyên, những rủi ro về mất mát thông tin, giấy tờ trong lưu trữ, gây lãng phí không nhỏ về kết quả thông tin lưu trữ...

Ngày nay, với sự phát triển không ngừng của CNTT đã đưa con người sang một kỷ nguyên mới, kỷ nguyên của công nghệ, kỷ nguyên của máy móc thay thế một phần hoạt động của con người. Với việc ứng dụng CNTT vào đời sống xã hội đã đem lại nhiều lợi ích cho các cơ quan, tổ chức, làm giảm chi phí, thời gian, công sức, tăng khả năng lưu trữ...

Các hoạt động quản lý rất đa dạng và xuất hiện rất nhiều trong xã hội, ở đâu có tổ chức là ở đó có nhu cầu quản lý. Khác với bài toán khoa học kỹ thuật, các bài toán quản lý có quy trình xử lý đơn giản nhưng khối lượng thông tin lưu trữ lại lớn. Một bài toán quản lý thường có những công việc sau:

- Tạo cơ sở dữ liệu (CSDL): Tập hợp các dữ liệu, thông tin cần được tổ chức lưu trữ của hệ thống cần quản lý và quản lý một cách thống nhất trên máy tính.
- Duy trì cơ sở dữ liệu: Cập nhật dữ liệu thường xuyên để đảm bảo dữ liệu phản ánh đúng và kịp thời hoạt động của hệ thống quản lý.
- Sử dụng cơ sở dữ liệu: Có hai hình thức sử dụng là tra cứu và thống kê.

Kiểu sử dụng dạng tra cứu nhằm tìm ra các thông tin vốn có trong CSDL theo một tiêu chuẩn nào đó. Ví dụ lập danh sách sinh viên có học bổng, lập danh sách các cán bộ nghỉ hưu... Việc tra cứu dữ liệu thường được thực hiện thông qua các chương trình và các chương trình đó sẽ truy cập trực tiếp vào CSDL và xử lý theo yêu cầu tra cứu. Do đó, hoạt động tra cứu thông thường chỉ trích ra các dữ liệu có sẵn trong cơ sở dữ liệu.

Kiểu sử dụng dữ liệu dạng thống kê thường thiên về tính đếm để rút ra các đặc trưng thống kê như tính tổng có điều kiện, lấy trung bình, tính các giá trị lớn nhất hay nhỏ nhất. Ví dụ, sau khi cập nhật kết quả của một kỳ thi có thể phải đánh giá chất lượng sinh viên thông qua những thống kê về điểm xuất sắc, giỏi, khá, trung bình, yếu, kém.

Mục đích cuối cùng của các hệ thông tin quản lý là hỗ trợ cho quá trình ra quyết định của một tổ chức hay cá nhân. Ví dụ, thông qua thống kê hàng tồn kho mà quyết định giảm giá, tra cứu những sinh viên đủ điều kiện để quyết định hình thức và mức khen thưởng. Vì thế các phần

mềm quản lý thường phải được xây dựng trên cơ sở các hoạt động hỗ trợ quyết định, chứ không đơn giản chỉ là tra cứu hay thống kê.

Quản lý là lĩnh vực sử dụng tin học nhiều nhất. Người ta ước tính 85% đầu tư tin học là dành cho quản lý. Những hệ thống như quản lý ngân hàng, kế toán xí nghiệp, quản lý bán hàng và kho tàng, quản lý nhân sự... đều là những ứng dụng trong lĩnh vực quản lý.

1.4.3. Tự động hóa

Trước đây, khi CNTT còn chưa phát triển, kỹ thuật tự động hóa khá đơn giản chủ yếu là theo kiểu điện cơ, do đó còn hạn chế trong việc đáp ứng các quá trình điều khiển phức tạp.

Ngày nay, với việc ứng dụng CNTT đã cho ra đời những loại hình điều khiển có tính thích nghi, hay chính là hệ hỗ trợ ra quyết định. Ví dụ: ứng dụng tự động hóa trong các nhà máy sản xuất ô tô, máy bay không người lái, các robot...

Mức độ tự động hoá: có 2 mức độ

Tự động hoá một phần: có sự phân chia việc xử lý thông tin giữa con người và máy tính.

Nhược điểm: Thường xảy ra mâu thuẫn khi kết nối từng phần nhỏ.

Tự động hoá toàn bộ: toàn bộ hệ thông tin được xử lý bằng máy tính, con người chỉ có vai trò phụ.

Ưu điểm: Xử lý thông tin tổng thể và tập trung, điều khiển chung nằm tại một khối nên rất hiệu quả. Dữ liệu tập trung ở một nơi và chỉ có một bản nên giảm được chi phí và tránh được sai lệch.

Nhược điểm: Khó xây dựng.

Các hệ thống nhúng là một ví dụ phổ biến nhất về tự động hóa trên cơ sở máy tính và đã trở nên rất phổ biến đến mức người ta ít khi để ý đến sự có mặt của nó.

1.4.4. Công tác văn phòng

Công tác văn phòng là công tác thường gặp tại các cơ quan, tổ chức, đoàn thể... Do đó, ứng dụng CNTT vào các hoạt động văn phòng được chú ý từ rất sớm và là lĩnh vực ứng dụng phổ biến của CNTT.

Việc ứng dụng CNTT đã giảm tải được rất nhiều thời gian, công sức và nhân lực và đã thay đổi hoàn toàn bộ mặt hoạt động của công tác văn phòng. Các hoạt động cơ bản của công tác văn phòng:

- Lưu trữ văn bản, tài liệu.
- Xử lý và lập kế hoạch.
- Nhận và lưu chuyển văn bản, tài liệu.
- Tạo và gửi văn bản, tài liệu.

Với sự ra đời phần mềm văn phòng điện tử đã đánh dấu sự thay đổi của việc ứng dụng CNTT vào trong các cơ quan, tổ chức, xí nghiệp.

1.4.5. Giáo dục

Giáo dục cũng không nằm ngoài xu thế phát triển của CNTT. Việc ứng dụng của CNTT vào trong giáo dục đã mở ra nhiều cơ hội mới hơn cho người học, người dạy và người quản lý.

Người học có nhiều cơ hội tiếp cận hơn với kiến thức của nhân loại mà không phải gói gọn trong nội dung bài học của mình, tiếp cận được với những phương pháp dạy và học tiên tiến trên thế giới, có nhiều cơ hội giao lưu học hỏi, tăng tính chủ động và sáng tạo.

Với người dạy, có nhiều phương pháp hỗ trợ hơn trong việc bổ sung kiến thức và việc truyền tải bài giảng đến người học cũng phong phú và sinh động hơn. Tạo được mối liên kết chặt chẽ hơn giữa nhà trường và gia đình. Thông qua đó cũng có cách đánh giá chính xác hơn đến người học.

Với người quản lý, ứng dụng CNTT cũng mang lại nhiều lợi ích trong việc quản lý giáo dục, quản lý người dạy, người học, nội dung chương trình dạy và học, có những đánh giá khách quan hơn.

Việc ra đời các hệ thống thư viện điện tử, bài giảng điện tử, sổ liên lạc điện tử, website của các cơ sở giáo dục... là minh chứng mạnh mẽ nhất của CNTT đã lan tỏa trong ngành giáo dục nước nhà và trên thế giới.

1.4.6. Thương mại điện tử

CNTT phát triển kéo theo sự phát triển của rất nhiều ngành nghề và thương mại điện tử cũng không nằm ngoài xu thế đó. Có thể kể đến một số hoạt động thương mại điện tử điển hình như:

- Quảng cáo trên mạng.
- Mua bán và thanh toán qua mạng.
- Thương thảo các hợp đồng qua mạng.

Cùng với đó là sự ra đời của các website bán hàng, website quảng cáo, website các công ty... Tuy nhiên, với sự phát triển của mình thì thương mại điện tử cũng phải đối diện với nhiều thách thức lớn hiện nay như vấn đề pháp lý, vấn đề bảo mật và an toàn thông tin...

CÂU HỎI VÀ BÀI TẬP

1. Nêu các khái niệm: Dữ liệu, Thông tin, Tin học, Công nghệ thông tin?
2. Thông tin số là gì?
3. Đơn vị đo dung lượng thông tin?
4. So sánh bảng mã ASCII và bảng mã UNICODE?
5. Mã hóa dữ liệu kiểu số?
6. Mã hóa dữ liệu phi số?
7. Các lĩnh vực ứng dụng của công nghệ thông tin?

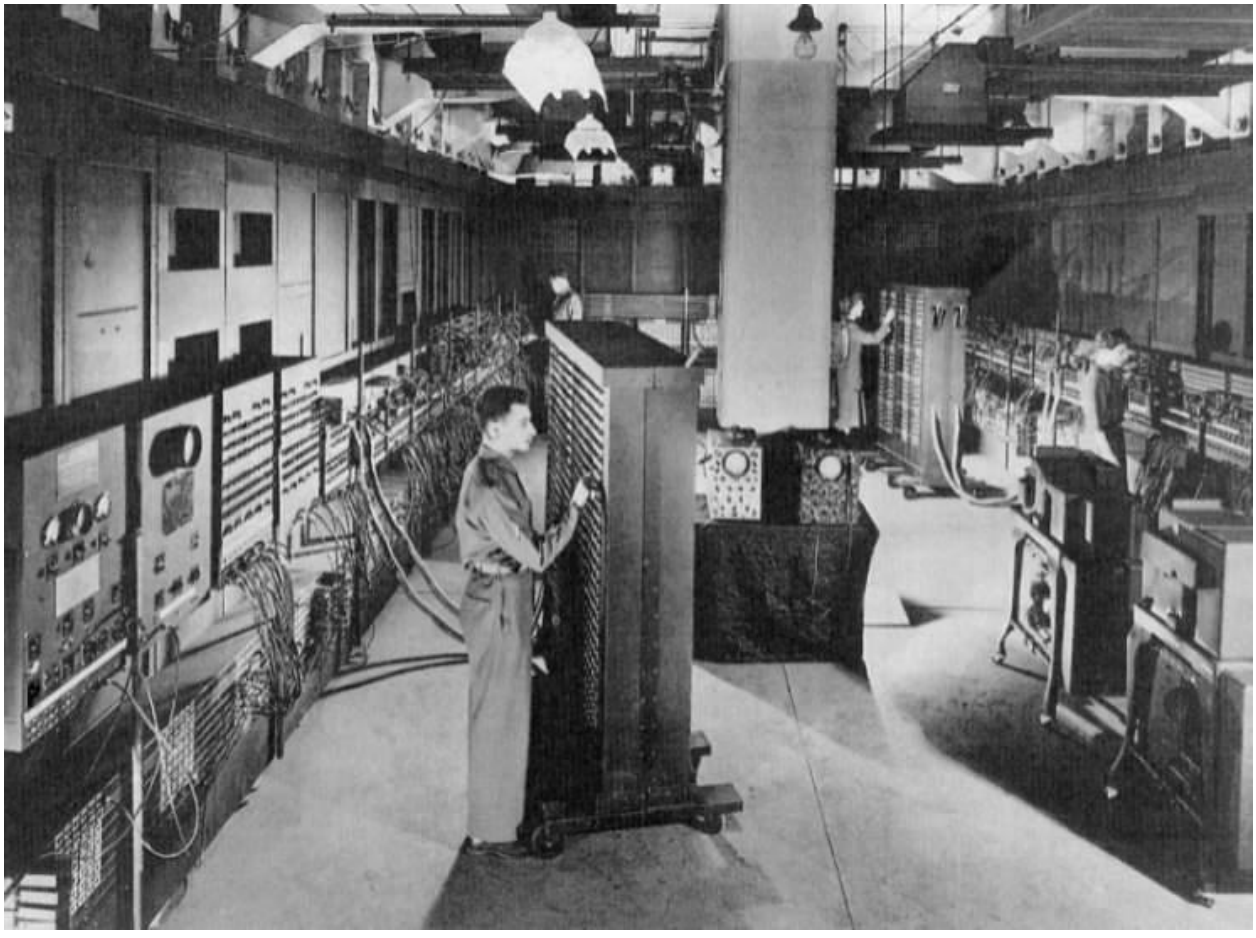
Chương 2

CẤU TRÚC MÁY TÍNH

Các loại máy tính như máy tính để bàn, máy tính xách tay, máy tính bảng hay điện thoại di động có cấu trúc chung gồm các khối chức năng: bộ xử lý trung tâm, bộ nhớ, hệ thống vào-ra và liên kết hệ thống. Chương này giới thiệu một số vấn đề cơ bản về cấu trúc máy tính gồm các khối nêu trên. Mục 2.1 giới thiệu sơ lược về lịch sử ra đời và quá trình phát triển của máy tính điện tử. Mục 2.2 trình bày về chức năng, sơ đồ cấu trúc chung và nguyên lý hoạt động của máy tính. Mục cuối cùng 2.3 sẽ giới thiệu về chức năng và cấu trúc của các thành phần cơ bản của máy tính.

2.1. GIỚI THIỆU

Máy tính điện tử (từ sau gọi tắt là máy tính) đầu tiên ra đời năm 1946, có tên là ENIAC (Electronic Numerical Intergator and Computer), là sản phẩm của một dự án thuộc Bộ Quốc phòng Hoa Kỳ phục vụ mục đích quân sự. ENIAC được thiết kế bởi John Mauchly và John Presper Eckert ở Đại học Pennsylvania. Chiếc máy tính này nặng 30 tấn, kích thước 140m^2, thực hiện được 5.000 phép cộng/giây, xử lý theo số thập phân, bộ nhớ chỉ lưu trữ dữ liệu và lập trình bằng cách thiết lập vị trí của các chuyển mạch và các cáp nối.



Hình 2.1. Máy tính điện tử đầu tiên - ENIAC

Năm 1952, máy tính von Neumann ra đời tại Học viện Nghiên cứu tiên tiến Princeton. Chiếc máy tính này được xây dựng theo ý tưởng "chương trình được lưu trữ", xử lý theo số nhị phân. Những nguyên lý của von Neumann (phần 2.2.3) đã trở thành mô hình cơ bản của máy tính cho đến ngày nay.

Năm 1980, hãng IBM cho ra đời chiếc máy tính cá nhân đầu tiên, sử dụng bộ vi xử lý 8 bit 8085 của Intel.

Với 70 năm phát triển, máy tính đã trải qua 4 thế hệ: dùng đèn điện tử (1943-1956), dùng transistor (1957-1965), dùng vi mạch tích hợp (1966-1980) và dùng siêu vi mạch tích hợp (1981-nay). Ngày nay, các máy tính cá nhân, máy tính bảng, điện thoại thông minh... có kích thước nhỏ gọn, cấu hình mạnh mẽ. Tất cả vẫn có cấu trúc chung và sử dụng những nguyên lý được đề ra bởi von Neumann.

Trong phần tiếp theo chúng tôi sẽ giới thiệu chức năng của máy tính, nguyên lý hoạt động và sơ đồ cấu trúc chung của các máy tính. Phần 2.3. sẽ giới thiệu cấu tạo và các đặc tính kỹ thuật chính của các thành phần cơ bản của máy tính gồm CPU, bộ nhớ và các thiết bị ngoại vi.

2.2. CHỨC NĂNG VÀ SƠ ĐỒ CẤU TRÚC CỦA MÁY TÍNH

2.2.1. Chức năng của máy tính:

Máy tính có những chức năng sau:

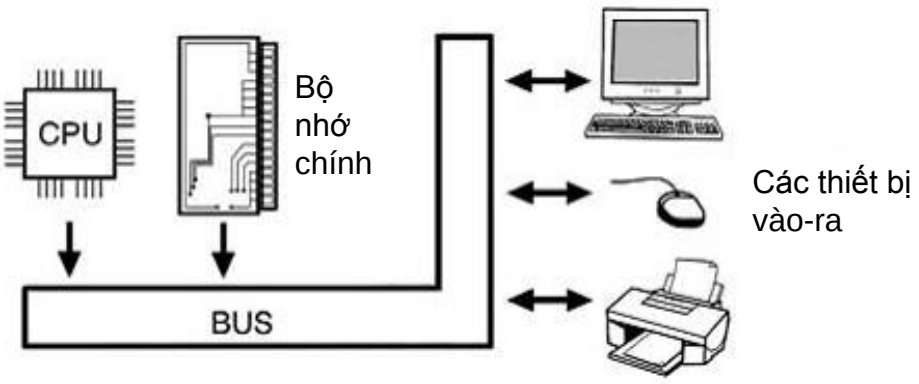
- Nhận thông tin vào (input) từ người sử dụng hoặc từ máy tính khác thông qua các thiết bị vào;

- Xử lý thông tin đã nhận theo dãy lệnh đã nhớ sẵn bên trong;

- Đưa thông tin sau xử lý (output) tới người sử dụng hoặc tới máy tính khác thông qua các thiết bị ra;

- Lưu trữ thông tin dạng số hóa.

2.2.2. Sơ đồ cấu trúc chung của máy tính



Hình 2.2. Sơ đồ cấu trúc của máy tính

Các khối chức năng:

Bộ xử lý trung tâm (CPU):

* Chức năng
  - Điều khiển hoạt động của máy tính;
  - Xử lý dữ liệu.

* Nguyên tắc hoạt động cơ bản
  - CPU hoạt động theo chương trình nằm trong bộ nhớ chính.
* Các thành phần chính
  - Đơn vị điều khiển (Control Unit);
  - Đơn vị số học và logic (Arithmetic and Logic Unit);
  - Tập các thanh ghi (Registers).

Bộ nhớ:

* Chức năng: Lưu trữ chương trình và dữ liệu.
* Các thao tác cơ bản với bộ nhớ
  - Đọc (Read);
  - Ghi (Write).
* Các thành phần chính
  - Bộ nhớ trong (Internal Memory);
  - Bộ nhớ ngoài (External Memory).

Hệ thống vào-ra:

* Chức năng: Trao đổi thông tin giữa máy tính với thế giới bên ngoài.
* Các thao tác cơ bản
  - Vào dữ liệu (Input);
  - Ra dữ liệu (Output).
* Các thành phần chính
  - Các thiết bị ngoại vi (Peripheral Devices) ;
  - Các môđun vào-ra (IO Modules).

2.2.3. Nguyên lý hoạt động của máy tính

Từ khi ra đời đến nay, các máy tính đều hoạt động theo những nguyên lý được đề xuất năm 1946 bởi nhà khoa học lỗi lạc người Mỹ gốc Hungary John von Neumann (1903-1957).

a. Nguyên lý Von Neumann

- Nguyên lý điều khiển bằng chương trình: máy tính hoạt động theo chương trình lưu trữ sẵn trong bộ nhớ của nó. Nguyên lý này đảm bảo cho máy tính có khả năng tự điều khiển không cần có sự can thiệp của người sử dụng trong quá trình xử lý thông tin.

- Nguyên lý truy cập theo địa chỉ: các chương trình, dữ liệu trước, trong và sau khi xử lý đều được đưa vào bộ nhớ trong những vùng nhớ được đánh địa chỉ. Việc truy cập dữ liệu là gián tiếp thông qua địa chỉ của nó trong bộ nhớ. Nguyên lý này đảm bảo tính mềm dẻo trong xử lý thông tin: người lập trình chỉ cần viết các yêu cầu một cách tổng quát theo vị trí các đối tượng mà không cần biết giá trị cụ thể của chúng.

b. Cấu trúc lệnh và quá trình thực hiện lệnh

Để xử lý thông tin tự động, mỗi máy tính cần được cài đặt sẵn một tập lệnh, thường vào trong bộ nhớ ROM. Mỗi lệnh máy là một chuỗi số nhị phân, yêu cầu CPU thực hiện một thao tác nào đó đối với các toán hạng. Các lệnh này phải chỉ ra đầy đủ các thông tin sau:

- Thao tác cần thực hiện: chuyển dữ liệu, xử lý số học với số nguyên/số dấu phẩy động, xử lý logic, điều khiển vào-ra, chuyển điều khiển (rẽ nhánh), điều khiển hệ thống, xử lý các dữ liệu chuyên dụng.

- Nơi đặt dữ liệu của lệnh và nơi đặt kết quả xử lý: tại bộ nhớ trong hoặc tại các thanh ghi trong CPU.

Cấu trúc chung của lệnh máy như sau:

Mã thao tác	Địa chỉ các toán hạng

Ví dụ:

Hình 2.3. Một lệnh cộng trong tập lệnh MIPS32

Một chương trình máy tính là một dãy các lệnh. Do chương trình cũng nằm trong bộ nhớ nên chính các lệnh cũng có địa chỉ, đó chính là địa chỉ byte đầu tiên của lệnh.

Quá trình thực hiện một chương trình thường là một quá trình thực hiện liên tiếp từng lệnh. Để quản lý thứ tự thực hiện các lệnh, trong bộ vi xử lý có một thanh ghi gọi là Bộ đếm chương trình (Program Counter – PC) để ghi địa chỉ của lệnh sẽ thực hiện tiếp theo. Giá trị khởi tạo của PC là địa chỉ lệnh đầu tiên của chương trình.

Máy tính điện tử được điều khiển bởi các lệnh của chương trình. Chu kỳ thực hiện một lệnh bao gồm các bước sau:

- **Nhận lệnh** (Fetch Instruction): Bộ điều khiển trong CPU gửi nội dung PC vào Bộ giải mã địa chỉ để đọc byte đầu tiên của lệnh lên thanh ghi lệnh. Nếu không có lệnh nhảy (ví dụ lệnh **goto** trong Pascal) PC sẽ tăng lên một đơn vị để bộ điều khiển chuẩn bị đọc byte tiếp theo, trường hợp ngược lại thì PC sẽ được nạp vào địa chỉ của lệnh kế tiếp sẽ nhảy đến.

- **Giải mã lệnh** (Decode Instruction): Bộ điều khiển căn cứ vào mã lệnh để biết lệnh dài bao nhiêu byte để đọc nốt các thông tin địa chỉ của lệnh và hoàn thành việc đọc lệnh. PC tiếp tục tăng theo số lượng byte đã đọc vào.

- **Nhận dữ liệu** (Fetch Data): Nhận dữ liệu từ bộ nhớ hoặc các cổng vào-ra.

- **Xử lý dữ liệu** (Process Data): Thực hiện phép toán số học hay phép toán logic với các dữ liệu.

- **Ghi dữ liệu** (Write Data): Ghi dữ liệu ra bộ nhớ hay cổng vào-ra.

Sau đó quay lại chu kỳ mới, bắt đầu từ nhận lệnh.

2.3. CÁC THÀNH PHẦN CƠ BẢN CỦA MÁY TÍNH

Như đã đề cập ở phần 2.2.2, một hệ thống máy tính gồm 4 khối chức năng: bộ xử lý trung tâm, bộ nhớ, các thiết bị vào-ra và liên kết hệ thống. Phần này chúng ta sẽ đề cập sâu hơn đến cấu tạo và những đặc tính kỹ thuật của các thiết bị trong các khối chức năng đó của các máy tính cá nhân. Người sử dụng thường quen thuộc với những thiết bị ngoại vi phổ biến như bàn phím, con chuột, màn hình, máy in; nhưng bên trong hộp máy có những thiết bị gì? Trước khi đề cập từng thiết bị cụ thể, chúng tôi giới thiệu một số thiết bị thường có trong hộp máy trong hình 2.4:

Hình 2.4. Các thành phần phổ biến bên trong hộp máy tính cá nhân

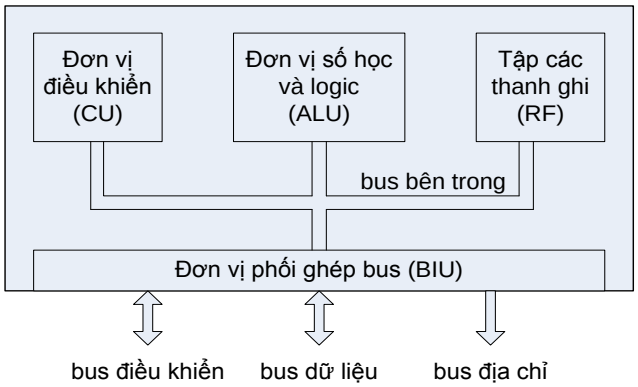
2.3.1. Bộ xử lý trung tâm

Bộ xử lý trung tâm (CPU: Central Processing Unit), hay bộ vi xử lý (microprocessor, processor) là một mạch xử lý dữ liệu theo chương trình được thiết lập trước. Nó là một mạch tích hợp phức tạp gồm hàng triệu transitor trên một bảng mạch nhỏ. Đây là thành phần quan trọng nhất, được xem như bộ não và thường là đắt nhất của một máy tính. Hai nhà sản xuất CPU lớn nhất hiện nay là Intel và AMD (Advanced Micro Devices).

Những chức năng của CPU:

- Nhận lệnh, giải mã lệnh và điều khiển các khối khác thực hiện lệnh;
- Thực hiện các phép tính số học, logic và các phép tính khác;
- Sinh ra các tín hiệu địa chỉ để truy nhập bộ nhớ.

Một bộ vi xử lý gồm những khối cơ bản là: Khối điều khiển, khối số học và logic, các thanh ghi (hình 2.5).



Hình 2.5. Sơ đồ khối của CPU

Khối điều khiển (CU: control unit): là khối có chức năng điều khiển sự hoạt động của máy tính theo chương trình định sẵn.

Khối số học và logic (ALU: arithmetic and logic unit): gồm các mạch chức năng để thực hiện các phép toán cơ sở như phép toán số học, phép toán logic, phép tạo mã...

Các thanh ghi (registers): được dùng như những bộ nhớ nhanh, có thể tương tác trực tiếp với các mạch xử lý của CPU. Có thanh ghi ghi địa chỉ lệnh sắp thực hiện, có thanh ghi ghi lệnh đang thực hiện, có thanh ghi ghi dữ liệu, có thanh ghi ghi kết quả xử lý...

Những yếu tố ảnh hưởng đến hiệu năng của bộ vi xử lý bao gồm: tốc độ đồng hồ, tốc độ bus, kích thước từ nhớ, dung lượng bộ nhớ cache, tập lệnh, số lượng lõi, các kỹ thuật xử lý.

Đồng hồ trong bộ vi xử lý (clock): là thiết bị thiết lập bước thực hiện lệnh. Mạch xung nhịp đồng hồ dùng để đồng bộ các thao tác xử lý trong và ngoài CPU theo các khoảng thời gian không đổi. Khoảng thời gian chờ giữa hai xung gọi là chu kỳ xung nhịp. Xung nhịp hệ thống tạo ra các xung tín hiệu chuẩn thời gian gọi là tốc độ xung nhịp – tốc độ đồng hồ tính bằng triệu hoặc tỷ đơn vị mỗi giây (MHz/GHz). Tuy nhiên, cần hiểu là tốc độ đồng hồ không bằng số lệnh mà bộ vi xử lý thực hiện trong một giây. Trong nhiều máy tính, mỗi chu kỳ có thể có vài lệnh, nhưng các lệnh khác có thể cần nhiều chu kỳ.

Bạn có thể nghĩ rằng máy tính có bộ vi xử lý 1,6 GHz thực hiện chậm hơn máy tính có bộ vi xử lý 2.3 GHz. Điều này chỉ đúng khi so sánh các bộ vi xử lý trong cùng họ chip. Ví dụ, bộ vi xử lý 1.87 GHz i7 840QM nhanh hơn bộ vi xử lý 1.6 GHz i7 720QM. Bạn có thể ngạc nhiên vì bộ vi xử lý i7 1.6 GHz nhanh hơn bộ vi xử lý i5 2.4 GHz vì i7 có nhiều lõi hơn i5.

Bộ vi xử lý nhiều lõi (multi-core processor): Một bộ vi xử lý có thể có nhiều hơn một đơn vị xử lý, được gọi là bộ xử lý nhiều lõi. Nhiều lõi thường có hiệu năng nhanh hơn. Bộ vi xử lý i5 2.4 GHz có 2 lõi, hiệu năng tương đương 4.8 GHz. Còn bộ vi xử lý i7 1.6 GHz có 4 lõi, hiệu năng tương đương 6.4 GHz.

Tốc độ bus (FSB: front side bus): là đường truyền dữ liệu đến và ra khỏi bộ vi xử lý. Bus tốc độ cao giúp chuyển dữ liệu nhanh, giúp CPU hoạt động với công suất lớn nhất. Tốc độ bus được đo bằng megahertz. Megahertz (MHz) có nghĩa là một triệu chu kỳ/giây. Các máy tính ngày nay có tốc độ bus từ 1000-1600 MHz.

Dung lượng cache: CPU cache là bộ nhớ đệm tốc độ rất cao, cho phép bộ vi xử lý truy cập dữ liệu nhanh hơn từ bộ nhớ RAM. Dung lượng cache lớn làm tăng hiệu năng của máy tính. CPU cache được chia thành 2-3 mức. Cache L1 (mức 1) có tốc độ nhanh nhất; cache L2, L3 có tốc độ chậm hơn một chút nhưng vẫn nhanh hơn tốc độ truy nhập bộ nhớ chính (RAM) hay các đĩa. Dung lượng cache thường được đo bằng megabytes (MB).

Kích thước từ nhớ: Là số bit mà bộ vi xử lý có thể thực hiện được mỗi lần. Kích thước từ nhớ được dựa trên kích thước của các thanh ghi trong khối số học và logic (ALU) và của các mạch dẫn đến các thanh ghi đó. Ví dụ, bộ vi xử lý 64-bit có các thanh ghi 64-bit và xử lý mỗi lần 64 bit. Kích thước từ nhớ lớn giúp cho bộ vi xử lý có khả năng xử lý nhiều dữ liệu hơn trong mỗi chu kỳ - một yếu tố làm tăng hiệu năng của máy tính. Các máy tính cá nhân ngày nay thường có bộ vi xử lý 32-bit hoặc 64-bit.

Tập lệnh: Khi các nhà thiết kế chip phát triển tập lệnh cho các bộ vi xử lý (VXL), họ ngày càng thêm các lệnh phức tạp mà mỗi lệnh cần vài chu kỳ đồng hồ để thực hiện. Bộ VXL có tập lệnh như vậy sử dụng công nghệ CISC (complex instruction set computer). Bộ VXL có tập lệnh rút gọn gồm các lệnh đơn giản sử dụng công nghệ RISC (reduced instruction set computer). Bộ VXL RISC thực hiện hầu hết các lệnh nhanh hơn so với bộ VXL CISC. Tuy nhiên, nó có thể

cần nhiều lệnh đơn giản để hoàn thành một tác vụ so với bộ VXL CISC. Đa số bộ VXL trong các máy tính cá nhân hiện nay sử dụng công nghệ CISC. Nhiều bộ VXL trong các thiết bị cầm tay như iPod, Droid, BlackBerry là ARM (advanced RISC machine).

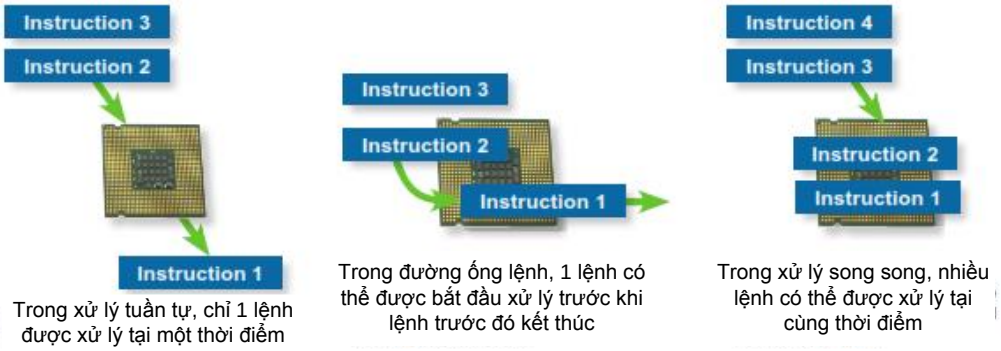
Các kỹ thuật xử lý:

- Serial processing (xử lý tuần tự): bộ VXL phải hoàn thành tất cả các bước của chu kỳ lệnh trước khi bắt đầu thực hiện lệnh kế tiếp.

- Pipelining (kỹ thuật đường ống lệnh): công nghệ này giúp cho bộ VXL có thể bắt đầu thực hiện một lệnh trước khi nó hoàn thành lệnh trước đó.

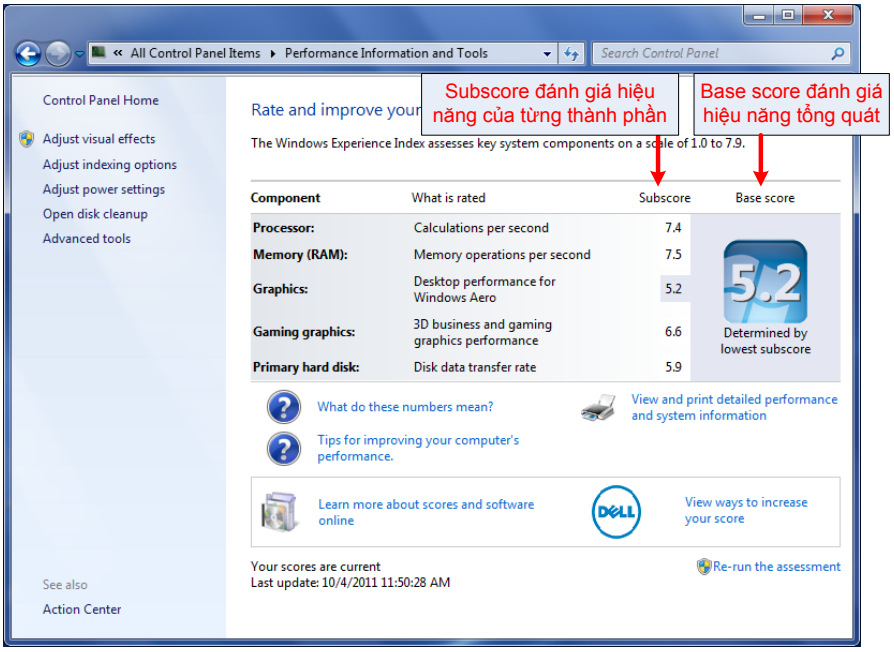
- Parallel processing (xử lý song song): công nghệ này giúp cho bộ VXL có thể thực hiện nhiều lệnh cùng một lúc.

Minh họa cho các kỹ thuật trên được thể hiện trong hình 2.6.



Hình 2.6. Các kỹ thuật xử lý lệnh của CPU

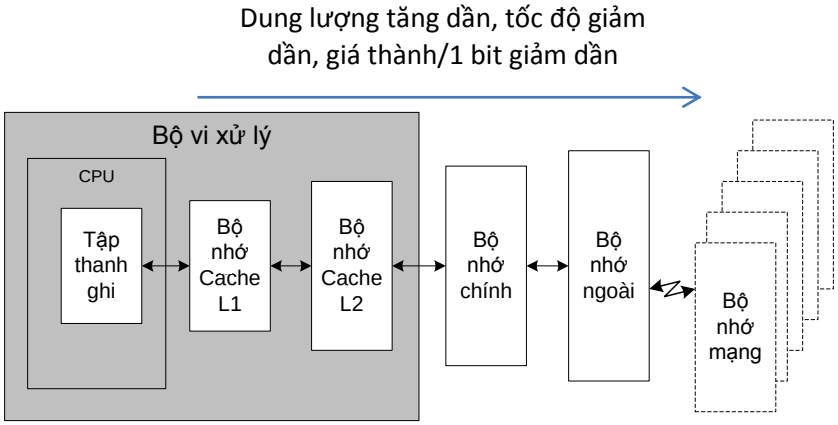
So sánh hiệu năng tổng quát của các bộ vi xử lý: Nhiều phòng thí nghiệm chạy một loạt các thí nghiệm để đánh giá tốc độ tổng quát của một bộ VXL. Các kết quả này được gọi là benchmark (điểm chuẩn) và có thể dùng để so sánh với các bộ VXL khác. Những kết quả thí nghiệm benchmark thường được đưa lên Web và được xuất bản trong các tạp chí máy tính.



Hình 2.7. Một báo cáo về điểm đánh giá của Windows Experience Index

2.3.2. Bộ nhớ

Hệ thống nhớ (hình 2.8) của máy tính gồm nhiều mức bộ nhớ để sử dụng ưu điểm, khắc phục nhược điểm của từng loại bộ nhớ.



Hình 2.8. Sơ đồ hệ thống nhớ

Các thanh ghi đã được đề cập khi nói về Bộ vi xử lý ở trên. Trong phần này chúng ta chỉ xem xét các mức bộ nhớ còn lại.

a. Bộ nhớ trong

Khái niệm:

Bộ nhớ trong (memory) là bộ nhớ có thời gian truy cập nhỏ, được dùng để nạp hệ điều hành, ghi chương trình và dữ liệu trong thời gian xử lý. Bộ nhớ trong gồm các mức bộ nhớ mà CPU có thể truy cập trực tiếp. Bộ nhớ trong gồm các loại: cache, RAM và ROM. Trong đó cache và RAM là các bộ nhớ có thể đọc và ghi dữ liệu, bị mất thông tin khi mất nguồn nuôi; còn ROM là bộ nhớ chỉ cho phép đọc, dữ liệu không bị xóa khi mất nguồn.

Cấu tạo:

Bộ nhớ trong được cấu tạo từ các phần tử vật lý có 2 trạng thái đối lập. Một trạng thái dùng để thể hiện bit 0, còn trạng thái kia thể hiện bit 1. Có nhiều kỹ thuật chế tạo các phần tử có 2 trạng thái như dùng từ tính, dùng mạch bán dẫn. Ngày nay, người ta dùng các bộ nhớ bán dẫn là các mạch bán dẫn điều khiển được có 2 trạng thái đóng/mở để thể hiện các bit.

Nhờ tiến bộ của công nghệ vi điện tử, các bộ nhớ bán dẫn có thể được chế tạo là các vi mạch tích hợp (vài cm^2) có dung lượng vài gigabyte (GB).

Tổ chức:

Ta có thể hình dung bộ nhớ trong như dãy liên tiếp các byte nhớ được đánh số thứ tự - là địa chỉ của byte nhớ. Địa chỉ được đánh số lần lượt từ 0, 1, 2... Mỗi byte gồm 8 bit, mỗi bit được thiết lập bằng 0 hoặc 1. Byte là đơn vị thông tin thuận lợi cho xử lý dữ liệu vì nó có thể chứa vừa đủ một ký tự mã hóa theo bảng mã ASCII hay một số nguyên nhỏ hơn 256 (= 2^8-1). Để thể hiện các dữ liệu dài hơn như một ký tự mã hóa theo bảng mã Unicode cần 2 byte, một số nguyên lớn hơn cần 2 hoặc 4 byte, một số thực cần 4, 8 hoặc 10 byte liền nhau.

Nội dung	Địa chỉ
00101011	0000
11010101	0001
00001010	0010
01011000	0011
11111011	0100
00001000	0101
11101010	0110
00000000	0111
10011101	1000
00101011	1001
11101011	1010
00101000	1011
11111111	1100
10101010	1101
00101011	1110
01010101	1111

Hình 2.9. Hình ảnh địa chỉ hóa bộ nhớ trong

Như vậy, mỗi byte nhớ có 2 đặc trưng:

- Địa chỉ: là thứ tự của vị trí byte nhớ trong Bộ nhớ trong. Địa chỉ của mỗi byte nhớ là cố định.
- Nội dung: là giá trị số dạng mã nhị phân, được lưu trữ bằng các trạng thái vật lý trong byte nhớ. Nội dung byte nhớ có thể thay đổi.

Do mỗi byte nhớ có địa chỉ riêng nên có thể truy cập tới dữ liệu trong từng byte nhớ không phụ thuộc vào các byte nhớ khác.

Đọc/ghi với bộ nhớ trong:

Khi đọc bộ nhớ, nội dung chứa trong ô nhớ không thay đổi. Quá trình đọc thông tin từ bộ nhớ trong diễn ra như sau:

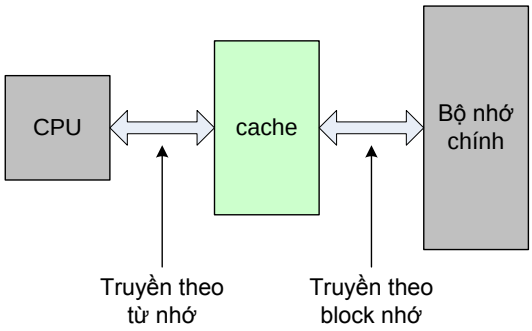
- Đầu tiên CPU gửi địa chỉ của vùng nhớ thông qua bus địa chỉ tới một mạch gọi là bộ giải mã địa chỉ.
- Tiếp theo, CPU gửi một tín hiệu điều khiển qua bus điều khiển tới kích hoạt bộ giải mã địa chỉ.
- Bộ giải mã địa chỉ mở mạch điện thực hiện chức năng sao chép dữ liệu trong vùng nhớ đưa ra bus dữ liệu, CPU ghi nhận dữ liệu vào các thanh ghi.

Quá trình ghi cũng tương tự nhưng xảy ra theo chiều ngược lại, dữ liệu đi từ CPU đến bộ nhớ. Khi ghi vào bộ nhớ thì nội dung có trong bộ nhớ đó bị xóa để lưu nội dung mới.

Do cơ chế địa chỉ hóa và do giá thành cao nên bộ nhớ trong thường có dung lượng không lớn lắm, từ vài megabyte (cache) đến vài gigabyte (RAM).

Các loại bộ nhớ trong:

Bộ nhớ cache là bộ nhớ đệm giữa CPU (chính xác là các thanh ghi trong CPU) và bộ nhớ chính (RAM), có tốc độ rất cao, cho phép CPU truy cập dữ liệu nhanh hơn từ bộ nhớ chính. Cache thường được đặt trên chip của CPU.



Hình 2.10. Cache đệm giữa CPU và bộ nhớ chính

Khi CPU cần đọc dữ liệu, nó tìm dữ liệu trong cache trước, nếu không thấy thì mới tìm trong bộ nhớ chính rồi đưa dữ liệu đó vào cache để tăng tốc độ xử lý dữ liệu trong các lệnh kế tiếp.

Cache được làm từ RAM tĩnh (SRAM, Static Random Access Memory – bộ nhớ truy cập ngẫu nhiên tĩnh), các bit được lưu trữ bằng các Flip-Flop, có cấu trúc phức tạp và giá thành cao.

RAM (Random Access Memory)

Thực chất, RAM là cách gọi tắt phổ biến của RAM động (DRAM, Dynamic RAM), các bit được lưu trữ trên tụ điện, có cấu trúc đơn giản hơn, tốc độ chậm hơn và giá thành thấp hơn so với SRAM. Khi tụ điện được tích điện, nó biểu diễn bit 1. Ngược lại, khi tụ điện xả hết sẽ biểu diễn bit 0.

Tương tự như cache, RAM là "phòng đợi" cho CPU. Nó được dùng để nạp vào hệ điều hành (đặt ở bộ nhớ ngoài, thường là đĩa cứng) khi khởi động máy tính, để chứa các lệnh chương trình ứng dụng, để lưu trữ dữ liệu tạm thời chờ được CPU đọc vào các mức bộ nhớ phía trong hoặc ghi lên các mức bộ nhớ ngoài.

Các máy tính cá nhân ngày nay thường có 2-8 GB RAM. Để đạt mức hiệu năng cơ bản tốt, máy tính cài hệ điều hành Windows 7 nên có ít nhất 1GB RAM. Các ứng dụng/trò chơi đồ họa, video để chạy tốt cần tối thiểu 2GB RAM. Lượng RAM mà máy tính cần phụ thuộc vào phần mềm được sử dụng. Dung lượng RAM yêu cầu thường được ghi trên nhãn của các gói phần mềm.

Các hệ điều hành ngày nay có khả năng phân phối rất tốt không gian RAM cho nhiều chương trình tại cùng một thời điểm. Trong trường hợp một chương trình vượt quá không gian cấp cho nó, hệ điều hành dùng một vùng trên đĩa cứng, gọi là **bộ nhớ ảo**, để chứa các phần của chương trình hoặc tệp dữ liệu đến khi chúng được cần đến. Bằng cách đổi dữ liệu trong RAM và bộ nhớ ảo, máy tính tạo ra dung lượng bộ nhớ chính gần như là không giới hạn. Tuy nhiên, vì tốc độ truy cập đĩa cứng nhỏ hơn rất nhiều so với RAM nên nếu phụ thuộc quá nhiều vào bộ nhớ ảo thì hiệu năng của máy tính sẽ bị giảm đáng kể.

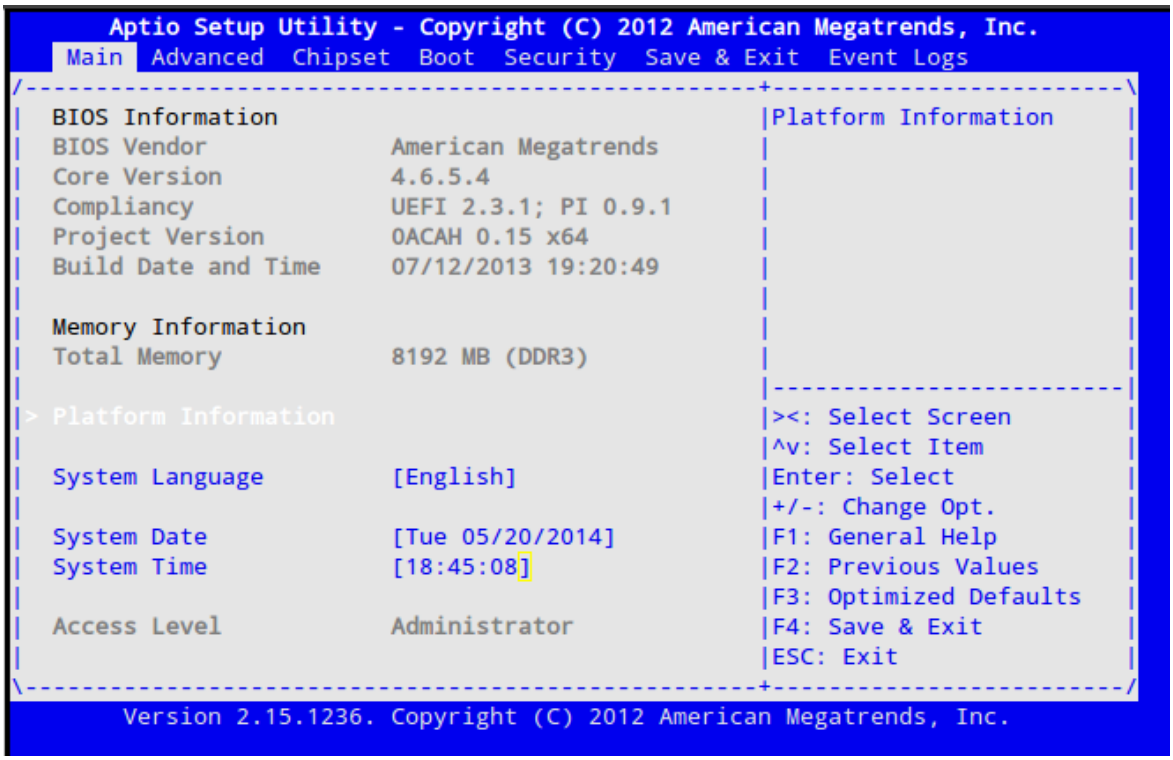
Đa số các máy tính cá nhân ngày nay sử dụng SDRAM (synchronous DRAM – RAM động làm việc được đồng bộ bởi xung đồng hồ), có tốc độ cao và tương đối rẻ. SDRAM được phân lớp tiếp thành DDR (Double Data Rate), DDR2 (2 kênh truyền dữ liệu), DDR3.

Hình 2.11. DDR3-SDRAM

ROM (Read Only Memory)

ROM là loại bộ nhớ có nội dung cố định, chỉ cho phép người dùng/máy tính đọc dữ liệu nhưng không cho phép ghi vào. Dữ liệu thường được ghi vào ROM trong lúc chế tạo, là tập các lệnh cốt lõi để khởi động máy tính như cách truy cập đĩa cứng, tìm hệ điều hành và nạp vào RAM. Tập lệnh này được gọi là BIOS (Basic Input/Output System).

EEPROM (Electrically Erasable Programmable ROM): là bộ nhớ có thể ghi chương trình theo từng byte và xóa được bằng điện. EEPROM thay thế công nghệ CMOS dùng một pin nhỏ cấp nguồn gắn trên bo mạch chủ. Loại bộ nhớ này được dùng để lưu trữ các thông tin hệ thống mà có thể bị thay đổi như thời gian, dung lượng RAM, dung lượng đĩa cứng. Khi người dùng thay đổi cấu hình của máy tính – ví dụ như lắp thêm RAM – dữ liệu trong EEPROM phải được cập nhật. Một số hệ điều hành nhận biết và thực hiện cập nhật tự động. Người dùng cũng có thể tự thay đổi thiết lập trong EEPROM bằng cách chạy chương trình cài đặt như hình 2.12.



Hình 2.12. EEPROM chứa các thiết lập cấu hình máy tính

b. Bộ nhớ ngoài

RAM chỉ dùng cho việc ghi dữ liệu khi đang xử lý, không giữ được dữ liệu khi không còn nguồn nuôi. Vì vậy, đối với các dữ liệu cần lưu giữ lâu dài, không thể để trên RAM được. Mặt khác, tuy tốc độ truy nhập trên RAM nhanh, nhưng dung lượng của nó nhỏ, không thể lưu trữ lượng thông tin lớn. Vì vậy, để có thể lưu trữ thông tin lâu dài với khối lượng lớn, ta phải sử dụng bộ nhớ ngoài.

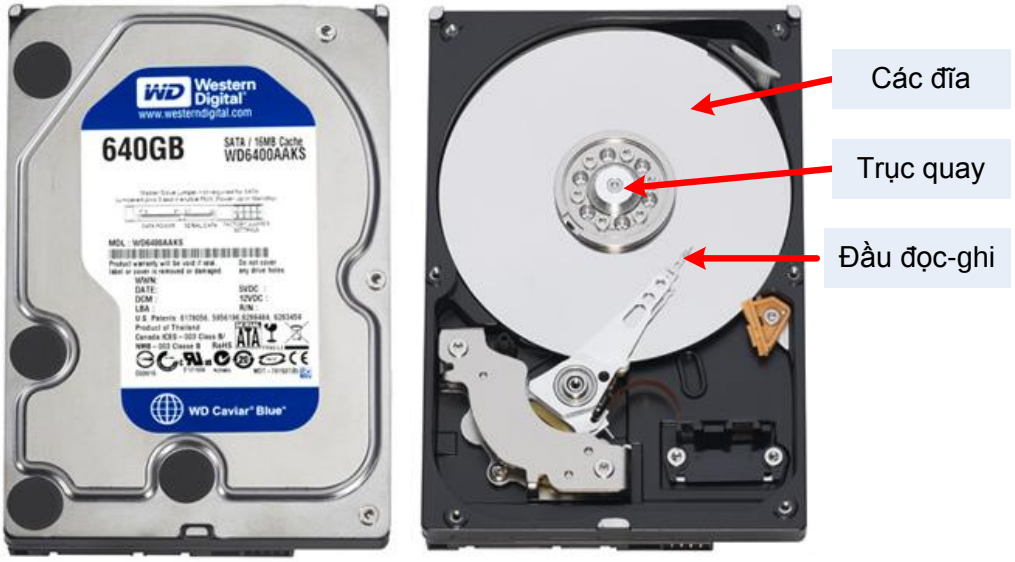
Bộ nhớ ngoài (storage devices) gồm các loại bộ nhớ mà CPU không thể truy cập trực tiếp, thông tin lưu trữ không bị xóa khi mất nguồn, có dung lượng lớn hơn bộ nhớ trong nhưng tốc độ truy cập thấp hơn. Bộ nhớ ngoài gồm các loại đĩa từ tính (đĩa cứng từ, đĩa mềm), đĩa quang (CD, DVD, Bluray), bộ nhớ flash (các loại thẻ nhớ, thanh nhớ usb, ổ cứng thể rắn).

Đặc điểm cơ bản của bộ nhớ ngoài là thông tin không được định vị bằng địa chỉ giống như bộ nhớ trong mà được tổ chức theo từng khối logic gọi là tệp (file). Do đó CPU không thể làm việc trực tiếp với dữ liệu ở bộ nhớ ngoài. Trước khi sử dụng, dữ liệu ở các file được chuyển dần vào bộ nhớ trong để CPU có thể xử lý.

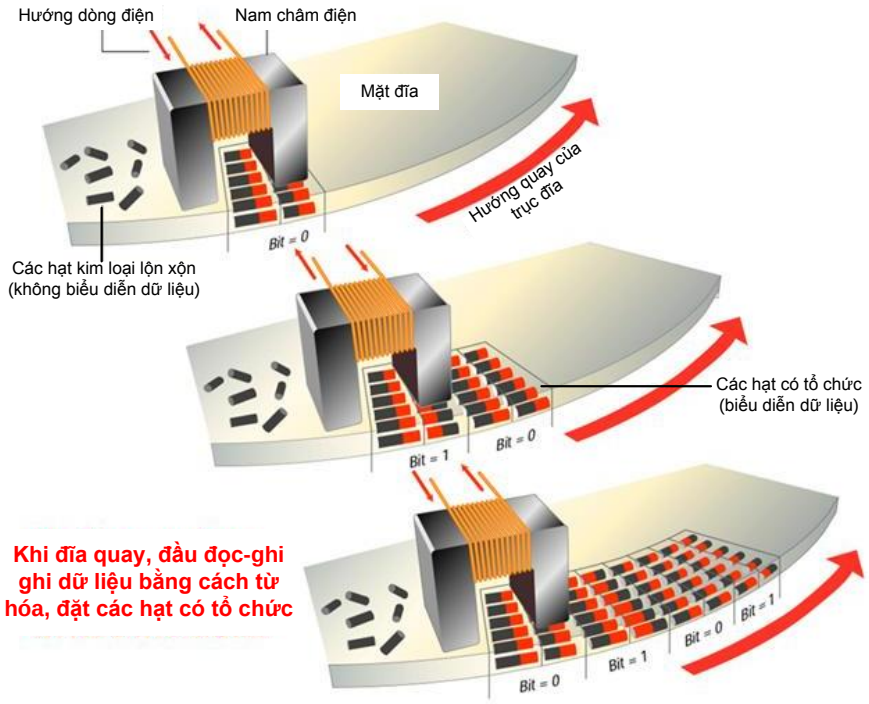
Bộ vi xử lý chỉ làm việc với các dữ liệu đã được mã hóa thành các bit 1 và 0. Khi dữ liệu được lưu trữ, các bit đó phải được chuyển thành dạng tín hiệu hay dấu hiệu nào đó lâu dài, nhưng có thể thay đổi được khi cần thiết. Dễ nhận thấy là dữ liệu không thể được ghi dạng số 0/1 theo nghĩa đen. Thay vào đó, các bit 0 và 1 phải được chuyển thành dạng nào đó thể hiện được trên bề mặt của các phương tiện lưu trữ. Có 3 công nghệ được dùng để chế tạo bộ nhớ ngoài là: từ tính, quang, thể rắn.

Đĩa cứng từ (Magnetic Hard Disk)

Đĩa cứng từ (hay gọi tắt là đĩa cứng) thường là một bộ đĩa hợp kim nhôm đường kính 3,5”, có phủ vật liệu từ tính trên mặt (hình 2.13). Các đĩa từ lưu thông tin bằng cách từ hóa các hạt rất nhỏ trên bề mặt đĩa. Các hạt đó duy trì hướng từ của chúng cho đến khi hướng bị thay đổi (hình 2.14). Vì vậy, thông tin trên đĩa từ được lưu trữ lâu dài nhưng cũng có thể thay đổi được hoặc xóa được. Tính chất này cung cấp sự linh hoạt trong việc sửa đổi dữ liệu, sử dụng lại những vùng nhớ chứa dữ liệu không cần thiết nữa.

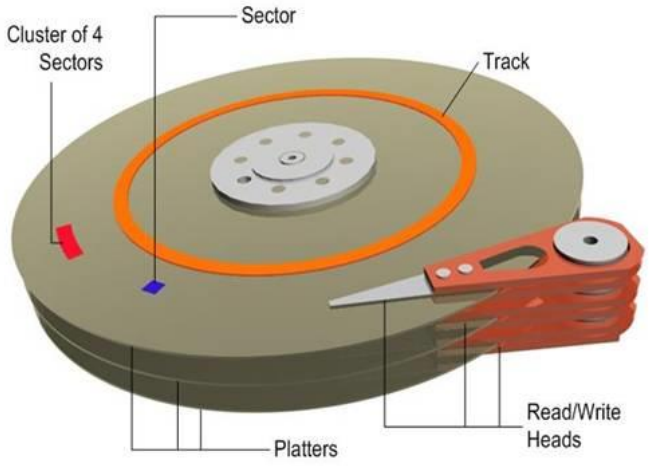


Hình 2.13. Ổ đĩa cứng từ khi nguyên trạng (trái) và khi tháo lớp vỏ bảo vệ (phải)



Hình 2.14. Từ hóa các hạt trên bề mặt đĩa từ

Đĩa cứng gồm nhiều đĩa được xếp thành chồng, đồng trục. Mỗi mặt đĩa được chia thành các đường tròn đồng tâm gọi là các đường ghi (track). Các đường ghi lại được chia thành các cung (sector). Dữ liệu được định vị trên đĩa theo địa chỉ, được xác định thông qua chỉ số của mặt đĩa, chỉ số đường ghi và chỉ số cung (hình 2.15).



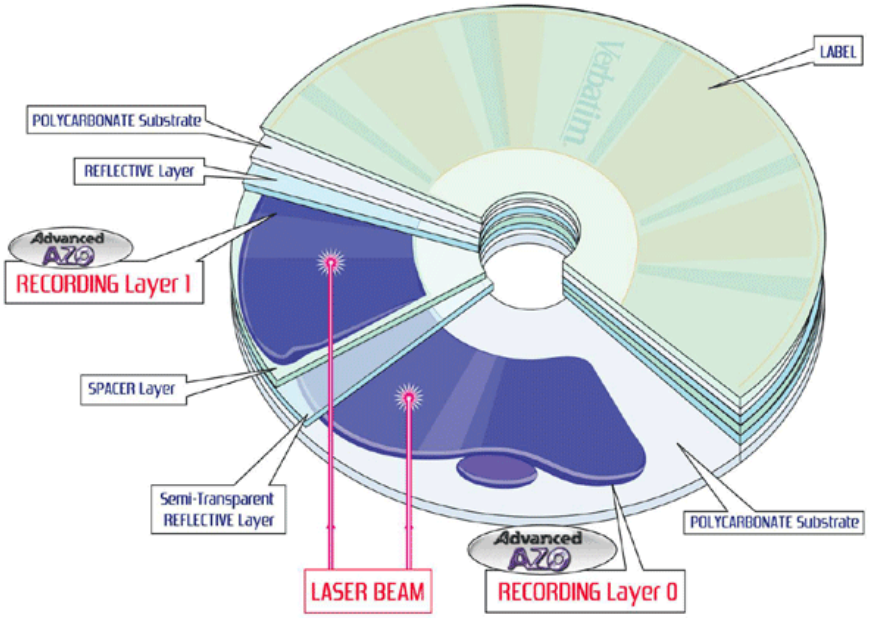
Hình 2.15. Cấu tạo đĩa cứng

Dữ liệu được đọc/ghi trên các mặt đĩa nhờ các đầu từ (còn được gọi là đầu đọc/ghi). Mỗi mặt đĩa có một đầu từ riêng. Chúng được gắn kết thành một khối và di chuyển đồng thời. Đầu từ dịch chuyển theo phương bán kính, đĩa thì quay tròn. Nhờ sự kết hợp đó, đầu từ có thể tiếp xúc với mọi vùng thông tin trên các đường ghi. Mặc dù tất cả các đầu từ đều đặt vào các mặt đĩa tương ứng nhưng đọc/ghi trên mặt đĩa nào thì đầu từ tương ứng sẽ được kích hoạt.

Một đĩa cứng hiện nay có dung lượng từ 40 GB đến 2 TB, thời gian truy cập 6-11 ms, tốc độ quay 5.400-7.200 vòng/phút, tốc độ chuyển dữ liệu trung bình khoảng 57.000 KB/s.

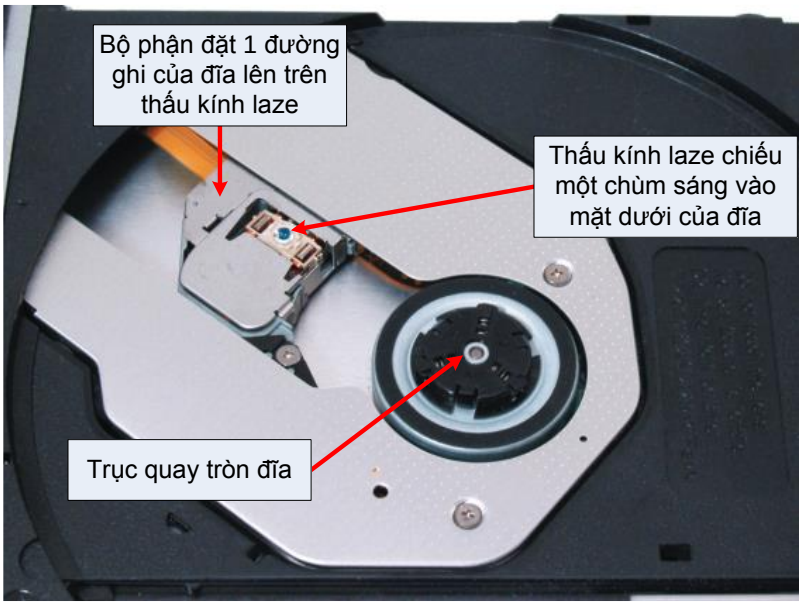
Đĩa quang

Đĩa quang gồm các loại đĩa CD (compact disc), DVD (digital video disc, hoặc digital versatile disc) và BD (bluray disc). Các đĩa quang thường có đường kính 4,75", làm bằng polycarbonate, có 1-2 lớp ghi dữ liệu, có phủ một lớp phim nhôm có tính phản xạ và một lớp bảo vệ (hình 2.16).



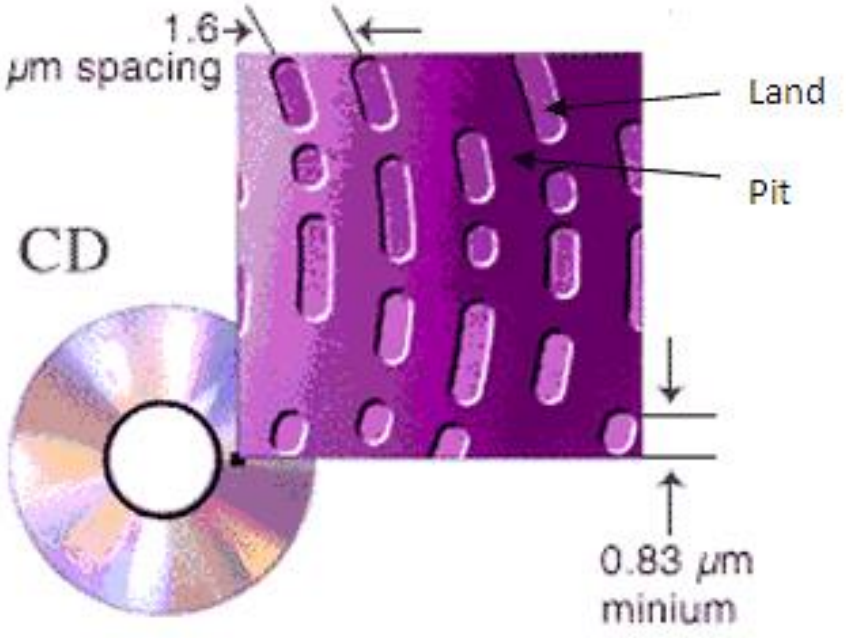
Hình 2.16. Các lớp của một đĩa quang

Ổ đĩa quang có một trục quay để quay tròn đĩa qua một đầu đọc/ghi bằng tia laze.



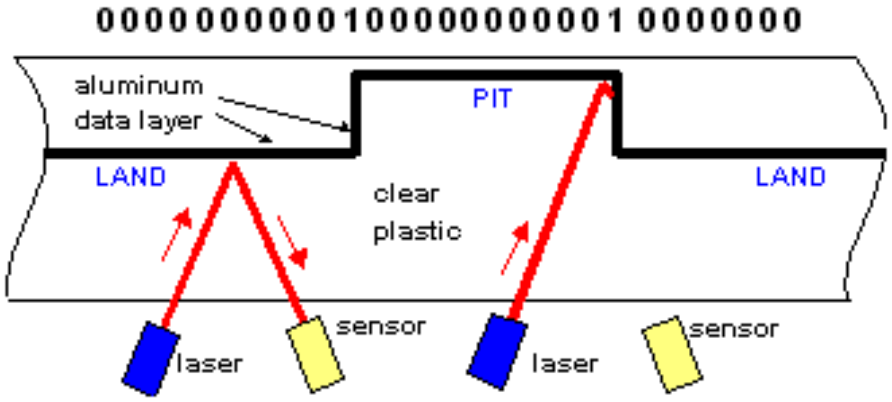
Hình 2.17. Bên trong một ổ đĩa quang

Để ghi dữ liệu lên các đĩa quang thì chúng cần được đốt bằng tia laze cường độ mạnh để tạo ra các vùng lõm (trong tiếng Anh gọi là pit) và các vùng nổi (land) trên lớp polycarbonate (hình 2.18). Việc này đòi hỏi phải có ổ đĩa có chức năng ghi.



Hình 2.18. Các vùng pitch và land trên đĩa quang

Để đọc dữ liệu từ đĩa quang thì dùng ổ đĩa với tia laze có cường độ yếu hơn. Khi đọc, đầu đọc chiếu tia laze lên đĩa và phân tích tín hiệu phản hồi để nhận biết các pit và land. Lưu ý là bản thân các pit và land không biểu diễn các bit "0" hay "1". Tại mỗi điểm chuyển đổi từ pit thành land hoặc ngược lại thì tia laze bị hấp thụ hoặc bị tán xạ, cảm biến không nhận được tia phản xạ, khi đó máy tính đọc thành bit "1". Tại các điểm khác (trên pit hoặc land) thì cảm biến nhận được tia phản xạ, máy tính sẽ đọc thành bit "0" (hình 2.19).



Hình 2.19. Nguyên tắc đọc dữ liệu trên đĩa quang

Dung lượng và tốc độ của các loại đĩa quang hiện nay được chỉ ra trong bảng sau:

Bảng 2.1. Dung lượng và tốc độ của các loại đĩa quang

Loại đĩa quang	Dung lượng/lớp	Tốc độ ghi cơ sở (1X)	Tốc độ lớn nhất hiện tại
CD	700 MB	150 KB/s	52X ~ 78000 KB/s
DVD	4,7 GB	1352.5 KB/s	24X ~ 32500 KB/s
BD	25 GB	4394.5 KB/s	16X ~ 70000 KB/s

Bộ nhớ bán dẫn dùng công nghệ flash

Loại bộ nhớ này còn được gọi là bộ nhớ thể rắn, bao gồm các loại thẻ nhớ (memory cards), thanh nhớ usb (usb flash drives, memory sticks) và ổ cứng thể rắn (SSD, solid-state drives) (hình 2.20). Các bộ nhớ này rất gọn, có thể dùng trực tiếp với máy tính hoặc với các thiết bị số cầm tay như máy ảnh, máy quay phim, điện thoại di động, máy nghe nhạc.

Hình 2.20. Các loại bộ nhớ thể rắn

Loại bộ nhớ này dùng mạng lưới các mạch bán dẫn với công nghệ flash, dữ liệu có thể bị xóa và ghi lại. Mỗi ô trên lưới có 2 transistor đóng vai trò là các cổng giữ các bit 0 và 1. Khi cổng mở, dòng điện có thể đi qua và ô đó tương ứng với bit 1. Ngược lại, khi cổng đóng, ô đó tương ứng với bit 0 (hình 2.21).

Hình 2.21. Các cổng giữ các bit 0/1 tại mỗi ô trong mạch bán dẫn

Bộ nhớ thể rắn tiêu thụ rất ít năng lượng, có tốc độ truy cập dữ liệu nhanh vì chúng không cần các thành phần chuyển động. Công nghệ này cũng rất bền vì nó không bị ảnh hưởng bởi các chấn động, từ trường hay sự thay đổi nhiệt độ bất thường. Tuy nhiên, hiện tại thì dung lượng bộ nhớ thể rắn ít hơn so với đĩa cứng từ và giá thành thì cao hơn nên ổ cứng thể rắn còn chưa phổ biến lắm.

2.3.3. Thiết bị vào/ra

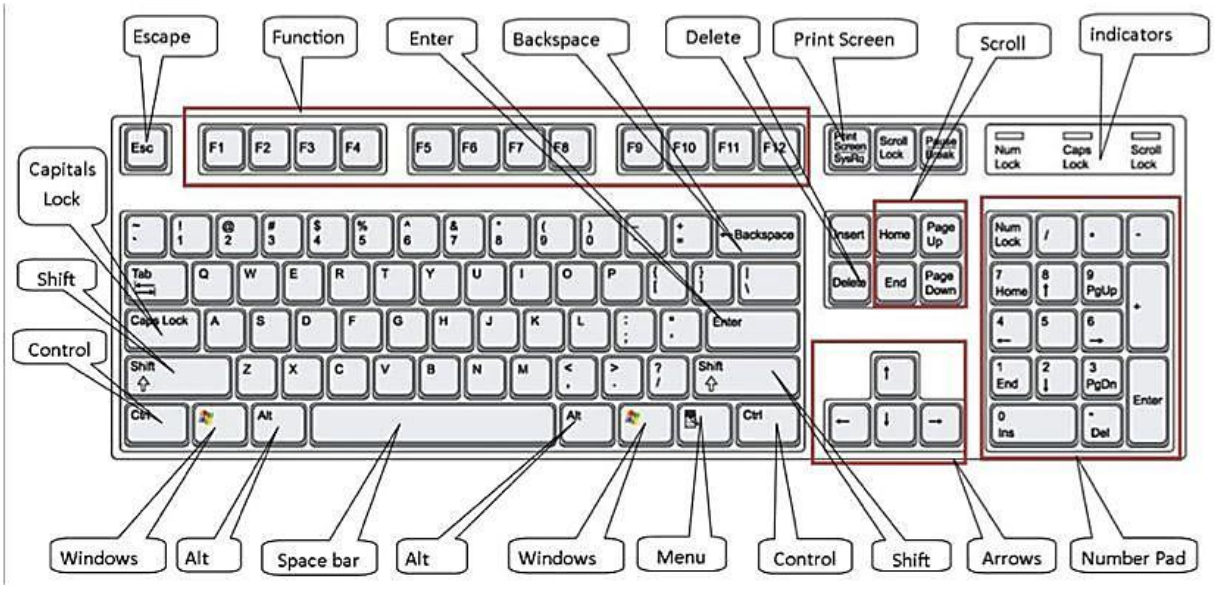
Các thiết bị vào/ra (Input/Output Devices) dùng để trao đổi dữ liệu giữa máy tính và môi trường bên ngoài. Cụ thể hơn, các thiết bị vào có chức năng chuyển dữ liệu từ bên ngoài vào bộ nhớ trong, còn các thiết bị ra dùng để chuyển thông tin từ bộ nhớ trong ra môi trường bên ngoài.

a. Thiết bị vào

Các thiết bị vào bao gồm bàn phím, con chuột, tay chơi game, máy quét ảnh, máy ảnh số, microphone, bút và màn hình cảm ứng, thiết bị đọc thẻ, đọc mã vạch... Chúng ta sẽ đề cập kỹ hơn đến 2 thiết bị vào cơ bản nhất là bàn phím và con chuột.

Bàn phím (Keyboard)

Là thiết bị dùng để đưa vào máy các lệnh điều khiển, dữ liệu. Các bàn phím thường được thiết kế tương tự như các máy đánh chữ (hình 2.22), ưu điểm là tránh sự mắc kẹt cơ khí của các phím. Thiết kế này được gọi là QWERTY (theo 6 phím chữ cái liên tục ở hàng trên bên trái).



Hình 2.22. Một bàn phím máy tính

Bàn phím có khoảng 104 phím, được chia thành 4 nhóm sau:

- Nhóm phím chữ: gồm các phím chữ cái, chữ số, các dấu.
- Nhóm phím chức năng: để thực hiện nhanh một số yêu cầu nào đó. Thường các phần mềm tự quy định những thao tác tương ứng với chúng. Bàn phím máy tính cá nhân thường để sẵn 12 phím chức năng F1, F2,… , F12.
- Nhóm phím điều khiển: xác định một số chức năng đặc biệt như thiết lập các chế độ khác nhau của bàn phím, thoát khỏi chương trình. Nhóm này gồm các phím: Esc (Escape), Caps Lock, Shift, Ctrl (Control), Alt (Alternate), Insert, Delete, Print Screen, Scroll Lock, Pause/Break, Num Lock.
- Nhóm phím điều khiển con trỏ màn hình: gồm các phím mũi tên lên, xuống, trái, phải, Home, End, Page Up, Page Down, Tab, Back space.

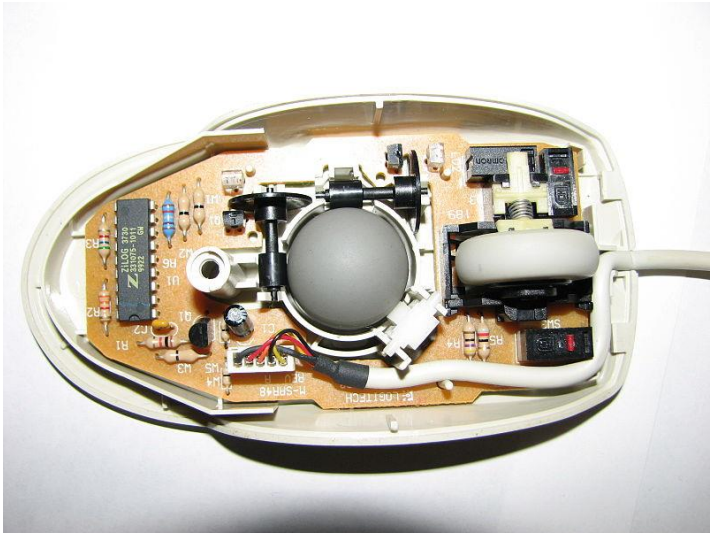
Khi ta ấn một phím, tín hiệu được truyền cho máy tính thông qua bộ lập mã, tương ứng với ký tự của phím được ấn đó.

Con chuột (Mouse)

Là thiết bị chỉ định điểm làm việc trên màn hình phổ biến nhất, hoạt động theo nguyên lý phát hiện chuyển động theo hai hướng so với bề mặt bên dưới. Chuyển động của con chuột trên bề mặt được phiên dịch thành chuyển động của một con trỏ trên màn hình giao diện đồ họa. Dạng phổ biến nhất của con chuột là gồm 2 nút bấm và 1 nút cuộn (hình 2.23). Thông thường, nút bên trái dùng cho thao tác lựa chọn, đặt vị trí của con trỏ màn hình, nút bên phải để hiện menu ngữ cảnh gồm các lệnh có thể được thực hiện với đối tượng tại vị trí con trỏ.

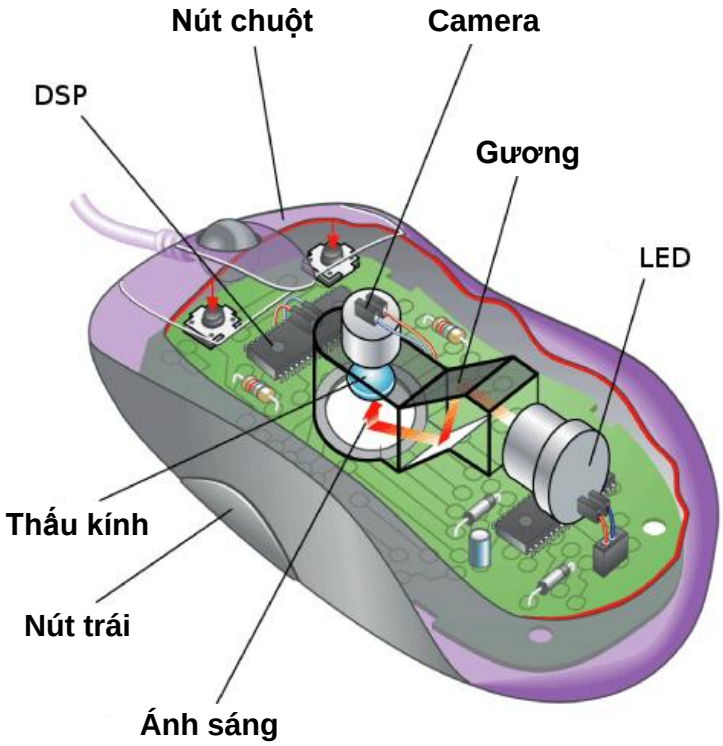
Hình 2.23. Con chuột máy tính

Chuột bi: loại này sử dụng cơ chế cơ học. Một viên bi hình cầu ở dưới con chuột, khi chuột di chuyển sẽ truyền chuyển động vào 2 trụ đặt vuông góc nhau (hình 2.24). Mỗi trụ này được gắn với một thiết bị đếm xung mà số lượng xung tỷ lệ với góc quay của nó. Các xung truyền vào trong máy tính sẽ được dùng để tính vị trí dịch chuyển của con trỏ màn hình. Loại chuột này có nhược điểm là dễ bị kẹt do bẩn.



Hình 2.24. Bên trong chuột bi

Chuột quang: loại chuột này chụp ảnh liên tiếp bề mặt bên dưới chuột (khoảng 1.000 ảnh mỗi giây), so sánh để phát hiện ra sự chuyển dịch. Chuột quang thường dùng đi-ốt phát quang hoặc phát laze hồng ngoại để chiếu sáng bề mặt bên dưới (hình 2.25). Ưu điểm của chuột quang là độ phân giải đạt được cao hơn nên cho kết quả chính xác hơn, hoạt động tốt trên nhiều loại bề mặt khác nhau (chuột laze thậm chí hoạt động trên cả bề mặt kính), không bị kẹt do bẩn giống như chuột bi.



Hình 2.25. Cấu tạo bên trong chuột quang

Chú thích: DSP (Digtal Signal Processor): Bộ xử lý tín hiệu số;

LED (Light-Emitting Diode): Đi-ốt phát quang;

Các máy tính xách tay thường có một bàn cảm ứng. Người sử dụng có thể dùng thay chuột bằng cách di ngón tay lên mặt bàn cảm ứng để điều khiển con trỏ di chuyển.

b. Thiết bị ra

Các thiết bị ra bao gồm màn hình, máy in, máy chiếu, máy vẽ, loa máy tính. Chúng ta sẽ tìm hiểu kỹ hơn về 2 loại thiết bị thông dụng nhất là màn hình và máy in.

Màn hình (Display hoặc Monitor)

Màn hình là thiết bị hiển thị chữ hay ảnh bằng cách tạo ra lưới các điểm ảnh (pixel) rất nhỏ có màu sắc khác nhau. Các yếu tố ảnh hưởng đến chất lượng hình ảnh bao gồm: kích thước màn hình, khoảng cách giữa các điểm ảnh (dot pitch), độ rộng góc nhìn, tốc độ đáp ứng, độ phân giải và độ sâu màu sắc.

Dot pitch là thước đo độ sắc nét của ảnh. Thông số này càng nhỏ thì ảnh càng sắc nét. Các màn hình ngày nay có khoảng cách giữa các điểm ảnh là khoảng 0,26-0,23mm.

Độ rộng góc nhìn được xác định bởi góc lớn nhất mà người sử dụng vẫn nhìn rõ ảnh màn hình. Độ rộng này từ 170^0 trở lên cho phép bạn nhìn màn hình từ các vị trí khác nhau mà không làm giảm chất lượng hình ảnh.

Tốc độ đáp ứng là khoảng thời gian cần thiết để một điểm ảnh thay đổi từ đen thành trắng rồi đổi lại thành đen. Màn hình có tốc độ đáp ứng nhanh hiển thị hình ảnh của các đối tượng chuyển động sắc nét với độ bóng mờ tối thiểu. Tốc độ này được đo bằng mili giây (ms). Với các máy chuyên chơi game thì tốc độ đáp ứng lý tưởng là 5ms hoặc nhỏ hơn.

Độ sâu màu sắc (hoặc độ sâu bit) là số màu mà màn hình có thể hiển thị. Các màn hình ngày nay có thể hiển thị hàng triệu màu. Khi được thiết lập độ sâu 24-bit màu, màn hình của bạn có thể hiển thị hơn 16 triệu màu (2^{24}).

Độ phân giải là thước đo khả năng thể hiện tinh tế của màn hình, được xác định bằng số lượng điểm ảnh theo chiều ngang và chiều dọc mà màn hình hiển thị. Độ phân giải chuẩn là theo tỷ lệ 4:3. Màn hình rộng thì tỷ lệ này là 16:9. Các màn hình máy tính cá nhân ngày nay thường có độ phân giải đạt chuẩn HD (high definition) là 1280 x 720 (720p), thậm chí đạt full HD là 1920 x 1080 (1080p), hoặc cao hơn. Thực ra, hai tính năng Độ sâu màu sắc và Độ phân giải không chỉ phụ thuộc vào chính màn hình mà còn phụ thuộc vào thiết bị điều khiển màn hình (video card).

Trước đây, loại màn hình phổ biến là đèn tia âm cực (đèn CRT) – là loại đèn dùng cho tivi. Các điểm ảnh được tạo bởi các súng bắn điện tử trong đèn hình có phủ các vật liệu phát quang. Loại màn hình này nặng, có độ dầy lớn, chiếm nhiều diện tích.

Ngày nay, chúng ta đang dùng phổ biến các loại màn hình mỏng, nhẹ, dùng công nghệ tinh thể lỏng (LCD – liquid crystal display) hoặc plasma. Màn hình tinh thể lỏng hiển thị hình ảnh bằng cách lọc ánh sáng qua một lớp các ô tinh thể lỏng. Nguồn sáng có thể dùng đèn huỳnh quang catốt lạnh (CCFL) hoặc dùng các điốt phát sáng (LED).

Hình 2.26. Màn hình CRT (trái) và LCD (phải)

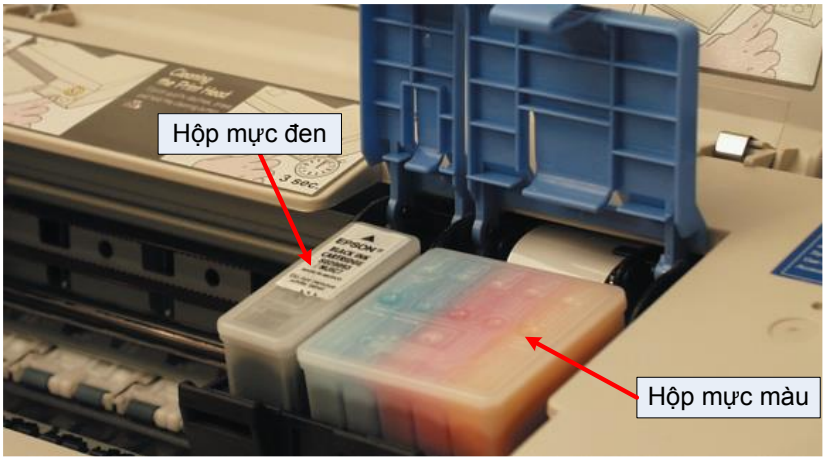
Máy in (Printer)

Máy in là thiết bị cho phép in chữ hay ảnh ra giấy. Có ba loại máy in gồm: máy in kim, máy in phun và máy in laze.

Máy in kim (Dot matrix printer) là loại ra đời đầu tiên, cùng với sự ra đời của máy tính cá nhân. Loại này sử dụng một bộ các kim in bố trí dạng ma trận, ảnh hay chữ được tạo bằng các chấm do kim in đập vào băng mực làm băng mực in lên giấy, mỗi chữ được thể hiện qua một tổ hợp các điểm tách ra từ ma trận điểm (hình 2.27). Mặc dù chất lượng hình ảnh in không mịn như các loại máy in khác nhưng có những công việc cần in những bản in nhiều liên bắt buộc phải in theo nguyên tắc va đập (như in hóa đơn) nên loại máy in này vẫn khá phổ biến ở các quầy thanh toán và trong các ngân hàng.

Hình 2.27. Bên trong máy in kim

Máy in phun (ink jet printer, hình 2.28) tạo các điểm trên giấy bằng cách phun tia mực siêu nhỏ. Công nghệ phổ biến nhất là dùng tinh thể áp điện để làm bơm mực. Một tinh thể áp điện sẽ co hay giãn tùy thuộc vào điệp áp đặt vào hai mặt đối diện của tinh thể. Một nguyên lý khác cũng được dùng là đầu in có các ống phun mực nhỏ li ti. Khi ống bị nóng thì mực bị sôi tạo thành bong bóng siêu nhỏ bắn vào giấy. Đa số các máy in phun dùng hệ màu CMYK – chỉ dùng 4 màu xanh cẩm thạch, đỏ tím vàng, đen (Cyan, Magenta, Yellow, Black) phối trộn với nhau để tạo thành hàng nghìn màu cho các bản in. Loại máy in này có ưu điểm là chất lượng bản in tốt, nhưng tốn nhiều mực và giá hộp mực khá cao.



Hình 2.28. Bên trong máy in phun màu

Máy in laze (laser printer, hình 2.29) là loại máy in dùng kỹ thuật laze để tạo từng trang ảnh bằng các hạt mực siêu nhỏ trên một trống tĩnh điện. Khi trống áp vào giấy in thì những hạt mực sẽ dính trở lại giấy và được nung nóng chảy ra thấm vào giấy. Vì công nghệ laze phức tạp

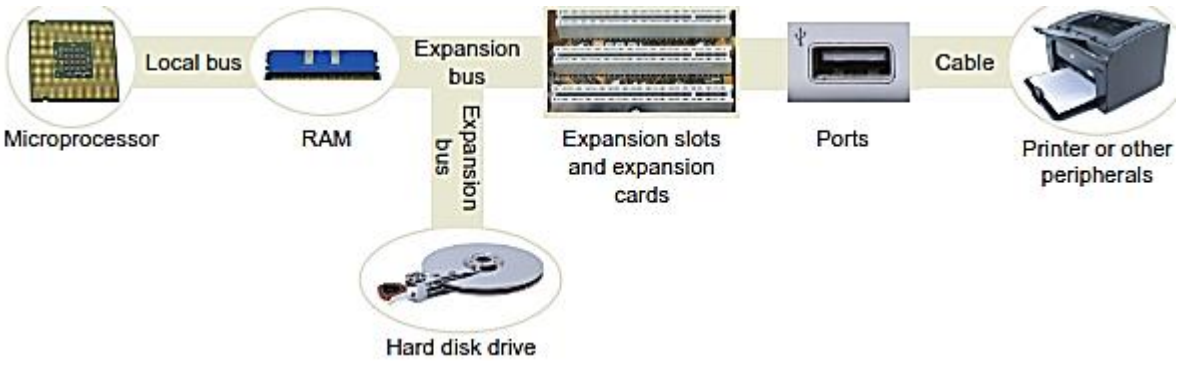
hơn công nghệ phun mực nên giá máy in laze cao hơn. Ưu điểm của loại máy in này là chất lượng ảnh rất cao, tốn ít mực hơn nên được dùng rất rộng rãi.

Hình 2.29. Máy in laze

Ngoài các thiết bị vào-ra kể trên thì có một số thiết bị vừa là thiết bị vào vừa là thiết bị ra như màn hình cảm ứng, modem, ổ đọc và ghi đĩa.

2.3.4. Liên kết hệ thống

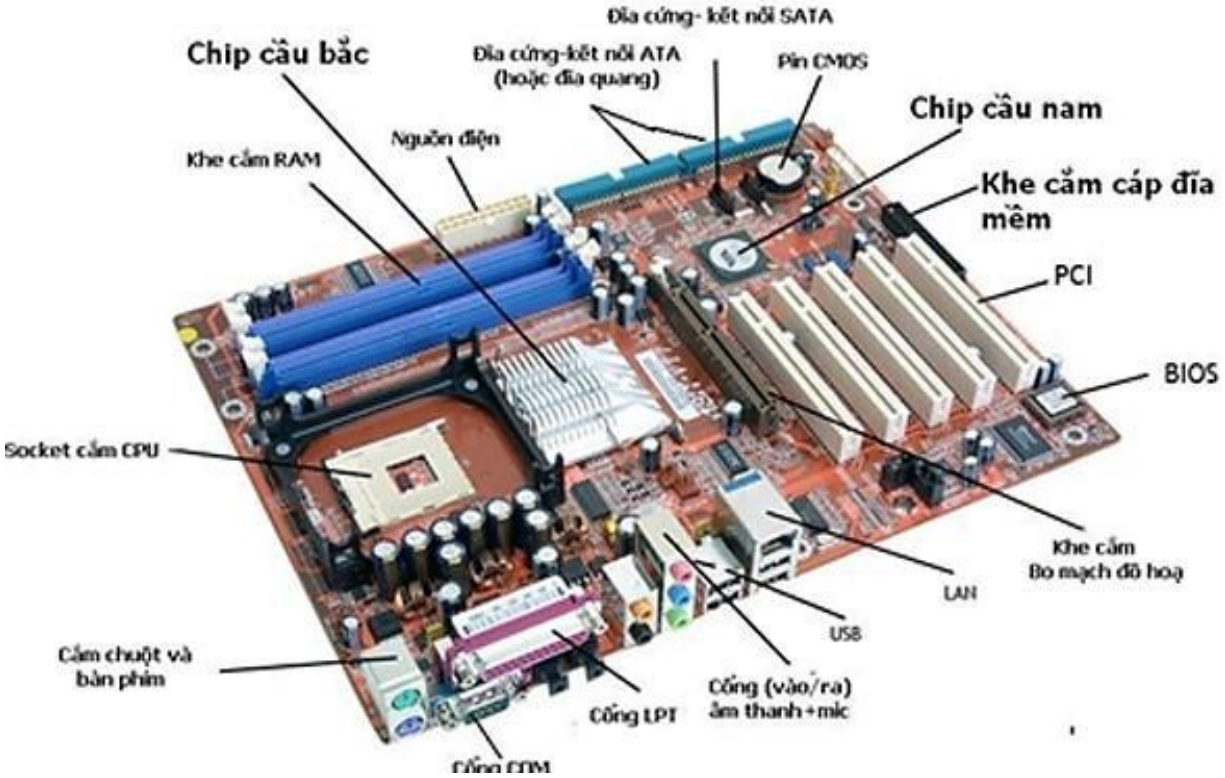
Các thiết bị máy tính được liên kết với nhau thông qua các đường bus, các khe cắm mở rộng hoặc các loại cổng kết nối (hình 2.30). Các thành phần này thường được thiết kế trên một bo mạch chủ (hình 2.31).



Hình 2.30. Liên kết các thành phần hệ thống

Bus là các tuyến đường để thông tin (dữ liệu, lệnh, địa chỉ) chạy trên đó. Chúng có thể là những đường mạch trên bo mạch chủ (ví dụ nối giữa CPU và RAM) hoặc các loại cáp mở rộng (ví dụ cáp nối ổ đĩa cứng với bo mạch chủ).

Các khe cắm mở rộng (expansion slot) được dùng để cắm các loại card điều khiển thiết bị vào-ra như card đồ họa, card âm thanh, modem.



Hình 2.31. Các thành phần kết nối hệ thống trên bo mạch chủ

Các cổng (port) gồm nhiều loại, được dùng để kết nối máy tính với các thiết bị vào-ra. Một số loại cổng phổ biến là: PS/2 kết nối chuột và bàn phím, VGA kết nối màn hình, LPT kết nối máy in, RJ45 kết nối modem, USB kết nối rất nhiều thiết bị giao tiếp qua chuẩn USB, các cổng âm thanh, cổng đọc thẻ nhớ...

CÂU HỎI VÀ BÀI TẬP

1. Nêu chức năng của Bộ xử lý trung tâm (CPU). Những yếu tố nào ảnh hưởng đến hiệu năng của CPU?
2. Cho biết sự khác nhau cơ bản về cách thức lưu trữ dữ liệu của bộ nhớ trong và bộ nhớ ngoài?
3. Kể tên và nêu những đặc điểm cấu tạo, lưu trữ của các loại bộ nhớ trong? So sánh tốc độ truy nhập dữ liệu của các bộ nhớ đó.
4. Kể tên và nêu những đặc điểm cấu tạo, lưu trữ của các loại bộ nhớ ngoài?
5. Cho biết tên và đặc điểm cấu tạo và hoạt động của một số loại thiết bị vào điển hình?
6. Cho biết tên và đặc điểm cấu tạo và hoạt động của một số loại thiết bị ra điển hình?
7. Nêu chức năng của thành phần liên kết hệ thống. Nêu tên một số cổng vào ra thông dụng và cho biết chúng có thể kết nối với những thiết bị ngoại vi nào?

Chương 3

PHẦN MỀM MÁY TÍNH VÀ HỆ ĐIỀU HÀNH

Máy tính không thể hoạt động nếu không được cài đặt hệ điều hành và các phần mềm ứng dụng. Chương này giới thiệu về các vấn đề cơ bản của hệ điều hành và phần mềm máy tính.

Phần 3.1 trình bày khái niệm phần mềm, quy trình sản xuất phần mềm và phân loại phần mềm. Phần 3.2 trình bày về lịch sử hình thành và phát triển của hệ điều hành, vai trò và hoạt động của hệ điều hành trên máy tính. Bên cạnh đó cũng giới thiệu về một số loại hệ điều hành điển hình cài đặt trên máy tính cá nhân và các thiết bị di động. Phần cuối cùng của chương giới thiệu các vấn đề liên quan tới tệp và thư mục.

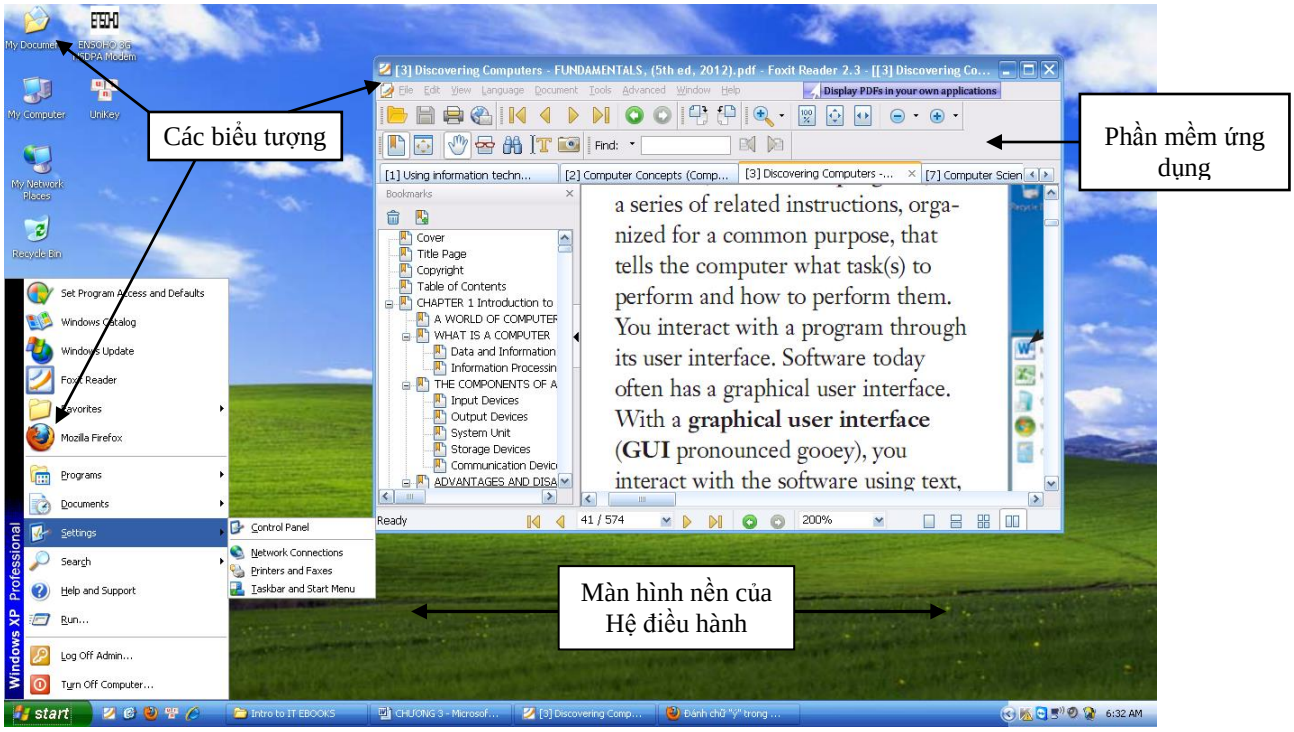
3.1. PHẦN MỀM MÁY TÍNH

3.1.1. Khái niệm về phần mềm

Phần mềm, hay còn gọi là chương trình, là một là một tập hợp những câu lệnh hoặc chỉ thị (Instruction) được viết bằng một hoặc nhiều ngôn ngữ lập trình theo một trật tự xác định, kết hợp với các dữ liệu hay tài liệu liên quan nhằm tự động thực hiện một số nhiệm vụ hay chức năng hoặc giải quyết một vấn đề cụ thể nào đó.

Phần mềm thực hiện các chức năng của nó bằng cách gửi các chỉ thị trực tiếp đến phần cứng máy tính (Computer hardware) hoặc bằng cách cung cấp dữ liệu để phục vụ các chương trình hay phần mềm khác.

Thông thường, người dùng có thể tương tác với phần mềm thông qua một giao diện. Ngày nay giao diện phần mềm thường là giao diện đồ họa, người dùng có thể tương tác với phần mềm thông qua các đoạn văn bản, hình ảnh hoặc các biểu tượng.



Hình 3.1. Giao diện đồ họa của phần mềm và hệ điều hành Windows XP

3.1.2. Phân loại phần mềm

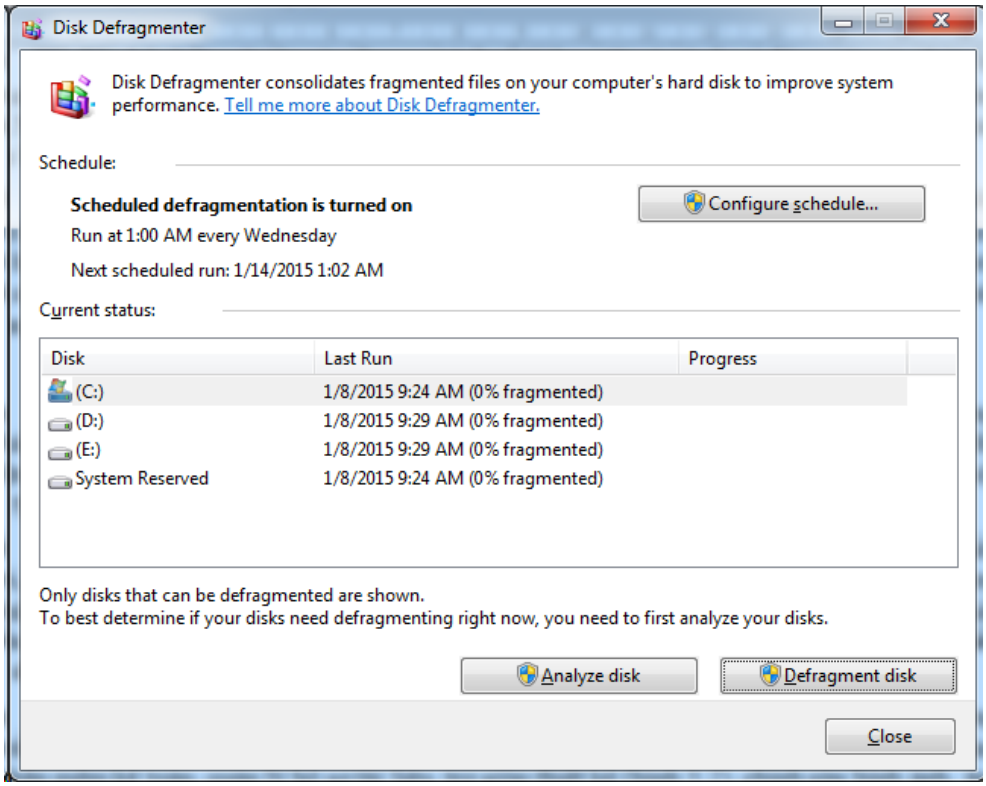
Có hai loại phần mềm cơ bản: phần mềm hệ thống và phần mềm ứng dụng, hình 3.1 là một minh họa về giao diện hệ điều hành và phần mềm ứng dụng soạn thảo văn bản.

a) Phần mềm hệ thống

Phần mềm hệ thống là các chương trình điều khiển hoặc duy trì các hoạt động của máy tính và các thiết bị liên quan. Phần mềm hệ thống hỗ trợ giao tiếp giữa người dùng, phần mềm ứng dụng và phần cứng máy tính. Có 2 kiểu phần mềm hệ thống: hệ điều hành và các chương trình tiện ích.

Hệ điều hành là một tập các chương trình phối hợp tất cả các hoạt động của các thiết bị phần cứng. Nó là một phương tiện cho người dùng để giao tiếp với máy tính và các phần mềm khác. Những dòng hệ điều hành phổ biến nhất hiện nay là Microsoft Windows, một trong số các phiên bản là Windows XP được thể hiện trong hình 3.1, hay Mac OS, hệ điều hành của Apple.

Chương trình tiện ích cho phép người dùng thực hiện các công việc liên quan tới việc bảo trì máy tính, các thiết bị và các chương trình được cài đặt trong máy. Hầu hết các hệ điều hành bao gồm nhiều chương trình tiện ích như: quản lý ổ đĩa, máy in và các thiết bị khác. Người dùng cũng có thể mua các chương trình tiện ích cho phép thực hiện chức năng quản lý máy tính bổ sung.



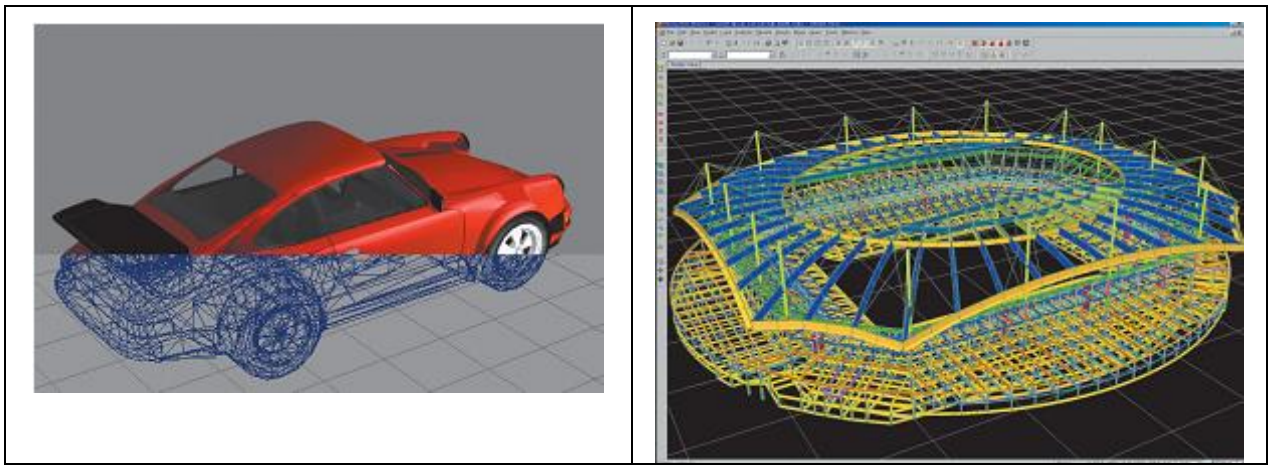
Hình 3.2. Chương trình tiện ích Disk Defragmenter giúp chống phân mảnh ổ cứng

b) Phần mềm ứng dụng

Phần mềm ứng dụng có thể là các chương trình được thiết kế giúp người dùng sử dụng một các hiệu quả hơn và/hoặc hỗ trợ các công việc cá nhân, ví dụ như phần mềm thiết kế một thí nghiệm, phần mềm điều khiển một dây chuyền sản xuất, phần mềm quản lý khách hàng của một công ty... Những phần mềm như vậy còn được gọi là *phần mềm đặt hàng*.

Có những phần mềm ứng dụng được thiết kế dựa trên những yêu cầu chung của nhiều người, không theo yêu cầu đặt hàng của riêng ai. Chúng được viết rất hoàn chỉnh và thường kèm theo những phương tiện để cài đặt lên máy một cách tự động. Người mua chỉ cần mua về, thiết lập các chế độ làm việc thích hợp là có thể sử dụng được. Những phần mềm như thế gọi là *phần mềm đóng gói*.

Phần mềm đóng gói cũng có nhiều loại khác nhau giúp người dùng thực hiện các công việc rất đa dạng. Ví dụ như các phần mềm quản lý thông tin cá nhân, nhắc việc, quản lý dự án, các phần mềm kế toán, quản lý hồ sơ tài liệu, trợ giúp thiết kế (hình 3.3), chỉnh sửa hình ảnh, âm thanh, video và các phần mềm đa phương tiện khác nhau. Bên cạnh đó cũng có những loại phần mềm giúp người dùng có thể tạo các trang Web cá nhân một cách đơn giản, các phần mềm quản lý tài chính cá nhân, pháp lý, thuế, các phần mềm mang tính giáo dục, các hệ thống tài liệu tham khảo và giải trí (ví dụ phần mềm trò chơi và mô phỏng)...



Hình 3.3. Các phần mềm hỗ trợ thiết kế

Các phần mềm đóng gói thường có sẵn tại các cửa hàng bán các sản phẩm máy tính hoặc có thể tải trực tuyến trên nhiều trang Web khác nhau.

Phần mềm phát triển ứng dụng là các phần mềm để tạo ra các phần mềm khác. Đây là các phần mềm mà các chuyên gia tin học thường sử dụng để phát triển phần mềm. Đối với những người làm việc trong lĩnh vực tin học thì phần mềm ứng dụng dành cho người dùng cuối, là sản phẩm và là mục tiêu của họ. Để hỗ trợ cho việc làm ra các sản phẩm phần mềm, họ lại dùng chính các phần mềm khác gọi là phần mềm hỗ trợ phát triển.

Ngày nay các thiết bị điện tử dân dụng trở nên thông minh hơn nhờ công nghệ vi xử lý. Các phần mềm điều khiển thiết bị được ghi trong ROM. Tivi, ô tô, điện thoại di động, lò vi sóng... đều sử dụng các hệ vi xử lý. Phần mềm được ghi vào trong ROM và dùng trong các hệ vi xử lý gắn liền với các thiết bị gọi là *phần mềm nhúng*. Ngày nay, phần mềm nhúng chiếm một tỷ trọng rất lớn trong thị trường phần mềm nói chung.

Việc phân loại phần mềm cũng mang tính chất tương đối, mỗi tài liệu có thể sẽ có quan niệm và cách phân loại khác nhau. Sự phân loại nói trên chỉ mang tính tương đối nhằm cung cấp một bức tranh tổng thể về các lớp phần mềm dựa trên mục đích và phương thức sử dụng. Ranh giới giữa chúng khá mờ, thậm chí còn xâm lấn vào nhau. Ví dụ phần mềm tiện ích cũng có thể được coi là một kiểu phần mềm ứng dụng.

3.1.3. Quy trình phát triển phần mềm

Phát triển phần mềm bao gồm bốn hoạt động cơ bản: đặc tả, phát triển (thiết kế và thực thi), kiểm thử và cài đặt bảo trì phần mềm. Các hoạt động này được tổ chức một cách khác nhau tùy theo tiến trình phát triển được lựa chọn.

Ở mô hình thác nước, các hoạt động này được tổ chức một cách tuần tự, trong khi đó, với mô hình phát triển tiến hóa, các hoạt động này lại được tổ chức một cách xen kẽ. Các hoạt động này được thực hiện như thế nào phụ thuộc vào kiểu phần mềm, tình hình nhân sự và cấu trúc của tổ chức.

a) Đặc tả phần mềm

Đặc tả phần mềm là một tiến trình để hiểu và xác định những dịch vụ nào cần có trong hệ thống cũng như xác định những ràng buộc đối với việc phát triển và chức năng của hệ thống. Đây là một trong những giai đoạn rất quan trọng trong tiến trình phần mềm, vì mỗi lỗi trong giai đoạn này sẽ không tránh khỏi việc dẫn đến những sai lầm trong các giai đoạn tiếp theo.

Tiến trình này sẽ sinh ra các tài liệu yêu cầu, đó là các bản đặc tả hệ thống. Các yêu cầu thường được trình bày ở hai mức độ chi tiết khác nhau trong tài liệu. Khách hàng và người dùng cuối cần những mô tả yêu cầu ở mức cao, còn người phát triển hệ thống lại cần những đặc tả hệ thống chi tiết hơn.

b) Thiết kế và thực thi phần mềm

Giai đoạn này liên quan tới việc chuyển những yêu cầu phần mềm thành những hệ thống có thể thực thi được. Thông thường nó liên quan đến việc thiết kế và lập trình.

Thiết kế phần mềm là việc mô tả cấu trúc của phần mềm được thực thi, dữ liệu của hệ thống, giao diện giao tiếp giữa các thành phần và đôi khi, đó là thuật toán sẽ được sử dụng. Người làm thiết kế cũng không phải ngay lập tức đưa ra được một bản thiết kế hoàn chỉnh, mà thông thường nó cũng phải được chỉnh sửa lặp đi lặp lại nhiều lần.

Tiến trình thiết kế có thể liên quan tới việc phát triển một vài mô hình của hệ thống ở những mức độ trừu tượng khác nhau. Khi thiết kế được phân tích, lỗi và các điểm thiếu sót ở các giai đoạn trước sẽ được bộc lộ, điều này cũng giúp cho việc chỉnh sửa những bản thiết kế trước đó.

Thực thi phần mềm là giai đoạn các lập trình viên dùng các ngôn ngữ lập trình để viết lệnh (mã nguồn) thực sự để tạo ra hệ thống dựa trên các bản đặc tả thiết kế chi tiết. Để đảm bảo chương trình được xây dựng thỏa mãn mọi yêu cầu trong bản đặc tả thiết kế, các kỹ sư lập trình cũng đồng thời tiến hành các thử nghiệm phần chương trình do mình tạo ra. Phần thử nghiệm trong giai đoạn này được gọi là kiểm thử đơn vị. Người viết mã nguồn chạy thử chương trình của mình với dữ liệu giả định để xem chương trình có cho ra các kết quả mong đợi. Giai đoạn này còn được gọi là kiểm thử hộp trắng. Bên cạnh đó, người ta cũng có thể tiến hành thử nghiệm đơn vị độc lập. Công việc này do một thành viên khác trong nhóm đảm nhiệm để đảm bảo tính "độc lập", được tiến hành dựa trên kế hoạch kiểm thử do người viết mã nguồn soạn ra.

c) Kiểm thử phần mềm

Kiểm thử phần mềm là quá trình vận hành chương trình để tìm ra lỗi, mục tiêu của người làm kiểm thử là thiết kế các trường hợp kiểm thử để có thể phát hiện một cách có hệ thống các loại lỗi khác nhau với chi phí thời gian và công sức ít nhất có thể.

Tuy nhiên, để đảm bảo phần mềm phát triển đúng theo đặc tả và đáp ứng tốt các yêu cầu người dùng thì ngoài hoạt động kiểm thử phần mềm, trong suốt tiến trình phát triển phần mềm người ta cần phải tiến hành các hoạt động xác minh và thẩm định phần mềm.

Xác minh là sự kiểm tra xem sản phẩm có đúng với đặc tả hay không, tức là chú trọng vào việc phát hiện lỗi của phần mềm qua từng giai đoạn phát triển.

Thẩm định là kiểm tra xem sản phẩn có đáp ứng được yêu cầu người dùng hay không, tức là chú trọng vào việc phát hiện sự khác biệt của sản phẩm làm ra với những gì mà người dùng mong đợi.

Xác minh và thẩm định tĩnh là việc kiểm tra phần mềm mà không thực hiện chương trình. (xét duyệt yêu cầu, xét duyệt thiết kế, thanh tra mã nguồn, sử dụng các biến đổi hình thức để kiểm tra tính đúng của chương trình).

Xác minh và thẩm định động là việc kiểm tra thông qua việc thực hiện chương trình, được tiến hành sau khi đã xây dựng được chương trình (mã nguồn). Đây là kỹ thuật kiểm tra được áp dụng phổ biến nhất, còn được gọi là kiểm thử phần mềm.

d) Cài đặt và bảo trì phần mềm

Trong giai đoạn này, hệ thống vừa phát triển sẽ được *cài đặt và triển khai* sao cho người dùng có thể sử dụng được. Trước khi người dùng thật sự bắt tay vào sử dụng hệ thống, nhóm phát triển cần tạo các tập dữ liệu cần thiết và huấn luyện cho người dùng sử dụng chương trình để đảm bảo hệ thống được sử dụng hữu hiệu nhất.

Bảo trì phần mềm là việc điều chỉnh các lỗi chưa được phát hiện trong các giai đoạn trước của chu kỳ sống của một phần mềm, nâng cấp tính năng sử dụng và an toàn vận hành của phần mềm. Bảo trì phần mềm có thể chiếm đến 65%-75% công sức trong chu kỳ sống của một phần mềm. Nhiệm vụ của giai đoạn bảo trì phần mềm là giữ cho phần mềm được cập nhật khi môi trường thay đổi và yêu cầu người sử dụng thay đổi.

3.1.4. Phần mềm mã nguồn đóng và mã nguồn mở

a) Phần mềm mã nguồn đóng

Phần mềm mã nguồn đóng là phần mềm mà mã nguồn không được công bố. Muốn sử dụng hợp pháp phần mềm nguồn đóng, người dùng cần được sự cho phép của người giữ bản quyền phần mềm, thường là những tổ chức hay cá nhân phát triển phần mềm đó. Phần mềm nguồn đóng thường là có phí, tuy nhiên một số phiên bản giản lược chức năng có thể là miễn phí. Ví dụ về phần mềm mã nguồn đóng có phí là các hệ điều hành Microsoft Windows, bộ phần mềm ứng dụng văn phòng Microsoft Office, phần mềm gõ tiếng Việt Vietkey, hệ quản trị cơ sở dữ liệu SQL Server, môi trường phát triển phần mềm Microsoft Visual Studio. Phần mềm nghe nhạc JetAudio là một ví dụ về phần mềm nguồn đóng miễn phí phiên bản giản lược (jetAudio Basic) và có phí phiên bản đầy đủ (jetAudio Plus VX). Trình duyệt web Google Chrome là ví dụ cho phần mềm mã nguồn đóng miễn phí.

Ưu thế của mã nguồn đóng như cái tên của nó, có thể che đậy toàn bộ công nghệ đằng sau giúp tăng cường bảo mật và giấu được công nghệ độc quyền. Hiển nhiên mã nguồn đóng luôn được phép học hỏi công nghệ mới nhất của mã nguồn mở nhưng điều ngược lại là khó có thể, từ yếu tố này dễ dàng nhận thấy mã nguồn đóng luôn sở hữu công nghệ tương tự mới nhất nếu không muốn nói là hơn so với mã nguồn mở. Ngoài ra việc độc quyền tạo điều kiện kiếm siêu lợi nhuận và một phần trong đó được quay trở lại đầu tư cho công nghệ đằng sau mỗi sản phẩm phát triển, chu trình này như một vòng khép kín giúp mã nguồn đóng đứng vững.

Hầu hết các công ty phát triển mã nguồn đóng đều cố gắng dùng các công nghệ mới nhất để hạn chế người dùng thay đổi hoặc sao chép phần mềm một cách bất hợp pháp. Ngay cả với phần mềm miễn phí mã nguồn cũng không được công khai vì tác giả không muốn người khác thay đổi mã nguồn của mình cũng như lợi dụng nó để biến thành sản phẩm của riêng mình (giống như một hình thức đạo văn).

b) Phần mềm mã nguồn mở

Phần mềm mã nguồn mở (open-source software) hiểu theo nghĩa rộng là một khái niệm chung được sử dụng cho tất cả các phần mềm mà mã nguồn của nó được công bố rộng rãi, công khai và cho phép mọi người tiếp tục phát triển phần mềm đó. Những phần mềm nguồn mở ngày nay rất phát triển cả về chất lượng và số lượng. Hệ điều hành Linux, trình duyệt web Mozilla Firefox, bộ phần mềm ứng dụng văn phòng Open Office, phần mềm gõ tiếng Việt Unikey, phần mềm máy chủ web Apache, hệ quản trị cơ sở dữ liệu MySQL, ngôn ngữ lập trình Perl là những ví dụ điển hình của phần mềm mã nguồn mở.

Mã nguồn mở không có nghĩa là chúng có thể được sao chép, sửa chữa thoải mái hay sử dụng vào mục đích nào cũng được. Mã nguồn mở được công bố dưới rất nhiều điều kiện khác nhau, một số trong đó cho phép phát triển, sử dụng và bán tùy ý miễn là giữ nguyên các dòng về nguồn gốc sản phẩm, một số bắt buộc tất cả các sản phẩm làm ra từ đó cũng phải là open-source, một số khác đòi hỏi phải công bố trọn vẹn mã nguồn, một số khác không cho phép sử dụng vào mục đích thương mại, một số khác lại không có ràng buộc gì đáng kể.

Qua đó ta thấy khái niệm *open source* không thể chuẩn xác mà muốn nói đến tính pháp lý của việc sử dụng các phần mềm mã nguồn mở, chúng ta phải xem xét đến điều kiện sử dụng đã được công bố. Một điều kiện hay được áp dụng nhất là GPL: GNU General Public License (http://www.fsf.org/licenses/gpl.html) của tổ chức Free Software Foundation.

GPL có 2 đặc điểm phân biệt, đó là:

- Tác giả gốc giữ bản quyền về phần mềm nhưng cho phép người dùng rất nhiều quyền khác, trong đó có quyền tìm hiểu, phát triển, công bố cũng như quyền khai thác thương mại sản phẩm.

- Tác giả sử dụng luật bản quyền để bảo đảm các quyền đó không bao giờ bị vi phạm đối với tất cả mọi người, trên mọi phần mềm có sử dụng mã nguồn của mình.

Đặc biệt điểm thứ hai thường được gọi là hiệu ứng virus (viral effect) vì nó biến tất cả các phần mềm có dùng mã nguồn GPL cũng biến thành phần mềm GPL. Trên thực tế điều này có ý nghĩa: bất kỳ tác giả nào sử dụng dù chỉ một phần rất nhỏ mã nguồn GPL trong chương trình của mình cũng phải công bố chương trình đó dưới điều kiện GPL. Điều kiện này quy định: mọi phần mềm GPL đều phải công bố mã nguồn của mình rộng rãi công khai và phải tạo điều kiện cho mọi người truy cập được mã nguồn ấy (ví dụ qua web hoặc qua việc bán CD giá rẻ); giữ nguyên mọi dòng chú thích về nguồn gốc tác giả, bản quyền của họ cũng như điều kiện được áp dụng đối với phần mềm (trong một tệp tin có tên LICENSE); cấm việc bán mã nguồn nhưng cho phép kinh doanh chương trình được tạo ra từ mã nguồn ấy hoặc là các dịch vụ hỗ trợ liên quan.

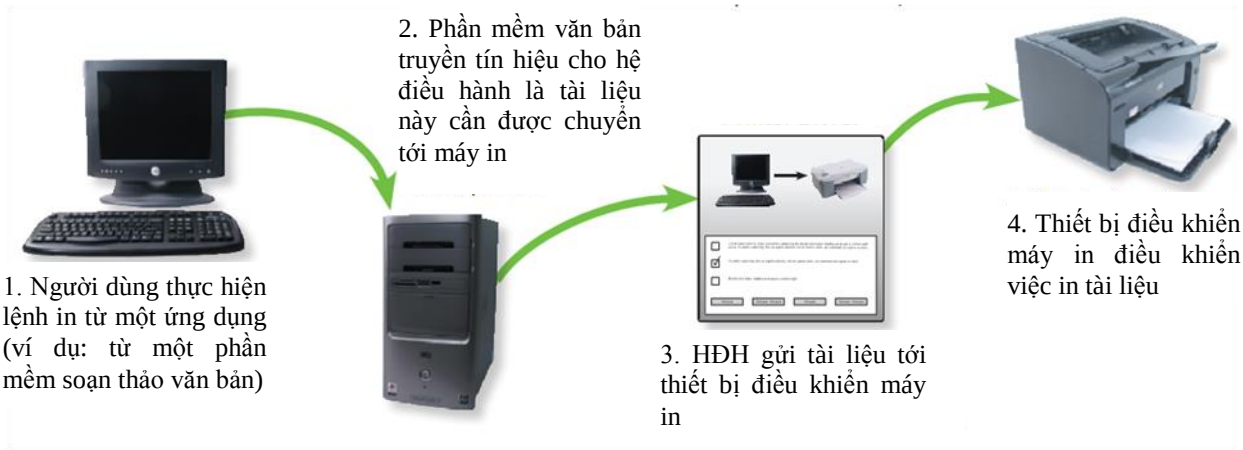
3.2. HỆ ĐIỀU HÀNH

3.2.1. Khái niệm hệ điều hành

Hệ điều hành (Operating system) là hệ thống các chương trình máy tính điều khiển, quản lý, phân phối việc sử dụng tài nguyên của máy tính và giao tiếp với người sử dụng. Như vậy, hệ

điều hành là hệ thống đứng giữa con người và máy tính, giúp con người thực hiện công việc xử lý của họ (thông qua chương trình ứng dụng) một cách hiệu quả. Hoạt động của máy tính không thể tách rời khỏi hệ điều hành. Hệ điều hành có các chức năng chính sau:

- Quản lý và điều phối các thiết bị của máy để phục vụ cho công việc xử lý. Hình 3.4 minh họa cách thức người dùng tiến hành in một văn bản dưới sự điều khiển của hệ điều hành.



Hình 3.4. Tiến trình in một văn bản từ máy tính

- Quản lý thông tin bộ nhớ ngoài: Các thông tin ở bộ nhớ ngoài được tổ chức thành các đơn vị lưu trữ gọi là *file* (tệp tin). Phân hệ thực hiện chức năng quản lý thông tin bộ nhớ ngoài gọi là phân hệ quản lý file (file management system). Các ứng dụng như tìm file, khai thác thông tin trên file, lưu trữ file vào bộ nhớ ngoài... đều phải thông qua hệ quản lý file. Chức năng này giải phóng các chương trình ứng dụng ra khỏi một công việc rất phức tạp và tỉ mỉ.

- Quản lý các tiến trình (process management). Về cơ bản, mỗi tiến trình là một chương trình đang thực hiện trên máy tính. Ngoài các chương trình của người dùng còn có các tiến trình hệ thống như quản lý thông tin giao tiếp với các thiết bị ngoại vi, điều phối tài nguyên... Thực chất quản lý tiến trình là lập lịch thực hiện các tiến trình phù hợp với yêu cầu tài nguyên của mỗi tiến trình.

- Cung cấp môi trường giao tiếp với người sử dụng và cung cấp các tiện ích cơ bản. Mỗi hệ điều hành thường cung cấp một ngôn ngữ giao tiếp với người sử dụng. Trước đây, ngôn ngữ giao tiếp thường là các lệnh mà người sử dụng phải gõ trực tiếp từ bàn phím. Ngày nay, các hệ điều hành thường cung cấp môi trường đồ họa để người dùng giao tiếp với máy, theo đó người dùng sử dụng chuột để chỉ ra các công việc thể hiện qua các biểu tượng chứ không phải gõ các lệnh.

Hệ điều hành phải được khởi động ngay trước khi máy tính làm việc với các chương trình khác. Hệ điều hành phải luôn thường trực cho tới khi máy ngừng hoạt động.

3.2.2. Lịch sử phát triển và phân loại hệ điều hành

a) Thế hệ 1 (1945 – 1955)

Vào khoảng giữa thập niên 1940, Howard Aiken ở Havard và John von Neumann ở Princeton, đã thành công trong việc xây dựng máy tính dùng ống chân không. Những máy này rất lớn với hơn 10.000 ống chân không nhưng chậm hơn nhiều so với máy rẻ nhất ngày nay.

Mỗi máy được một nhóm thực hiện tất cả từ thiết kế, xây dựng lập trình, thao tác đến quản lý. Lập trình bằng ngôn ngữ máy, thường là bằng cách dùng bảng điều khiển để thực hiện

các chức năng cơ bản. Ngôn ngữ lập trình chưa được biết đến và hệ điều hành cũng chưa nghe đến.

Vào đầu thập niên 1950, phiếu đục lỗ ra đời và có thể viết chương trình trên phiếu thay cho dùng bảng điều khiển. Không có sự phân biệt giữa chương trình của người sử dụng và chương trình điều khiển.

b) Thế hệ 2 (1955 – 1965)

Sự ra đời của thiết bị bán dẫn vào giữa thập niên 1950 làm thay đổi bức tranh tổng thể. Máy tính trở nên đủ tin cậy hơn, nó được sản xuất và cung cấp cho các khách hàng. Lần đầu tiên có sự phân chia rõ ràng giữa người thiết kế, người xây dựng, người vận hành, người lập trình và người bảo trì.

Để thực hiện một công việc (một chương trình hay một tập hợp các chương trình), lập trình viên trước hết viết chương trình trên giấy (bằng hợp ngữ hay FORTRAN) sau đó đục lỗ trên phiếu và cuối cùng đưa phiếu vào máy. Sau khi thực hiện xong nó sẽ xuất kết quả ra máy in.

Hệ thống xử lý theo lô ra đời, nó lưu các yêu cầu cần thực hiện lên băng từ và hệ thống sẽ đọc và thi hành lần lượt. Sau đó, nó sẽ ghi kết quả lên băng từ xuất và cuối cùng người sử dụng sẽ đem băng từ xuất đi in.

Hệ thống xử lý theo lô hoạt động dưới sự điều khiển của một chương trình đặc biệt là tiền thân của hệ điều hành sau này. Ngôn ngữ lập trình sử dụng trong giai đoạn này chủ yếu là FORTRAN và hợp ngữ.

c) Thế hệ 3 (1965 – 1980)

Trong giai đoạn này, máy tính được sử dụng rộng rãi trong khoa học cũng như trong thương mại. Máy IBM 360 là máy tính đầu tiên sử dụng mạch tích hợp, từ đó kích thước và giá cả của các hệ thống máy giảm đáng kể và máy tính càng phổ biến hơn. Các thiết bị ngoại vi dành cho máy tính xuất hiện ngày càng nhiều và thao tác điều khiển bắt đầu phức tạp.

Hệ điều hành ra đời nhằm điều phối, kiểm soát hoạt động và giải quyết các yêu cầu tranh chấp thiết bị. Chương trình hệ điều hành dài cả triệu dòng hợp ngữ và do hàng ngàn lập trình viên thực hiện.

Một cải tiến quan trọng của thế hệ máy tính thứ 3 là *chế độ đa chương trình* (multi-program). Trong thế hệ máy tính thứ 2, khi một chương trình đang sử dụng thiết bị xuất nhập dữ liệu thì CPU phải chờ công việc xuất nhập đó kết thúc mới tiếp tục công việc xử lý. Chế độ đa chương trình nhằm song song hóa các thiết bị ngoại vi để tận dụng thời gian CPU. Trong chế độ này, nhiều chương trình cùng được nạp vào trong bộ nhớ. Trong khi chương trình này sử dụng CPU thì chương trình thứ hai có thể đọc dữ liệu từ đĩa, còn chương trình thứ ba có thể sử dụng máy in... Vai trò của hệ điều hành trong chế độ xử lý đa chương trình là điều phối các tài nguyên của hệ thống một cách hợp lý phục vụ cho các chương trình để giảm thời gian "chết" của các thiết bị. Việc phân phối bộ nhớ cho các chương trình cùng được nạp vào bộ nhớ cũng là một việc điều phối "tài nguyên".

Trong các hệ điều hành đa chương trình, các chương trình được xử lý có một thứ tự ưu tiên nào đó, thông thường là thứ tự chương trình được nạp vào máy. Một số hệ điều hành tự động xếp các chương trình có dự báo thời gian chạy ít lên trước để giảm thời gian chờ đợi tổng thể. Như vậy, các chương trình không bắt buộc phải cùng tiến triển, điều đó dẫn đến tình trạng có các chương trình hầu như không được phục vụ cho đến khi một số chương trình có độ ưu tiên cao hơn đã thực hiện xong. Trong nhiều trường hợp, chế độ này tỏ ra khá bất tiện.

Một phát minh quan trọng khác lần đầu tiên được đưa vào trong hệ điều hành của các máy tính thế hệ thứ 3 là *cơ chế phân chia thời gian* (time sharing). Chế độ này thường áp dụng với các máy tính có nhiều người sử dụng thông qua các trạm cuối. Mỗi trạm cuối thường có một màn hình và một bàn phím, người sử dụng giao tiếp trực tiếp với máy tính qua hệ điều hành. Tốc độ làm việc của mỗi người đều rất chậm so với khả năng của máy. Trong chế độ phân chia thời gian, hệ điều hành lần lượt cấp CPU cho mỗi người sử dụng một khoảng thời gian nhất định.

Hệ điều hành phân chia thời gian là một kiểu phát triển cao hơn của hệ điều hành đa chương trình. Điểm mới là việc xử lý các chương trình là luôn phiên chứ không đợi xử lý xong chương trình này mới chuyển đến chương trình khác. Tính năng này giúp cho nhiều nhiệm vụ có thể thực thi được tại cùng một thời điểm. Chính vì lý do này mà hệ điều hành có cơ chế phân chia thời gian còn gọn là *hệ điều hành đa nhiệm* (multi-task).

Bên cạnh đó, có một số phiên bản của hệ điều hành còn cho phép nhiều người đồng thời sử dụng hệ thống, người ta gọi đó là các *hệ điều hành nhiều người sử dụng*. Trong chế độ này, khi nhiều người cùng dùng một phần mềm thì máy tính phải nạp phần mềm đó vào máy tính nhiều bản chạy động thời, mỗi bản chỉ phục vụ một người dùng vì nếu không thì người này đóng ứng dụng mà hoàn toàn không biết người kia vẫn đang dùng nó.

Trong thời kỳ các thế hệ máy tính thế hệ 3 người ta cũng đưa vào hệ điều hành một chế độ vận hành gọi là *bộ nhớ ảo*. Để máy tính có thể chạy được các chương trình có kích thước lớn hơn bộ nhớ trong, người ta tổ chức một cơ chế phân trang. Mỗi trang bộ nhớ đều có độ dài như nhau, một chương trình khi làm việc được hệ điều hành cấp cho một số trang trong bộ nhớ. Như vậy, chỉ một phần chương trình đang chạy có mặt trong bộ nhớ trong, phần còn lại của chương trình vẫn nằm ở đĩa từ. Khi nào cần sử dụng đến các lệnh đang nằm ở đĩa thì hệ điều hành tổ chức hoán chuyển phần đang nằm trong bộ nhớ trong lên đĩa và nạp phần chương trình cần thiết từ đĩa vào các trang dành cho chương trình.

d) Thế hệ 4 (1980 - nay)

Phù hợp với hai khuynh hướng xây dựng các máy tính nhỏ - máy tính cá nhân và xây dựng các máy tính lớn thì hệ điều hành cũng được phát triển theo hai khuynh hướng khác nhau.

Với máy vi tính, hệ điều hành có những hướng phát triển ưu tiên mới, đó là tính năng thân thiện vì đối tượng sử dụng thường là những người không chuyên nghiệp. Một số hệ điều hành nổi tiếng trong thời kỳ này là DOS, WINDOWS trên các dòng máy PC, MAC OS trên các dòng máy Macintosh.

Đối với các dòng máy tính lớn như mini và mainframe thì hệ điều hành chủ đạo vẫn là UNIX, được kế thừa từ MULTICS. UNIX được thiết kế là một hệ điều hành đa nhiệm và nhiều người dùng với cơ chế phân chia thời gian. Do quan niệm có nhiều người dùng nên nó có cơ chế kiểm soát thẩm quyền nghiêm ngặt để đảm bảo an toàn cho mỗi chương trình cùng chạy trên máy tính.

Hệ thống tệp tin của UNIX cũng được phân cấp theo một cây thư mục có các thuộc tính để kiểm soát thẩm quyền: quyền đọc, quyền sửa, quyền thực hiện cho bản thân người tạo ra tệp tin, nhóm người sử dụng và cho những người khác. UNIX cung cấp nhiều tiện ích dưới dạng các lệnh, bao gồm các lệnh thao tác với tệp tin và thư mục, các phương tiện để lọc, các phương tiện để lập trình, các hệ soạn thảo văn bản, các lệnh để quản trị hệ thống.

Ngày nay hầu hết các máy tính lớn đều sử dụng UNIX, SOLARIS, SUN, AIX của IBM, alpha UNIX của DEC, HP Unix của Hewlette Parkard đều là các phiên bản khác nhau của hệ điều hành UNIX. UNIX cũng được dùng trên các máy tính cá nhân như các phiên bản SOLARIS

trên PC, SCO UNIX và gần đây là LINUX, một hệ điều hành nguồn mở, tạo điều kiện để những người quan tâm có thể cùng tham gia và phát triển UNIX. Ngoài giao diện truyền thống của UNIX, người ta cũng sử dụng một giao diện đồ họa là X-WINDOWS.

Do khuôn khổ của một giáo trình tin học đại cương, ở đây chúng tôi không giới thiệu thêm về các hệ điều hành khác như *hệ điều hành song song*, hay trong môi trường mạng máy tính như *hệ điều hành mạng, hệ điều hành phân tán.*

3.2.3. Một số hệ điều hành điển hình

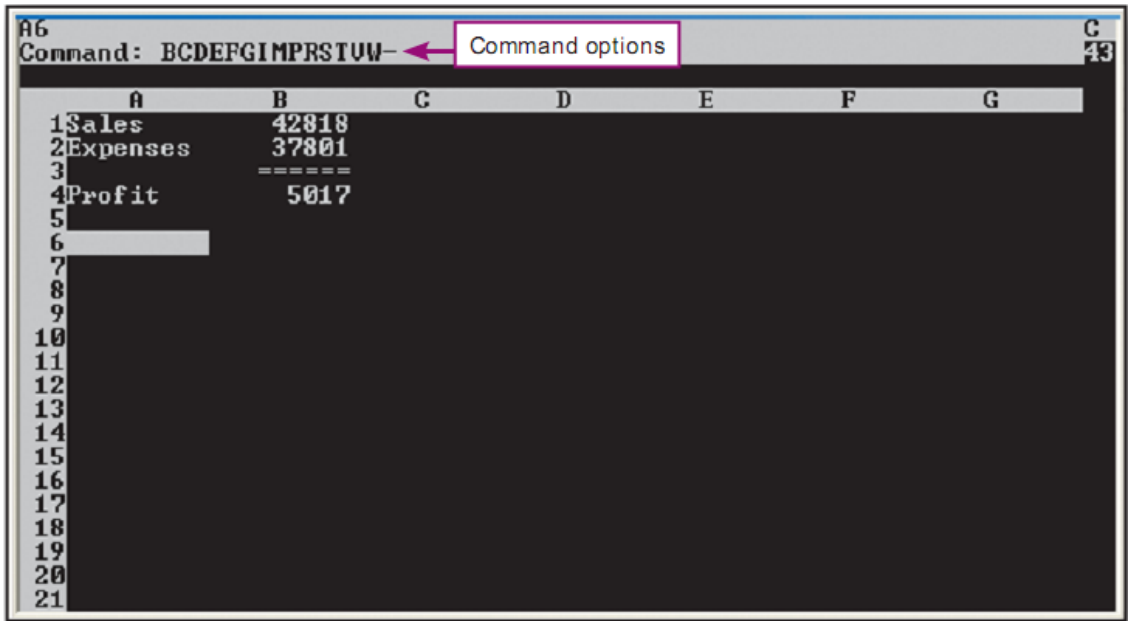
a) Hệ điều hành cho máy tính

Hệ điều hành MS-DOS

Hệ điều hành hướng đĩa MS-DOS (Microsoft Disk Operating System) là hệ điều hành của hãng phần mềm Microsoft. Đây là một hệ điều hành có giao diện dòng lệnh (command-line interface) được thiết kế cho các máy tính cá nhân (PC - Personal Computer). MS-DOS đã từng rất phổ biến trong suốt thập niên 1980 và đầu thập niên 1990, cho đến khi Windows 95 ra đời.

Phiên bản DOS đầu tiên ra đời vào tháng 8 năm 1981, với tên chính thức là PC DOS 1.0. Tên gọi MS-DOS chỉ được biết đến kể từ tháng 5 năm 1982 (MS-DOS 1.25). Sau đó, Microsoft lần lượt cho ra đời các phiên bản tiếp theo của MS-DOS song song cùng với PC-DOS.

MS-DOS 5.0 ra đời vào tháng 6 năm 1991 bao gồm nhiều tính năng mới như quản lý bộ nhớ (MEMMAKER.EXE), trình soạn thảo văn bản (MS-DOS Editor), ngôn ngữ lập trình QBASIC đã trở nên phổ biến một thời trước khi MS-DOS 6.22 ra đời vào tháng 6 năm 1994. MS-DOS 6.22 cũng là phiên bản DOS cuối cùng được chạy như một hệ điều hành độc lập. Sau khi Windows 95 ra đời vào năm 1995, các phiên bản MS-DOS tiếp theo đều được phát hành đi kèm với Windows, chẳng hạn như MS-DOS 7.0 (8/1995) là nền để cho Windows 95 khởi động và MS-DOS 8.0 đi kèm với Windows ME. Đây cũng là phiên bản cuối cùng của hệ điều hành này.



Hình 3.5. Giao diện làm việc của hệ điều hành MS-DOS

MS-DOS là hệ điều hành đơn nhiệm. Tại mỗi thời điểm chỉ thực hiện một thao tác duy nhất. Nói một cách khác, MS-DOS chỉ cho phép chạy một ứng dụng duy nhất tại mỗi thời điểm.

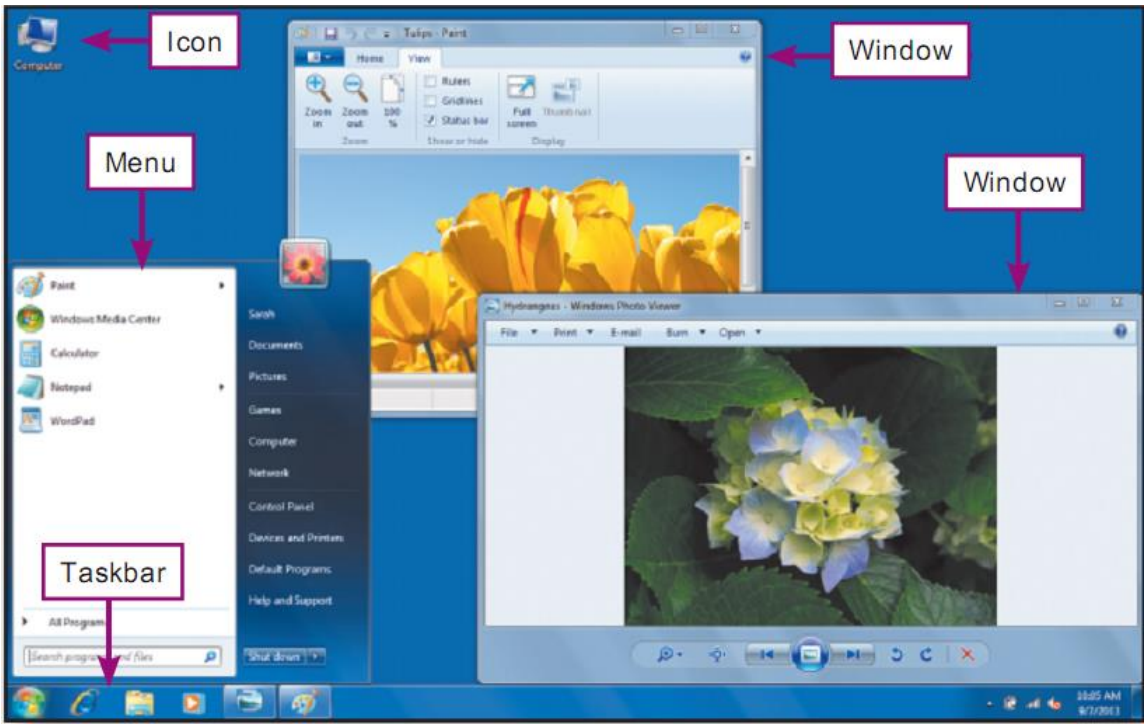
Điều này khác hẳn với Windows, vốn là một hệ điều hành đa nhiệm - người dùng có thể thi hành nhiều ứng dụng cùng một lúc. Mặc dù vậy, về sau người ta đã thiết kế một số ứng dụng chạy thường trú cho MS-DOS. Các ứng dụng này có thể chạy trên nền của các ứng dụng khác, khiến người sử dụng có thể thực hiện nhiều thao tác cùng lúc. Ở Việt Nam, phần mềm chạy thường trú trên MS-DOS rất phổ biến là chương trình hỗ trợ gõ tiếng Việt, VietRes.

Một số môi trường làm việc đa nhiệm như Deskmate hay Desqview đã được thiết kế để chạy trên DOS. Những phiên bản Windows đầu tiên cũng đều phải khởi động từ dấu nhắc DOS. Tuy nhiên, ngày nay, MS-DOS đã trở nên ít phổ biến hơn. Nó chỉ còn tồn tại trong các phiên bản Windows sau này (2000, XP) dưới dạng một ứng dụng cho phép người dùng kích hoạt chế độ dòng lệnh và thường chỉ được dùng để thực hiện những tác vụ liên quan mật thiết đến hệ thống mà giao diện đồ họa của Windows không làm được.

Hệ điều hành Microsoft Windows

Microsoft Windows là tên của các dòng phần mềm hệ điều hành độc quyền của hãng Microsoft. Lần đầu tiên Microsoft giới thiệu một môi trường điều hành mang tên Windows (Cửa sổ) là vào tháng 11 năm 1985 với những tính năng thêm vào Hệ điều hành MS-DOS giao diện đồ hoạ - đang được sự quan tâm cao vào thời điểm đó, đồng thời để cạnh tranh với hãngApple Computer.

Windows khởi đầu được phát triển cho những máy tính tương thích với IBM (dựa vào kiến trúc x86 của Intel) và ngày nay hầu hết mọi phiên bản của Windows đều được tạo ra cho kiến trúc này (tuy nhiên Windows NT đã được viết như là một hệ thống xuyên cấu trúc cho bộ xử lý Intel và MIPS) và sau này đã xuất hiện trên các cấu trúc PowerPC và DEC Alpha. Sự phổ biến của Windows đã khiến bộ xử lý của Intel trở nên phổ biến hơn và ngược lại.



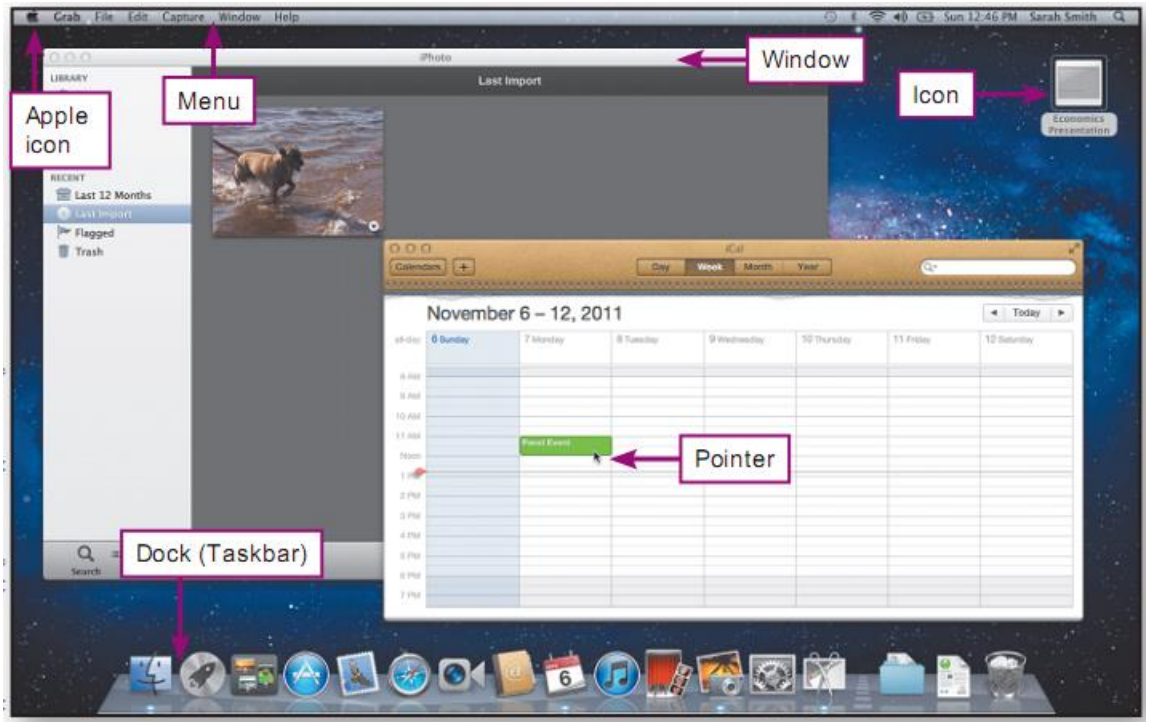
Hình 3.6. Giao diện của hệ điều hành Windows

Từ đó đến nay, Microsoft Windows dần dần chiếm ưu thế trong thị trường máy tính cá nhân trên toàn thế giới, hiện nay nó được cài đặt trên khoảng 80% số lượng máy tính trên thế giới. Hệ điều hành Windows được đặt tên xuất phát từ các khu vực làm việc hình chữ nhật xuất

hiện trên màn hình máy tính. Mỗi cửa sổ làm việc có thể hiển thị một tài liệu hoặc chương trình khác nhau, cung cấp một mô hình trực quan thể hiện khả năng đa nhiệm của hệ điều hành (hình 3.6).

Hệ điều hành MAC OS

Mac OS là tên viết tắt của hệ điều hành Macintosh và nó là hệ điều hành được thiết kế riêng cho dòng máy tính Macintosh của hãng Apple Computer. Mặc dù hệ điều hành Mac đã được phát triển vài năm trước khi Windows ra đời, cả hai hệ điều hành đều có nhiều khu vực làm việc hình chữ nhật để phản ánh khả năng xử lý đa nhiệm. Cả Windows và Mac OS đều cung cấp các dịch vụ mạng cơ bản. Tính năng độc đáo của máy tính để bàn Mac là chúng bao gồm các biểu tượng của Apple, Dock và một thanh trình đơn ứng dụng cố định ở phía trên cùng của màn hình. Hình 3.7 minh hoạ một số tính năng cơ bản của hệ điều hành dùng cho máy tính để bàn Mac.



Hình 3.7. Giao diện làm việc của hệ điều hành Mac OS

Mac OS là một hệ điều hành có nhiều nét tương đồng với Microsoft Windows. Tuy nhiên, nó có một số điểm vượt trội so với hệ điều hành Windows như: độc lập về độ phân giải, có phần mềm nghe nhìn Quick Time, có hệ thống sao lưu dữ liệu Time Machine, tiện ích tìm kiếm dữ liệu trong máy tính cùng hơn 200 tiện ích mới. Là một trong những hệ điều hành mang tính cách mạng và bảo mật cao nhất từ trước đến nay, không cần phần mềm bảo vệ riêng, ít bị virus tấn công trong khi đó Windows thì bị hàng loạt lỗi bảo mật nguy hiểm.

Nhược điểm lớn nhất của Mac OS là đòi hỏi sự tương đồng cả về phần cứng và các phần mềm ứng dụng. Trong những năm 90, Apple đã từng cho phép các hãng sản xuất khác trong liên minh PowerPC sử dụng hệ điều hành Mac OS, nhưng việc này đã chấm dứt vào khoảng năm 1998 và Mac OS trở về mã nguồn đóng.

Việc cài đặt Mac OS X lên các máy tính không do Apple sản xuất bắt đầu được chú ý vào khoảng năm 2005, khi Apple chuyển sang dùng bộ xử lý Intel cho máy tính của họ và Mac OS X bắt đầu bộc lộ những ưu điểm so với Windows. Hiện nay, để làm việc này có 2 cách, một là cài

Mac OS X vào máy ảo trên nền Windows (sử dụng một vài phần mềm máy ảo như VMWare, VirtualBox), hai là cài song song (dual-boot) với Windows. Cả 2 cách đều không thể dùng đĩa cài đặt gốc Mac OS X do Apple sản xuất, mà phải dùng các phiên bản đã được tùy chỉnh bởi cộng đồng OSX86 để cho tương thích với các máy không thuộc Apple (có thể kể đến như Leo4all, iPC, Hazard, Kalyway, JaS...) hoặc tự thêm các driver vào đĩa cài đặt gốc để việc cài đặt và sử dụng diễn ra suôn sẻ.

Hệ điều hành UNIX và LINUX

Hệ điều hành UNIX đã được phát triển trong năm 1969 tại AT&T Bell Labs. Ưu điểm nổi bật của UNIX là độ tin cậy cao trong môi trường đa người dùng, nhiều phiên bản của UNIX được cài đặt trên các hệ thống máy tính lớn (mainframes) và máy vi tính (microcomputers).

Phiên bản Linux đầu tiên do Linus Torvalds viết vào năm 1991, lúc ông còn là một sinh viên của Đại học Helsinki tại Phần Lan. Ông làm việc một cách hăng say trong vòng 3 năm liên tục và cho ra đời phiên bản Linux 1.0 vào năm 1994. Bộ phận chủ yếu này được phát triển và tung ra trên thị trường dưới bản quyền GNU General Public License. Do đó mà bất cứ ai cũng có thể tải và xem mã nguồn của Linux.

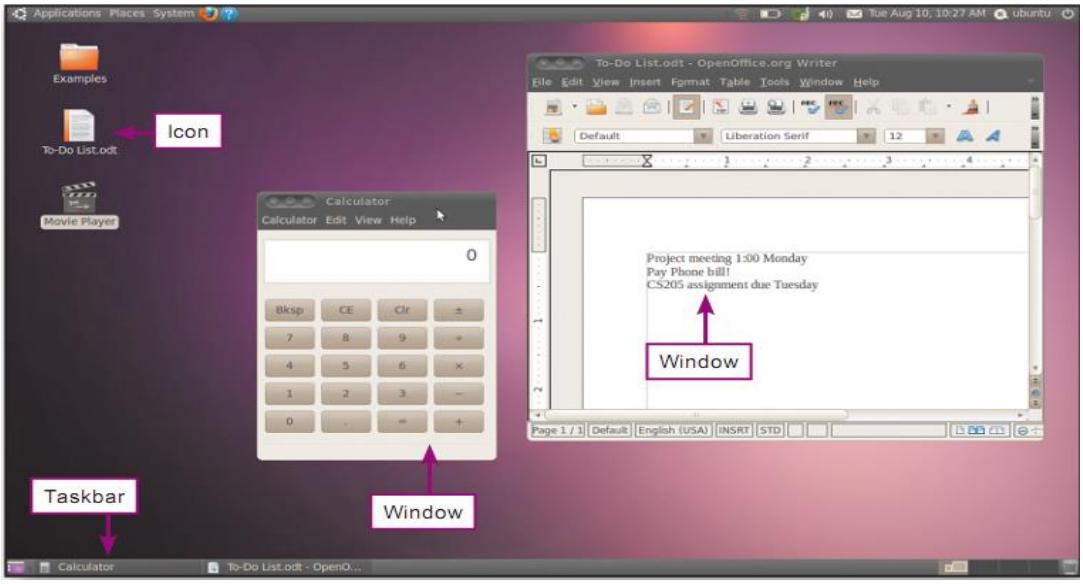
Linux được lấy cảm hứng và dựa trên một dẫn xuất UNIX gọi là Minix, được tạo ra bởi Andrew Tanenbaum. Linux thường được sử dụng như một hệ điều hành cho các máy chủ, nó không phổ biến cho các ứng dụng máy tính để bàn như Windows hay Mac OS.

Linux là khá độc đáo vì nó được phân phối cùng với mã nguồn của mình theo các điều khoản của mã nguồn mở GPL (General Public License), cho phép tất cả mọi người tự sửa đổi để tạo ra các bản sao cho riêng mình, cho những người khác hoặc với mục đích thương mại. Chính sách cấp phép này đã khuyến khích các lập trình viên phát triển các tiện ích, phần mềm và cải tiến Linux. Linux được phân phối chủ yếu trên web.

Linux chia sẻ một số tính năng kỹ thuật với UNIX, chẳng hạn như khả năng đa nhiệm và đa người dùng, an toàn và đáng tin cậy. Hệ điều hành Android, Symbian và Chromium cho các thiết bị cầm tay được xây dựng dựa trên nền Linux.

Tuy nhiên, Linux không có được giao diện thân thiện đối với người dùng như các hệ điều hành Windows và Mac. Các hệ thống phần mềm chạy trên nền Linux cũng còn hạn chế và chủ yếu hướng tới đối tượng khách hàng là các doanh nghiệp hoặc các kỹ thuật viên.

Một phiên bản của Linux có thể được tải về bao gồm các hạt nhân Linux, các tiện ích, giao diện đồ họa, các ứng dụng và cách thức cài đặt. Một số phiên bản của Linux tương đối thân thiện với người dùng là Fedora, Mandriva, openSUSE và Ubuntu (Hình 3.8).



Hình 3.8. Giao diện đồ họa của phiên bản Ubuntu

b) Hệ điều hành cho thiết bị di động

Có 6 loại hệ điều hành thống trị các thiết bị cầm tay, đó là: ***iOS, Symbian, BlackBerry OS, Android OS, Windows Phone 7 và HP webOS*** (như trong hình 3.9)

Hình 3.9. Sáu loại hệ điều hành điển hình cho các thiết bị cầm tay

IOS là hệ điều hành trên các thiết bị di động của Apple, ban đầu hệ điều hành này chỉ được phát triển để chạy trên iPhone, nhưng sau đó nó đã được mở rộng để chạy trên các thiết bị

của Apple như iPod touch, iPad và Apple TV. Ngày 31/5/2011, App Store của Apple chứa khoảng 500.000 ứng dụng iOS và được tải về tổng cộng khoảng 15 tỷ lần. Trong quý 4 năm 2010, có khoảng 26% điện thoại thông minh chạy hệ điều hành iOS, sau hệ điều hành Android của Google và Symbian của Nokia, giao diện người dùng của iOS dựa trên cơ sở thao tác bằng tay. Người dùng có thể tương tác với hệ điều hành này thông qua rất nhiều động tác bằng tay trên màn hình cảm ứng của các thiết bị của Apple.

Symbian là hệ điều hành được viết và sử dụng cho một số điện thoại di động. Symbian được sử dụng nhiều nhất bởi các điện thoại cao cấp của Nokia. Hãng này đã rất thành công với hệ điều hành này và có thời đã giúp Symbian trở thành hệ điều hành dành cho thiết bị di động phổ biến nhất thế giới. Tuy nhiên kể từ tháng 9/2012, Symbian đã đi vào giai đoạn thoái trào do không cạnh tranh nổi với các hệ điều hành mới tân tiến hơn như iOS, Android... Số người dùng càng ngày càng ít khiến thị phần Symbian thu hẹp và trở nên khiêm tốn so với các nền tảng khác. Đến ngày 25/1/2013, tập đoàn Nokia đã chính thức khai tử nền tảng Symbian, thiết bị cuối cùng chạy Symbian là Nokia 808 pureview.

BlackBerry OS là nền tảng phần mềm tư hữu do Research In Motion phát triển cho dòng sản phẩm cầm tay BlackBerry. BlackBerry OS cung cấp khả năng đa nhiệm và được thiết kế cho các thiết bị sử dụng phương pháp nhập đặc biệt, thường là trackball hoặc màn hình cảm ứng. Hệ điều hành được hỗ trợ MIDP 1.0 và WAP 1.2. Các phiên bản trước đó cho phép đồng bộ hóa không dây thư điện tử và lịch với Microsoft Exchange Server và với cả Lotus Domino. Phiên bản OS 4 hiện tại hỗ trợ MIDP 2.0, có khả năng kích hoạt không dây hoàn toàn và đồng bộ thư điện tử, lịch, công việc, ghi chú và danh bạ với Exchange và khả năng hỗ trợ Novell GroupWise, Lotus Notes khi kết hợp với BlackBerry Enterprise Server.

Các bản cập nhật cho BlackBerry OS có thể có nếu nhà mạng cung cấp thông qua dịch vụ BlackBerry OTASL.

Android là một hệ điều hành dựa trên nền tảng Linux được thiết kế dành cho các thiết bị di động có màn hình cảm ứng như điện thoại thông minh và máy tính bảng. Ban đầu, Android được phát triển bởi Tổng công ty Android, với sự hỗ trợ tài chính từ Google và sau này được chính Google mua lại vào năm 2005.

Android có mã nguồn mở và Google phát hành mã nguồn theo Giấy phép Apache. Chính mã nguồn mở cùng với một giấy phép không có nhiều ràng buộc đã cho phép các nhà phát triển thiết bị, mạng di động và các lập trình viên nhiệt huyết được điều chỉnh và phân phối Android một cách tự do. Ngoài ra, Android còn có một cộng đồng lập trình viên đông đảo chuyên viết các ứng dụng để mở rộng chức năng của thiết bị, bằng một loại ngôn ngữ lập trình Java có sửa đổi.

Những yếu tố này đã giúp Android trở thành nền tảng điện thoại thông minh phổ biến nhất thế giới. Mặc dù được thiết kế để chạy trên điện thoại và máy tính bảng, Android đã xuất hiện trên TV, máy chơi game và các thiết bị điện tử khác. Bản chất mở của Android cũng khích lệ một đội ngũ đông đảo lập trình viên và những người đam mê sử dụng mã nguồn mở để tạo ra những dự án do cộng đồng quản lý. Những dự án này bổ sung các tính năng cao cấp cho những người dùng thích tìm tòi hoặc đưa Android vào các thiết bị ban đầu chạy hệ điều hành khác.

Windows Phone 7 là thế hệ kế tiếp của dòng điện thoại chạy hệ điều hành Microsoft Windows Mobile. Windows Phone 7 được phát triển dựa trên phần lõi là Windows CE 7 giống Zune HD, trong khi các phiên bản trước lại dựa trên Windows CE 5.

Như đã nói ở trên, Windows Phone 7 hoàn toàn khác Windows Mobile, khác cả về phần cứng lẫn phần mềm: Giao diện sử dụng dạng lật mở hoàn toàn mới lạ, chú trọng tính năng nhập liệu bằng ngón tay, kết hợp và mở rộng đầy đủ với các thành phần của Zune và Xbox, đòi hỏi

cấu hình phần cứng rất khắt khe đối với các đối tác sản xuất. Ngoài ra, Windows Phone 7 hỗ trợ cả Outlook và Office.

HP WebOS là một hệ điều hành di động dựa trên hạt nhân Linux, ban đầu được phát triển bởi Palm, mà sau này được mua lại bởi Hewlett-Packard. Nói theo một cách dễ hiểu nhất, WebOS là một hệ điều hành ảo chạy trong trình duyệt Web. Đến nay, LG Electronics đã mua lại hệ điều hành này nhằm phát triển dòng tivi thông minh của hãng. Nếu như đối với Android hay iOS, người dùng chỉ có thể chạy được một ứng dụng trên màn hình, việc chuyển qua chuyển lại giữa các ứng dụng phải thông qua một danh sách các ứng dụng đang chạy được gọi lên bằng một thao tác nào đó thì ở WebOS các ứng dụng đang chạy được quản lý dưới dạng các cửa sổ. Hoặc hệ thống hiển thị thông báo từ email, tin nhắn và thông báo từ các ứng dụng của WebOS cũng được đánh giá tốt hơn hẳn so với các hệ điều hành khác như iOS hay Windows Phone.

3.2.4. Quản lý dữ liệu trên bộ nhớ ngoài

Dữ liệu trên bộ nhớ ngoài mà Windows nói riêng và các hệ điều hành nói chung quản lý là tệp và thư mục.

a) Khái niệm tệp tin

Tệp, hay tệp tin (file) là một tập hợp các thông tin có liên quan đến nhau do người dùng tạo ra trong máy tính hoặc các thiết bị số hóa khác (ảnh, video), được lưu trữ trong máy. Các thông tin này là các giá trị số, một hay nhiều chuỗi ký tự, ký hiệu giống hoặc khác nhau. Tệp tin được đặt tên và lưu trữ trong các thiết bị lưu trữ khác nhau như đĩa cứng, đĩa CD, USB...

Tên tệp tin gồm hai phần, phần tên chính và phần mở rộng, được cách nhau bằng dấu chấm (.). Phần tên chính thông thường do người dùng đặt, còn phần mở rộng thường được quy định bởi phần mềm tạo ra tệp, ban đầu phần tên chính chỉ bao gồm tám ký tự và phần mở rộng từ một đến ba ký tự, hiện nay tên tệp có độ dài tùy ý tùy thuộc vào hệ thống tệp tin và hệ điều hành, trong một số trường hợp có thể đặt tên có dấu tiếng Việt.

Một tệp tin luôn luôn kết thúc bằng 1 ký tự đặc biệt (hay dấu kết thúc) có mã ASCII là 255 ở hệ thập phân. Ký tự này thường được ký hiệu là EOF (từ chữ End Of File).

Một tệp tin có thể không chứa một thông tin nào ngoại trừ tên và dấu kết thúc. Tuy nhiên, điều này không hề mâu thuẫn với định nghĩa vì bản thân tên của tệp tin cũng đã chứa thông tin. Những tệp tin này gọi là tệp tin rỗng hay tệp tin trống.

Độ dài (kích thước) của tệp tin có thể chỉ phụ thuộc vào khả năng của máy tính, khả năng của hệ điều hành cũng như vào phần mềm ứng dụng dùng nó, đơn vị nhỏ nhất dùng để đo độ dài của tệp tin là byte, độ dài của tệp tin không bao gồm độ dài của tên tệp tin và dấu kết thúc.

b) Khái niệm thư mục

Để có thể tổ chức quản lý tốt tệp tin trên đĩa người ta lưu các tệp tin thành từng nhóm và lưu trong từng chỗ riêng gọi là thư mục. Nói cách khác, thư mục là một dạng tệp tin đặc biệt có công dụng như là một ngăn chứa, được dùng trong việc quản lý và sắp xếp các tệp tin. Thư mục có thể chứa các tệp tin và các thư mục con bên trong, các thư mục con này cũng có thể chứa thêm các tệp tin và các thư mục con khác nữa... tạo thành một *cây thư mục* trên đĩa. Có thể tạo nhiều thư mục dùng để chứa các tệp tin khác nhau giúp phân loại chúng để thuận tiện trong việc tìm kiếm, sử dụng.

Thư mục gốc là thư mục do định dạng đĩa tạo ra và chúng ta không thể xóa được, mỗi đĩa chỉ có một thư mục gốc, từ đây người sử dụng có thể tạo ra các thư mục con, thư mục hiện hành là thư mục mà người sử dụng đang thao tác.

Cũng giống như tệp tin, thư mục có thể được đặt tên tùy ý nhưng không cần phải có phần mở rộng, độ dài của tên cũng tùy thuộc vào hệ thống tệp tin và hệ điều hành, trong một số trường hợp có thể đặt tên có dấu tiếng Việt. Bản thân thư mục chiếm một dung lượng không đáng kể trên thiết bị lưu trữ, nhưng nếu có chứa các tệp tin bên trong thì nó sẽ có dung lượng bằng tổng dung lượng tất cả các tệp tin cộng dồn lại.

CÂU HỎI VÀ BÀI TẬP

1. Khái niệm phần mềm?

2. Có mấy loại phần mềm? Đặc điểm cơ bản để phân loại phần mềm ứng dụng và phần mềm hệ thống là gì?

3. Kể tên một số loại phần mềm nhúng điển hình.

4. Kể tên một số loại phần mềm tiện ích điển hình.

5. Phần mềm nguồn mở là gì? Hãy nêu những điểm ưu việt của phần mềm nguồn mở.

6. Trình bày khái niệm và chức năng của hệ điều hành?

7. Hệ điều hành đã phát triển qua mấy thế hệ? Thế hệ nào bắt đầu xuất hiện hệ điều hành đa nhiệm? Ưu điểm của hệ điều hành đa nhiệm là gì?

8. Trình bày khái niệm tệp và thư mục. Nêu các nguyên tắc đặt tên cho tệp và thư mục.

Chương 4

MẠNG MÁY TÍNH VÀ INTERNET

Chương này giới thiệu sơ lược về mạng máy tính, các thành phần cơ bản của mạng máy tính, cách kết nối mạng và cách phân loại mạng máy tính. Từ đó giúp sinh viên có kiến thức tổng quan về mạng máy tính, đồng thời sinh viên còn được trang bị các kiến thức về mạng Internet, lịch sử và các dịch vụ cơ bản của Internet.

4.1. MẠNG MÁY TÍNH

Ngày nay, nhu cầu sử dụng máy tính không ngừng được tăng lên về cả số lượng và ứng dụng, đặc biệt là sự phát triển hệ thống mạng máy tính, kết nối các máy tính lại với nhau thông qua môi trường truyền tin để cùng nhau chia sẻ tài nguyên trên mạng góp phần làm tăng hiệu quả của các ứng dụng trong tất cả các lĩnh vực khoa học kỹ thuật, kinh tế, quân sự, văn hoá. Sự kết hợp của máy tính với hệ thống truyền thông (communication), đặc biệt là viễn thông (telecommunication) đã tạo ra một sự chuyển biến có tính cách mạng trong vấn đề tổ chức khai thác và sử dụng các hệ thống máy tính. Từ đó đã hình thành các môi trường trao đổi thông tin tập trung, phân tán, cho phép đồng thời nhiều người cùng trao đổi thông tin với nhau một cách nhanh chóng và hiệu quả từ những vị trí địa lý khác nhau. Các hệ thống như thế được gọi là mạng máy tính (computer network). Như vậy, *mạng máy tính là một tập hợp gồm nhiều máy tính hoặc thiết bị xử lý thông tin được kết nối với nhau qua các đường truyền vật lí theo một kiến trúc mạng nhất định và có sự trao đổi dữ liệu với nhau. Nhờ có mạng máy tính, thông tin từ một máy tính có thể được truyền sang máy tính khác.*

Ví dụ về mạng:

- Mạng máy tính của Học viện Nông nghiệp Việt Nam.
- Mạng của Công ty FPT.

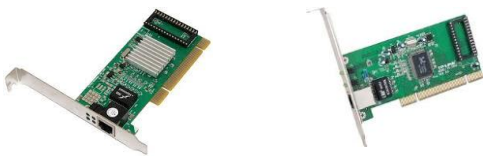
4.1.1. Các thành phần cơ bản của mạng máy tính

Một mạng máy tính có thể có các thành phần sau:

- Các máy tính (Computer): Được dùng để xử lý, lưu trữ và trao đổi thông tin. Mỗi máy tính trong mạng máy tính là một nút của mạng.

Máy tính

- Cạc mạng (Network Interface Card, NIC): Là một bản mạch cung cấp khả năng truyền thông mạng cho một máy tính.



Cạc mạng máy tính

- Đường truyền vật lý: Là phương tiện (media) truyền tải dữ liệu, là nơi trên đó dữ liệu được truyền đi. Ta có thể chia đường truyền thành hai loại là hữu tuyến và vô tuyến.

Dây cáp mạng

Sóng vô tuyến

- Các thiết bị kết nối mạng: Là các thiết bị để liên kết các máy tính và các mạng với nhau như HUB, SWITCH, ROUTER...

Switch

Hub

Router

- Các thiết bị đầu cuối (terminal) như: Máy photo, máy in, máy scan, camera máy tính...

Máy photo

Máy in

Máy scan

Webcam

- Các phụ kiện mạng như giắc cắm, ổ cắm...

Giắc cắm mạng

Ổ cắm mạng

Hình 4.1. Các thành phần cơ bản của mạng

- Hệ điều hành mạng là một phần mềm điều khiển sự hoạt động của mạng.

- Các ứng dụng trên mạng (ví dụ: email, tìm kiếm, www, hệ quản trị cơ sở dữ liệu...).

- Kiến trúc mạng máy tính (network architecture) thể hiện cách kết nối máy tính với nhau và qui ước truyền dữ liệu giữa các máy tính như thế nào. Cách nối các máy tính với nhau gọi là hình trạng (topology) của mạng. Tập các qui ước truyền thông gọi là giao thức (protocol).

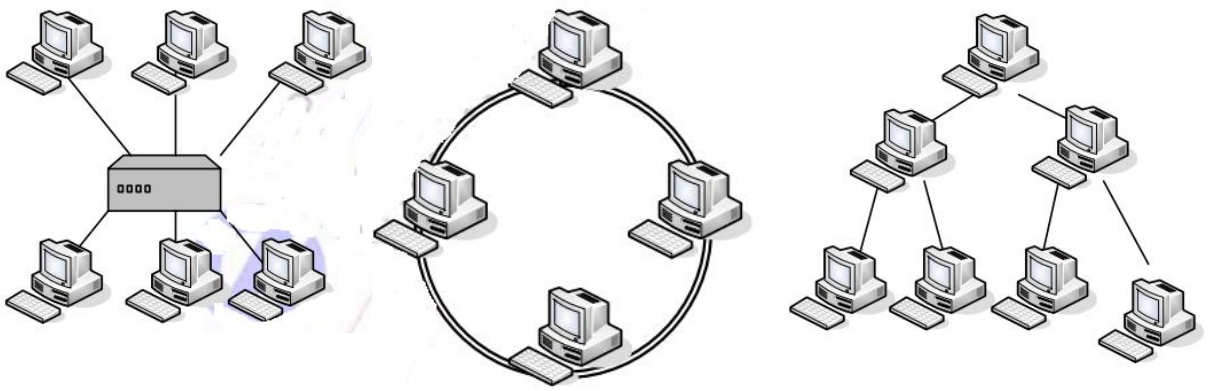
4.1.2. Mô hình kết nối và giao thức mạng

a) Mô hình kết nối (Topo mạng)

Dựa theo kĩ thuật truyền tải thông tin, người ta có thể chia mạng thành hai kiểu nối mạng chủ yếu là điểm-điểm (*point-to-point network*) và quảng bá (*broadcast hay point-to-multipoint network).*

- *Kiểu kết nối điểm-điểm*

Trong hệ thống mạng này, các đường truyền nối các nút thành từng cặp và mỗi nút đều có trách nhiệm lưu trữ dữ liệu tạm thời, sau đó chuyển tiếp dữ liệu đến nút lân cận nó (nút được nối trực tiếp với nó). Nút lân cận sẽ chuyển tiếp dữ liệu như vậy cho đến khi dữ liệu đến đích, do cách thức làm việc như thế nên mạng kiểu này còn được gọi là mạng "Lưu và chuyển tiếp" (Store and Forward). Kiểu nối mạng điểm - điểm có ba dạng chính là: hình sao (star), chu trình (loop) và hình cây (tree).

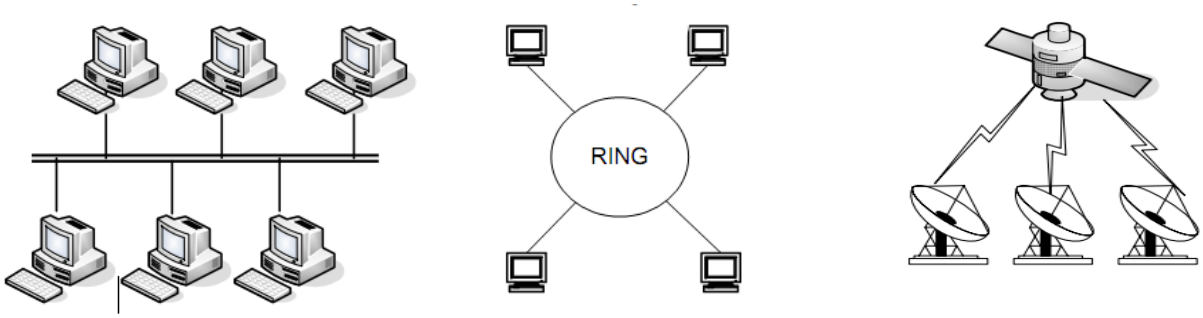


Topo mạng hình sao – Star Topo mạng chu trình – Loop Topo mạng hình cây – Tree

Hình 4.2. Các mạng có cấu trúc điểm – điểm

- *Kiểu kết nối quảng bá*

Trong hệ thống mạng quảng bá chỉ tồn tại một kênh truyền vật lý, tất cả các nút trong mạng cùng truy cập chung trên kênh truyền vật lý này. Dữ liệu được gửi đi từ một nút nào đó sẽ được tất cả các nút còn lại tiếp nhận, do đó cần chỉ ra địa chỉ đích của dữ liệu để mỗi nút căn cứ vào đó kiểm tra dữ liệu xem có phải là gửi đến mình hay không. Kiểu kết nối quảng bá có một số dạng chính: bus và vòng tròn.



Topo mạng – Bus Topo mạng – Ring Vệ tinh

Hình 4.3. Các mạng có cấu trúc quảng bá

b) Giao thức mạng (Network protocol)

Việc trao đổi thông tin, cho dù là đơn giản nhất cũng đều phải tuân theo những quy tắc nhất định. Việc truyền tín hiệu trên mạng cần phải có những quy tắc, quy ước về nhiều mặt, từ khuôn dạng (cú pháp, ngữ nghĩa) của dữ liệu cho tới các thủ tục gửi, nhận dữ liệu, kiểm soát hiệu quả, chất lượng truyền tin và xử lý các lỗi. Yêu cầu về xử lý và trao đổi thông tin của người sử

dụng càng cao thì các quy tắc càng nhiều và phức tạp hơn. Tập hợp tất các quy tắc, quy ước đó được gọi là giao thức (protocol) của mạng. Các thành phần chính của một giao thức bao gồm:

- Cú pháp: Định dạng dữ liệu, phương thức mã hóa và các mức tín hiệu.
- Ngữ nghĩa: Thông tin điều khiển, điều khiển lưu lượng và xử lý lỗi...

4.1.3. Phân loại mạng máy tính

Có nhiều cách phân loại mạng khác nhau tùy theo yếu tố chính được chọn để làm chỉ tiêu phân loại như theo khoảng cách địa lý, theo topo kết nối hay kiểu truyền thông mà mạng sử dụng...

Cách 1: Phân loại mạng theo khoảng cách địa lý

Nếu lấy "*khoảng cách địa lý"* làm yếu tố chính thì mạng được phân chia thành mạng cục bộ, mạng đô thị, mạng diện rộng, mạng toàn cầu.

- **Mạng cục bộ** (LAN: Local Area Network): là mạng được cài đặt trong một phạm vi tương đối nhỏ (ví dụ trong một cơ quan, công ty, trường học...).
- **Mạng đô thị** (MAN: Metropolitan Area Network): là mạng được cài đặt trong phạm vi một thành phố, một trung tâm kinh tế, phạm vi địa lý là hàng trăm km.
- **Mạng diện rộng** (WAN: Wide Area Network): phạm vi hoạt động của mạng có thể vượt qua biên giới một quốc gia, có thể cả một khu vực.
- **Mạng toàn cầu** (VAN: Vast Area Network): phạm vi của mạng trải rộng trên khắp các lục địa.

Khoảng cách địa lý có tính chất tương đối, đặc biệt trong thời đại ngày nay. Những tiến bộ và sự phát triển của công nghệ truyền dẫn và quản lý mạng khiến ranh giới khoảng cách địa lý giữa các mạng trở nên mờ nhạt. Hiện tại, khi phân loại mạng theo khoảng cách địa lý người ta thường đồng nhất 4 loại mạng trên thành 2 loại sau:

- WAN là mạng lớn trên diện rộng, hệ thống mạng này có thể truyền thông và trao đổi dữ liệu trong một phạm vi rộng lớn như trong một quốc gia hay quốc tế.
- LAN là mạng cục bộ được bố trí trong phạm vi hẹp như một cơ quan, một bộ, ngành... Một số mạng LAN có thể nối lại với nhau để tạo thành một mạng LAN lớn hơn.

Cách 2: Phân loại mạng theo mối quan hệ giữa các máy trong mạng

- Mạng bình đẳng (peer-to-peer, còn được gọi là mạng ngang hàng): Các máy có quan hệ ngang hàng, một máy có thể yêu cầu một máy khác phục vụ.
- Mạng khách/chủ (client/server): Một số máy là server (máy chủ) chuyên phục vụ các máy khác gọi là máy khách (client) hay máy trạm (work station) khi có yêu cầu. Các dịch vụ có thể là cung cấp thông tin, tính toán hay các dịch vụ Internet.

4.2. INTERNET

4.2.1. Một số khái niệm

a) Internet

Internet là một hệ thống thông tin toàn cầu gồm các mạng máy tính được liên kết với nhau. Hệ thống này sử dụng giao thức truyền thông TCP/IP (Transmission Control Protocol/ Internet Protocol) để truyền dữ liệu.

Điểm khác với các mạng máy tính thông thường là ở chỗ Internet không thuộc sở hữu của ai cả và không có mạng nào điều hành mạng nào, chỉ có các ủy ban điều phối và kĩ thuật (các tổ chức phi lợi nhuận, phi chính phủ - non profit, non govermental) quản lý việc cấp địa chỉ và nghiên cứu các chính sách cũng như công nghệ trên Internet.

Hệ thống Internet bao gồm hàng ngàn mạng máy tính của các doanh nghiệp, các viện nghiên cứu và các trường đại học, người dùng cá nhân và các chính phủ trên toàn cầu. Chúng cung cấp một khối lượng thông tin và dịch vụ khổng lồ trên Internet, mạng Internet mang lại rất nhiều tiện ích hữu dụng cho người sử dụng, một trong các tiện ích phổ thông của Internet là hệ thống World Wide Web (WWW), thư điện tử (email), trò chuyện trực tuyến (chat), máy tìm kiếm (search engine), các dịch vụ thương mại và chuyển ngân và các dịch vụ về y tế, giáo dục như hình thức đào tạo trực tuyến, chữa bệnh từ xa...

- **Lịch sử Internet**

Nguồn gốc của Internet là ARPANET, một dự án do bộ Quốc phòng Mỹ khởi đầu năm 1969 thực hiện kết nối mạng giữa Bộ Quốc phòng Mỹ với một số cơ sở nghiên cứu khoa học lớn ở Mỹ. Giao thức truyền thông lúc đó được dùng là NCP (Network Control Protocol), đến giữa những năm 70, họ giao thức TCP/IP được Vincerf (đại học Stanford) và Robert Kahn phát triển, đến năm 1983 thì họ giao thức này hoàn toàn thay thế NCP trong ARPANET.

Sau một thời gian kế hoạch sử dụng ARPANET không thực hiện được như mong muốn vì một số lý do kỹ thuật và chính trị. Do vậy, năm 1986 Hội đồng Khoa học Quốc gia Hoa kỳ (National Science Foundation - NSF) đã xây dựng mạng NSFNET, một mạng riêng và hoạt động nhanh hơn nhiều để nối với các trung tâm siêu tính toán (tốc độ đường truyền là 1,5 Mb/s thay vì 560 Kb/s trong ARPANET). Sự xuất hiện của mạng NSFNET đã thúc đẩy sự tăng trưởng của Internet, một xa lộ thông tin mới hình thành và nhiều trường đại học, viện nghiên cứu đã tham gia vào cộng đồng Internet. Sau đó các tổ chức chính phủ, giới kinh doanh cũng vào cuộc và ngày càng chiếm tỷ trọng đáng kể trong thế giới Internet. Về mặt địa lý Internet cũng đã nhanh chóng vượt ra khỏi nước Mỹ và trở thành mạng toàn cầu với vài chục triệu người dùng như hiện nay, đến năm 1990 thì quá trình chuyển đổi sang Internet hoàn tất và ARPANET ngừng hoạt động.

Về kiến trúc, Internet cũng đã có những thay đổi. Trước đây mạng chỉ sử dụng giao thức IP thì ngày nay nhiều mạng với kiến trúc khác (không dùng IP) nhờ có cầu kết nối đa giao thức (multiprotocol gateway) vẫn có thể kết nối được vào Internet và sử dụng đầy đủ các dịch vụ thông tin trên Internet.

b) Máy chủ và máy khách

Máy chủ (server) là một máy tính hoặc một hệ thống máy tính cung cấp các tài nguyên và dịch vụ cho cả hệ thống mạng sử dụng. Các máy chủ thường có cấu hình mạnh (tốc độ xử lý nhanh, bộ nhớ lưu trữ lớn) hoặc là các máy chuyên dụng. Dựa vào chức năng có thể chia thành các loại server như sau:

- File Server: phục vụ các yêu cầu hệ thống tệp tin trong mạng.
- Print Server: phục vụ các yêu cầu in ấn trong mạng.
- Application Server: cho phép các ứng dụng chạy trên các server và trả về kết quả cho client.
- Mail Server: cung cấp các dịch vụ về gửi nhận e-mail.
- Web Server: cung cấp các dịch vụ về web.
- Database Server: cung cấp các dịch vụ về lưu trữ, tìm kiếm thông tin.
- Communication Server: quản lý các kết nối từ xa.

Máy khách (client) là máy tính/hệ thống máy tính sử dụng các tài nguyên và các dịch vụ được cung cấp bởi máy chủ.

c) Địa chỉ IP và tên miền

* *Địa chỉ IP*

Để tham gia Internet, các thực thể truyền thông (máy tính và các thiết bị mạng có các hoạt động xử lý – host) cần được cấp một địa chỉ gọi là địa chỉ IP. Nói một cách đơn giản: IP là một địa chỉ của một máy tính khi tham gia vào mạng nhằm giúp cho các máy tính có thể chuyển thông tin cho nhau một cách chính xác, tránh thất lạc. Có thể coi địa chỉ IP trong mạng máy tính giống như địa chỉ nhà của bạn để nhân viên bưu điện có thể đưa thư đúng cho bạn chứ không phải một người nào khác. Giao thức tầng mạng trong bộ giao thức TCP/IP hiện tại đang là IPv4 (Internet-working protocol verision 4), IPv4 cung cấp truyền thông host-to-host giữa những hệ thống trên Internet, mặc dù được thiết kế khá tốt, tuy nhiên với sự phát triển nhanh chóng của các ứng dụng máy tính, IPv4 đang tiến tới giới hạn của nó, cụ thể với nhu cầu địa chỉ IP tăng nhanh, không gian địa chỉ 32 bit trong phiên bản 4 đang bị cạn kiệt. Phiên bản IPv6 là một phiên bản mới của Internet. Nó được xây dựng trên cơ sở của giao thức IPv4 nhằm tận dụng các ưu điểm và khắc phục hạn chế của IPv4.

Địa chỉ IPv4

Địa chỉ IP đang được sử dụng hiện tại là IPv4 có độ dài 4 byte (32 bit), mỗi byte tách biệt nhau bằng dấu chấm (.) để dễ đọc địa chỉ, theo đó giá trị trong mỗi byte được viết thành một số thập phân (một byte có giá trị nằm trong đoạn từ 0 đến 255), ví dụ 192.168.10.1.

Một địa chỉ IPv4 bao gồm 3 thành phần chính như sau:

Class Bit	NetID	HostID

Bit 1 2 3 …… 32

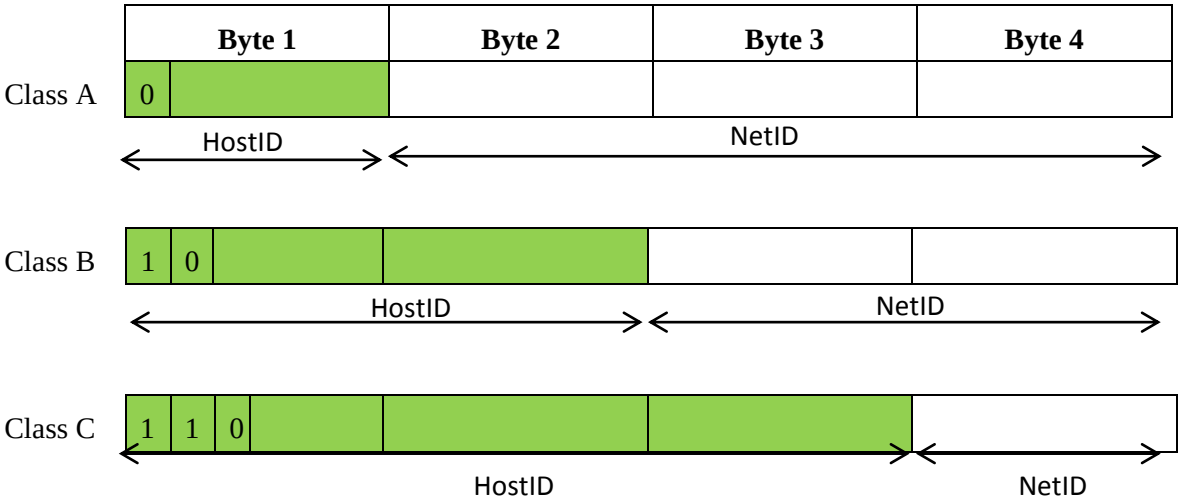
Trong đó:

Class bit: Bit nhận dạng lớp để phân biệt địa chỉ ở lớp nào.

NetID: Địa chỉ của mạng.

HostID: Địa chỉ của các máy trên mạng.

Địa chỉ IP được chia ra thành 5 lớp A, B, C, D, E. Sự khác nhau cơ bản giữa các lớp địa chỉ này là ở khả năng tổ chức các cấu trúc con của nó. Hiện tại đã dùng hết lớp A, lớp B và gần hết lớp C; lớp D, E tổ chức Internet đang để dành cho mục đích khác nên không cấp địa chỉ, ở đây chúng ta chỉ nghiên cứu 3 lớp địa chỉ đầu.



Hình 4.4. Cấu trúc địa chỉ IP các lớp A-B-C

Qua cấu trúc của các lớp địa chỉ IP chúng ta có những nhận xét sau:

- Bit nhận dạng lớp là những bit đầu tiên: của lớp A là 0, lớp B là 10, lớp C là 110.
- Lớp A có 127 địa chỉ mạng, mỗi mạng cho phép đánh địa chỉ cho 2^{24} máy.
- Lớp B có 2^{14} địa chỉ mạng, mỗi mạng cho phép đánh địa chỉ cho 2^{16} máy.
- Lớp C có 2^{21} địa chỉ mạng, mỗi mạng cho phép đánh địa chỉ cho 2^{8} máy.

Địa chỉ IPv6

Trước tình hình cạn kiệt địa chỉ IPv4, đã có nhiều phương pháp được đưa ra để khắc phục tình trạng này như NAT (Network Address Translation - biên dịch địa chỉ mạng), CIDR (Classless Inter-Domain Routing – lược đồ địa chỉ mới của Internet)... nhưng nguy cơ cạn dần địa chỉ vẫn chưa được khắc phục. Việc triển khai phiên bản mới - phiên bản IPv6 với không gian địa chỉ lớn hơn và có nhiều ưu điểm hơn phiên bản cũ là điều cần thiết. Địa chỉ IPv6 được thiết kế với nhiều ưu điểm như: không gian địa chỉ lớn, hỗ trợ tự cấu hình, tích hợp khả năng xác thực

và bảo vệ an ninh, hỗ trợ tốt hơn tính năng di động, khả năng mở rộng... IPv6 sử dụng 128 bit để đánh địa chỉ, có chiều dài gấp bốn lần địa chỉ IPv4, có thể tạo ra hơn $3,4*10^{38}$ tổ hợp địa chỉ IP và thường được biểu diễn ở hệ hexadecimal (hệ 16).

Địa chỉ IPv6 có 128 bit, nên việc nhớ được địa chỉ này rất khó khăn. Do đó, để viết địa chỉ IPv6, người ta đã chia 128 bit ra thành 8 nhóm, mỗi nhóm chiếm 2 bytes, gồm 4 số được viết dưới hệ số 16 và mỗi nhóm được ngăn cách nhau bằng dấu hai chấm.

Ví dụ: FEDL:8435:7356:EADC:BA98:2010:3280:ABCD

Trường hợp có nhiều số 0 ở các bit đầu trong một nhóm, chúng ta có thể lược bỏ các số 0 này đi. Ví dụ với địa chỉ 1088:0000:0000:0000:0008:0800:200C:463A, ta có thể viết 0 thay vì phải viết là 0000, viết 8 thay vì phải viết 0008, viết 800 thay vì phải viết là 0800. Kết quả của địa chỉ sau khi thu gọn:1088:0:0:0:8:800:200C:463A.

Trường hợp có nhiều nhóm 0 liên tiếp, chúng ta có thể nhóm các số 0 lại thành 2 dấu hai chấm "::" (chỉ thực hiện một lần), địa chỉ ở trên, ta có thể viết lại như sau: 1088::8:800:200C:463A

* Tên miền (DNS – Domain Name System)

Mỗi máy tính trong mạng muốn liên lạc hay trao đổi thông tin, dữ liệu cho nhau cần phải biết rõ địa chỉ IP của nhau. Nếu số lượng máy tính nhiều thì việc nhớ những địa chỉ IP này rất là khó khăn. Chính vì vậy người ta sử dụng một hệ thống đặt tên gọi là tên miền (domain name) để đặt tên cho các máy tính trên mạng. Đối với con người việc nhớ tên máy dù sao cũng dễ dàng hơn vì chúng có tính trực quan và gợi nhớ hơn địa chỉ IP. Vì thế, người ta nghĩ ra cách làm sao ánh xạ địa chỉ IP thành tên máy tính, đó là dịch vụ đánh tên miền DNS. DNS cho phép người sử dụng có thể truy nhập tới một máy tính bằng tên của nó thay vì bằng địa chỉ IP.

Mỗi tên miền có thể gồm nhiều trường phân cách nhau bởi một dấu chấm. Theo quy ước tên miền được đặt theo một cây phân lớp mà trường đầu tiên là trường địa lý (thường là theo cách viết tắt của tên nước). Ví dụ vn chỉ Việt Nam, th chỉ Thái Lan, fr chỉ Pháp, jp chỉ Nhật... Các tên miền không có lớp địa lý được ngầm hiểu là Mỹ.

Trường thứ hai thường là lớp lĩnh vực hoạt động:

Tên miền	**Mô tả**
.com	Các tổ chức, công ty thương mại
.org	Các tổ chức phi lợi nhuận
.net	Các trung tâm hỗ trợ về mạng
.edu	Các tổ chức giáo dục
.gov	Các tổ chức thuộc chính phủ
.mil	Các tổ chức quân sự

Từ trường thứ 3 trở đi do các tổ chức tự đặt tên, miễn là không trùng nhau và được các tổ chức quản lý tên miền chấp nhận.

Ví dụ: Máy chủ Học viện Nông nghiệp Việt Nam có tên miền là vnua.edu.vn, có địa chỉ IP là 220.221.107.132. Máy chủ Đại học Bách khoa Hà Nội có tên miền là hust.edu.vn có địa chỉ IP là 202.191.56.197.

c) Trang web và website

Trang web (web page): là một tài liệu HTML, trong đó lưu trữ các nội dung và định dạng văn bản, hình ảnh, âm thanh theo định dạng HTML (Hypertext Markup Language, ngôn ngữ đánh dấu siêu văn bản).

Website: là một tập hợp các trang web liên kết với nhau và được quản lý bởi một cá nhân hay một tổ chức nào đó, ví dụ như website của Học viện Nông nghiệp Việt Nam.

Về kích thước: một website có thể có kích thước rất lớn, gồm hàng trăm trang trên đó, ví dụ như website của Học viện Nông nghiệp Việt Nam.

Về nội dung: một website có thể chứa nhiều thông tin dành cho nhiều người, cho nhiều mục đích và nhu cầu khác nhau. Tuy nhiên, một trang web chỉ phục vụ cho một mục đích hoặc một nhu cầu cụ thể.

4.2.2. Kết nối Internet

Để kết nối được Internet ta cần:

- Máy tính có Modem (Dial-up, ADSL) hoặc card mạng.
- Có thuê bao kết nối với Internet: qua mạng, qua đường điện thoại, đường thuê riêng của bưu điện. Thông thường hiện nay kết nối qua điện thoại hoặc qua ADSL.
- Có tài khoản Internet ở trên mạng hay ở một nhà cung cấp dịch vụ Internet (Internet Service Provider, ISP), ví dụ như VNPT, FPT, Viettel.
- Có phần mềm Internet thông dụng như Web browser để xem trang web, ví dụ: IE, FireFox, Chrome, phần mềm để xem thư hay chat như Outlook, Messenger.

4.3. MỘT SỐ DỊCH VỤ CƠ BẢN CỦA INTERNET

4.3.1. WWW (World Wide Web)

Đây là khái niệm mà người dùng Internet quan tâm nhiều nhất hiện nay, từ viết tắt là WWW hay gọi ngắn gọn là Web. Web là một công cụ, hay đúng hơn là dịch vụ của Internet. Web được tạo ra năm 1989 bởi Sir Tim Berners-Lee, làm việc tại CERN – Thụy Sỹ, là một hệ thống các trang văn bản được liên kết với nhau thông qua Internet.

Khác với các dịch vụ trước đây của Internet, Web chứa thông tin bao gồm văn bản, hình ảnh, âm thanh và thậm chí cả video được kết hợp với nhau. Web cho phép bạn có thể truy cập vào mọi ngõ ngách trên Internet, là những điểm chứa cơ sở dữ liệu gọi là Website. Nhờ có Web, nên dù không phải là một chuyên gia, bạn vẫn có thể sử dụng Internet.

Phần mềm sử dụng để định hướng Web gọi là trình duyệt Web (Web browser). Hiện nay, trình duyệt thông dụng nhất là Google Chrome của Google, tiếp đó là Internet Explorer của Microsoft, ngoài ra còn có Opera, Mozilla Firefox... Người dùng sử dụng một phần mềm Web

browser để xem thông tin trên các máy chủ WWW. Tại máy chủ phải có một phần mềm Web server, phần mềm này thực hiện nhận các yêu cầu từ Web browser gửi lên và thực hiện yêu cầu đó.

4.3.2. Tìm kiếm

Tìm kiếm thông tin là hoạt động phổ biến đối với người sử dụng Internet. So với thông tin được lưu trữ trên những phương tiện khác, việc truy cập và tìm kiếm thông tin được lưu trữ trên Internet là dễ dàng hơn. Ngoài ra, kết quả tìm kiếm đạt được nhiều hơn, đây là điểm mạnh nhưng đôi khi cũng là điểm yếu của Internet vì khi có quá nhiều kết quả tìm kiếm, ta phải tốn thời gian để lọc lại những thông tin phù hợp.

Trên Internet có rất nhiều trang Web cung cấp các công cụ tìm kiếm (search engine). Mỗi công cụ tìm kiếm có những điểm mạnh và điểm yếu riêng, do đó, khi tìm thông tin ta nên bắt đầu bằng công cụ quen thuộc nhất. Nếu kết quả tìm kiếm chưa tốt, ta có thể thực hiện lại việc tìm kiếm với công cụ tìm kiếm khác.

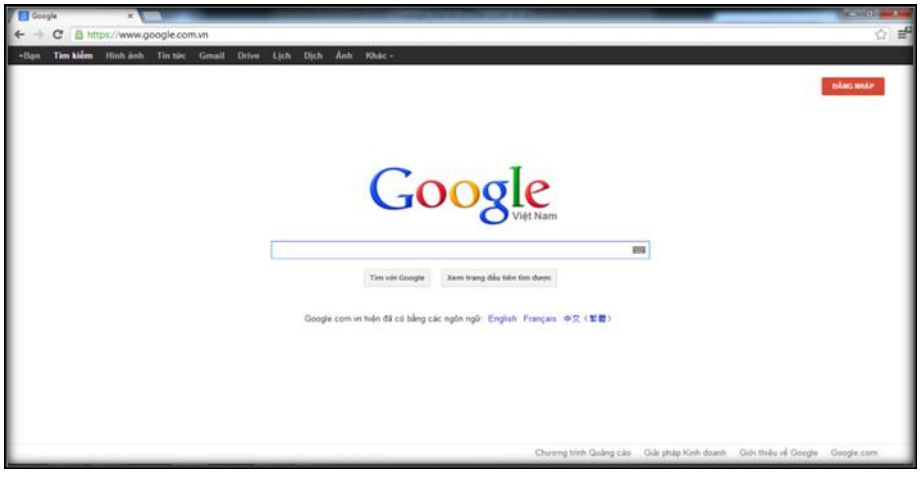
Có rất nhiều website được thiết kế chuyên để tìm kiếm thông tin trên Internet, trong đó có thể kể đến một số trang nổi tiếng sau:

http://www.google.com	http://www.yahoo.com	http://www.altavista.com
http://www.msn.com	http://www.hotbot.com	http://www.lycos.com

Để sử dụng, bạn gõ địa chỉ trang tìm kiếm vào thanh address của trình duyệt và nhấn Enter (tương tự như khi bạn vào một trang web thông tin).

Tìm kiếm với http://www.google.com.vn

Trang tìm kiếm google được sử dụng khá phổ biến, hỗ trợ tính năng tìm kiếm tốt và có thể lựa chọn giao diện theo nhiều ngôn ngữ của nhiều quốc gia. Để thực hiện tìm kiếm với trang google chúng ta thực hiện: mở trình duyệt web, nhập địa chỉ http://www.google.com.vn.



Hình 4.5. Giao diện chính của cửa sổ trang web Google

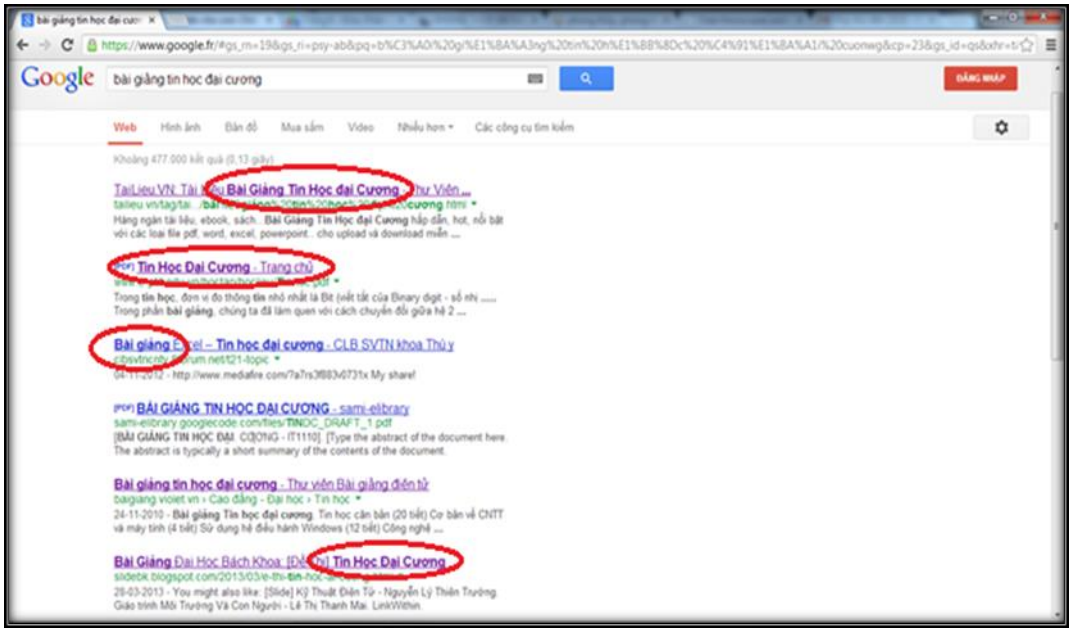
Tại đây chúng ta có thể tìm kiếm nhiều đối tượng khác nhau như: địa chỉ các trang web (Internet), hình ảnh, nhóm, thư mục... Nhập cụm từ khóa liên quan đến vấn đề cần tìm kiếm vào ô tìm kiếm và nhấn **"Tìm với Google".**

Các phép toán trong điều kiện tìm kiếm

Nhiều người cho rằng khi sử dụng một công cụ tìm kiếm (google, yahoo...), trang tìm kiếm sẽ tự động tìm tất cả các trang web và hiển thị tất cả các thông tin mà bạn đang cần tìm, điều này không đúng. Ở đây các máy tìm kiếm (các trang có chức năng tìm kiếm) chỉ tìm kiếm trong một danh sách các website mà chúng lưu trữ, những website này có thể do máy tìm được trước đó hoặc do các website đó đăng ký với máy tìm kiếm. Do vậy, kết quả mà bạn tìm được bằng những trang web tìm kiếm khác nhau là khác nhau.

- *Không nên tìm kiếm theo một từ đơn*: thường thì bạn nên sử dụng hai từ hoặc nhiều hơn, hay một cụm từ ngắn. Những từ bạn chọn làm từ khóa phải là những từ liên quan trực tiếp tới vấn đề bạn đang cần tìm.

Ví dụ: Nếu bạn muốn tìm kiếm bài giảng về môn Tin học đại cương thì cụm từ bạn cần gõ là "bài giảng tin học đại cương"



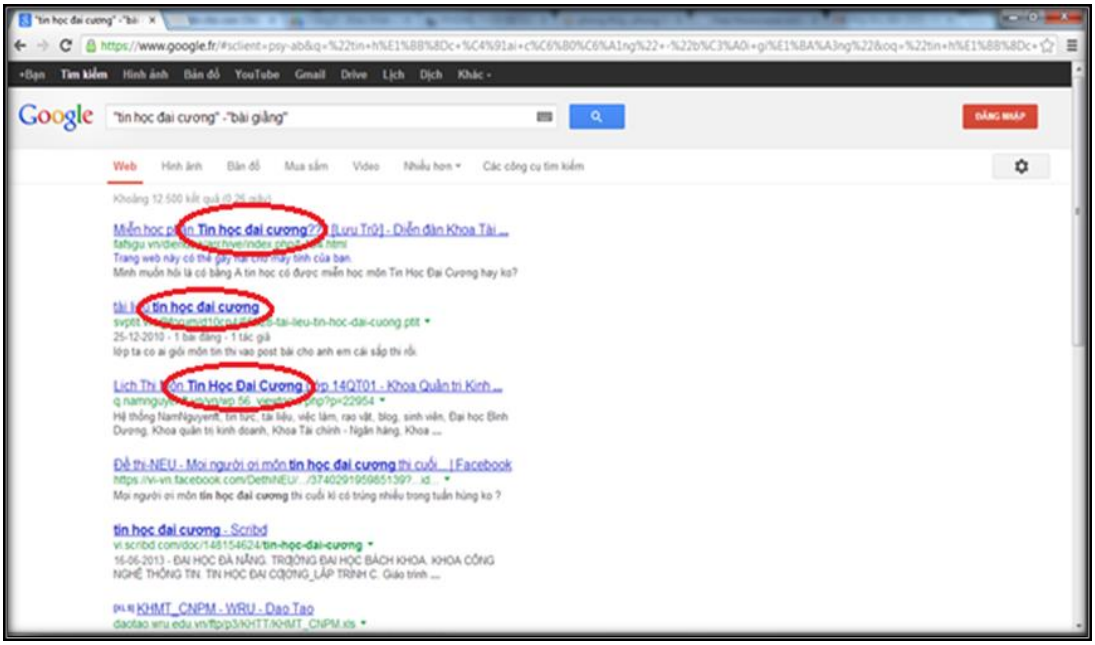
Hình 4.6. Cửa sổ kết quả tìm kiếm: bài giảng tin học đại cương

- *Sử dụng dấu + và "" để thu hẹp phạm vi tìm kiếm.*

Nhìn vào hình 4.6 chúng ta thấy máy tìm ra tất cả các trang có cả cụm "bài giảng tin học đại cương" hoặc có một số từ ở trong cụm từ như "tin học đại cương", "bài giảng"... Như vậy đây là kết quả mà chúng ta hoàn toàn không mong đợi máy tìm kiếm không tự nhận biết cả sáu từ bạn gõ vào là một từ khóa.

Để máy tìm kiếm tìm được chính xác cả cụm từ mong muốn, chúng ta hãy đặt cụm từ đó trong cặp dấu ngoặc kép (ở ví dụ trên là: "bài giảng tin học đại cương"). Ngoài ra, để kết hợp các từ khóa, chúng ta sử dụng dấu + để tìm kiếm, sao cho nội dung các trang tìm được vừa có từ khóa này, vừa có từ khóa kia.

- *Ghép thêm toán tử dấu (-)* vào trước một từ sẽ cho kết quả là cấm từ đó xuất hiện trong kết quả tìm kiếm.



Hình 4.7. Cửa sổ kết quả tìm kiếm "tin học đại cương"-"bài giảng"

Tìm kiếm nâng cao trong Google: Để thu hẹp kết quả web cho các tìm kiếm phức tạp bằng cách sử dụng trang Tìm kiếm nâng cao. Ví dụ: chúng ta muốn biết tình hình dịch sởi diễn ra tại Hà Nội trong 24 giờ qua. Hoặc tìm hình ảnh đen trắng của New York.

Thực hiện tìm kiếm nâng cao với Google (hình 4.8)

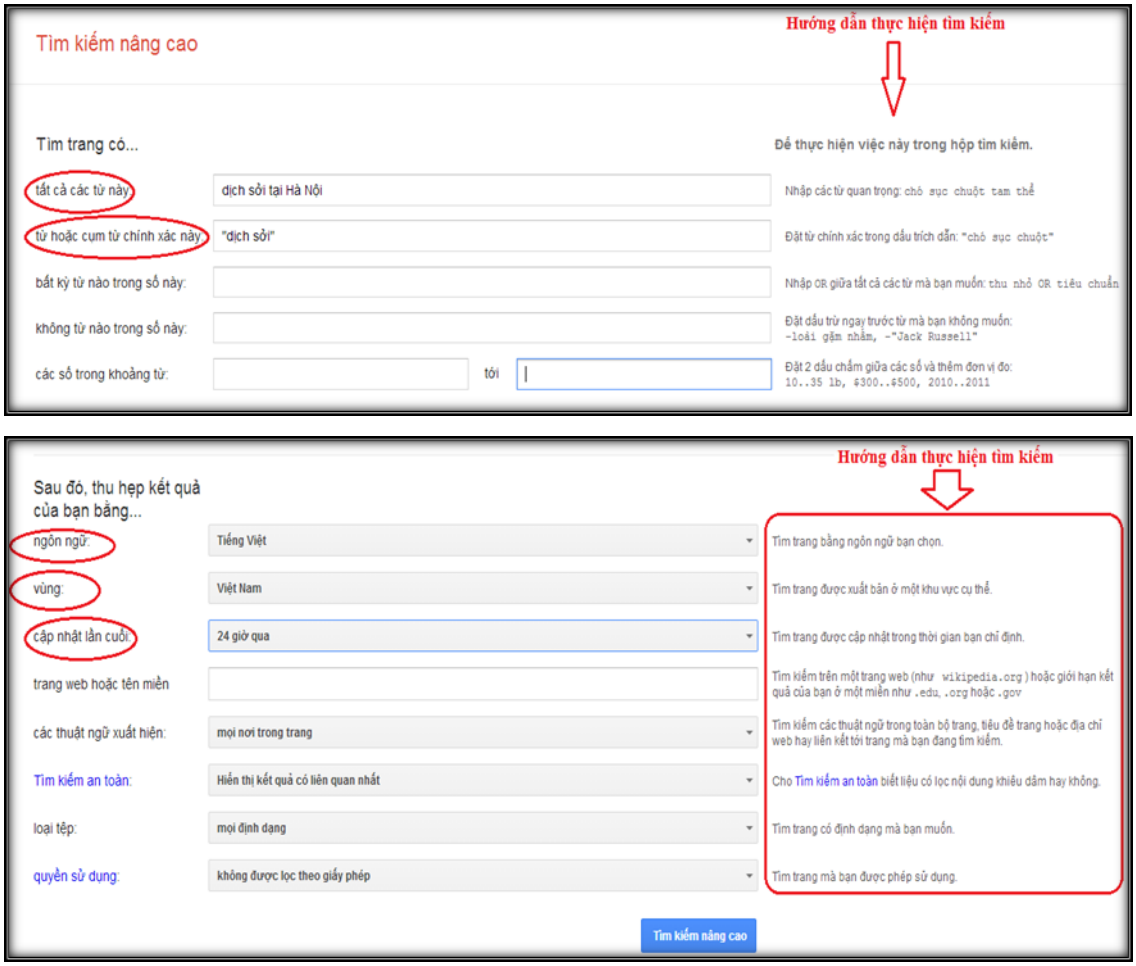
- Truy nhập vào trang Tìm kiếm nâng cao bằng cách

  - Truy cập trực tiếp vào www.google.com/advanced_search.
  - Nhấp vào biểu tượng bánh răng ⚙ ở góc trên cùng bên phải của trang kết quả tìm kiếm > Tìm kiếm nâng cao.

- Nhập cụm từ tìm kiếm của bạn vào phần "Tìm trang có".
- Chọn bộ lọc mà bạn muốn sử dụng trong phần "Sau đó, thu hẹp kết quả của bạn bằng". Bạn có thể sử dụng một hoặc nhiều bộ lọc (những bộ lọc này trong hộp tìm kiếm sử dụng toán tử tìm kiếm).
- Nhấp vào Tìm kiếm nâng cao.

Bộ lọc có thể sử dụng:

- Ngôn ngữ;
- Vùng;
- Ngày cập nhật lần cuối;
- Trang web hoặc miền;
- Nơi cụm từ tìm kiếm xuất hiện trên trang;
- Tìm kiếm an toàn;
- Cấp độ đọc;
- Loại tệp;
- Quyền sử dụng.

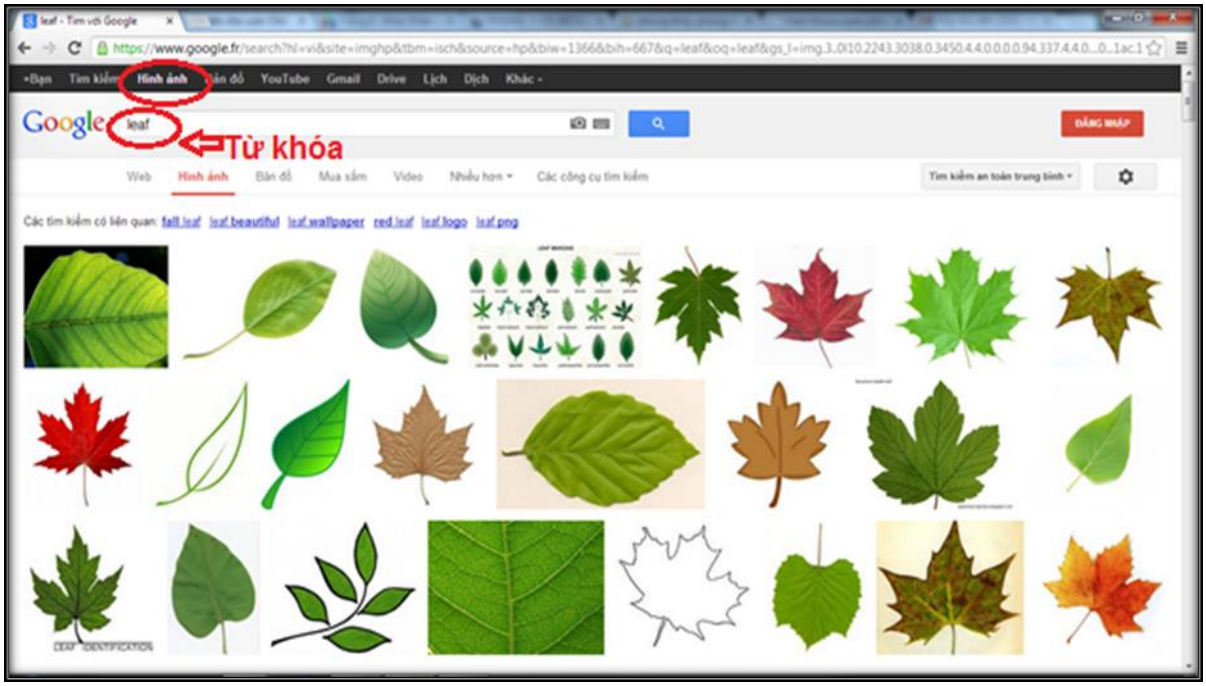


Hình 4.8. Tìm kiếm nâng cao trong Google

Với trang tìm kiếm nâng cao này, người dùng dễ dàng thu hẹp phạm vi tìm kiếm, làm cho việc tìm kiếm có kết quả nhanh hơn, chính xác hơn với mong muốn của người dùng.

Tìm kiếm hình ảnh

Tại trang chủ của google, nhấn chọn mục hình ảnh và gõ từ khóa tìm kiếm vào ô nhập, các hình ảnh sau khi tìm thấy có thể lưu vào trong máy tính của bạn.



Hình 4.9. Tìm kiếm hình ảnh với google

4.3.3. Thư điện tử

Thư điện tử (Email) là phương tiện liên lạc vô cùng tiện lợi trong thời đại công nghệ thông tin ngày nay. Sử dụng Email ta có thể trao đổi thông tin với bạn bè, đồng nghiệp trên toàn cầu. Ưu điểm nổi bật nhất của việc sử dụng Email là nhanh, rẻ, mọi lúc mọi nơi.

Phần này sẽ giới thiệu về một số khái niệm khi sử dụng Email, cách tạo và sử dụng Email. Muốn sử dụng thư điện tử thì người dùng phải có máy tính nối kết Internet hoặc nối kết vào máy chủ cung cấp dịch vụ Email (Mail Server). Ngoài ra, để gửi hoặc nhận Email thì người sử dụng phải có tài khoản Email và danh sách địa chỉ Email của người nhận.

Cấu trúc một địa chỉ Email

Địa chỉ Email (Email Address) là một định danh trên Internet cho phép người sử dụng Internet nhận biết được chính xác người cần liên hệ, giao dịch, trao đổi thông tin và ra lệnh gửi các thông điệp, tài liệu, hình ảnh (Email message) tới định danh này.

Cấu trúc một địa chỉ Email: **<Tên tài khoản>@<Tên miền>**

- Tên miền: Tên của máy tính làm Server lưu và quản lý địa chỉ Email này.

- Tên tài khoản: Tên được đăng ký, để phân biệt với các địa chỉ Email khác có cùng tên miền.

Ví dụ: mayxaydung@yahoo.com; ptthong@vnua.edu.vn; webmaster@vnua.edu.vn …

Cấu trúc một Email

Phần tiêu đề

From:	Phạm Quang Dũng <pqdung.hau1@gmail.com>
To:	Thang Tran <tranhuythang@gmail.com>
Cc:	Ngo Cong Thang <ncthang@hua.edu.vn>, Hoang Thi Ha <htha@hua.edu.vn>, ptvan <ptvan@hua.edu.vn>, phan thi thu hong <ptthong@hua.edu.vn>, Hoang Nguyen <startnewday85@gmail.com>
Date:	14/06/2013 09:57 AM
Subject:	Mẫu viết bài giảng tin đại cương chung cho 2 bộ môn

Phần nội dung

Tôi xin gửi các thầy cô template để sau viết bài giảng THĐC chung cho 2 BM.
PQDung.

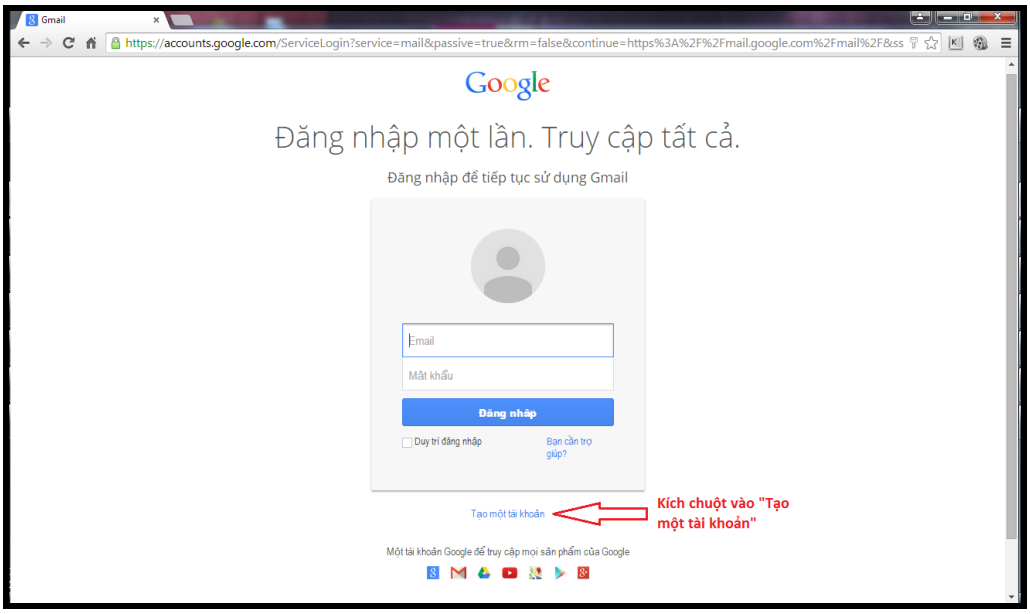
[2] File: Baigiang THDC.pptx Size: 105k Content Type:application/vnd.openxmlformats-officedocument.presentationml.presentation

Webmail

Webmail là hệ thống cung cấp các dịch vụ Email (nhận, gửi, lọc Email) thông qua một website nào đó trên mạng Internet. Thông thường, đây là hệ thống cung cấp địa chỉ Email miễn phí. Để gửi và nhận Email, người sử dụng Internet chỉ có một cách duy nhất là dùng trình duyệt Web truy cập vào địa chỉ Website của nhà cung cấp dịch vụ, sử dụng tài khoản đã được cung cấp để kiểm tra Email và sử dụng các dịch vụ Email thông thường khác. Ví dụ về các nhà cung cấp các dịch vụ Email: mail.vnua.edu.vn; www.hotmail.com; mail.yahoo.com; mail.google.com...

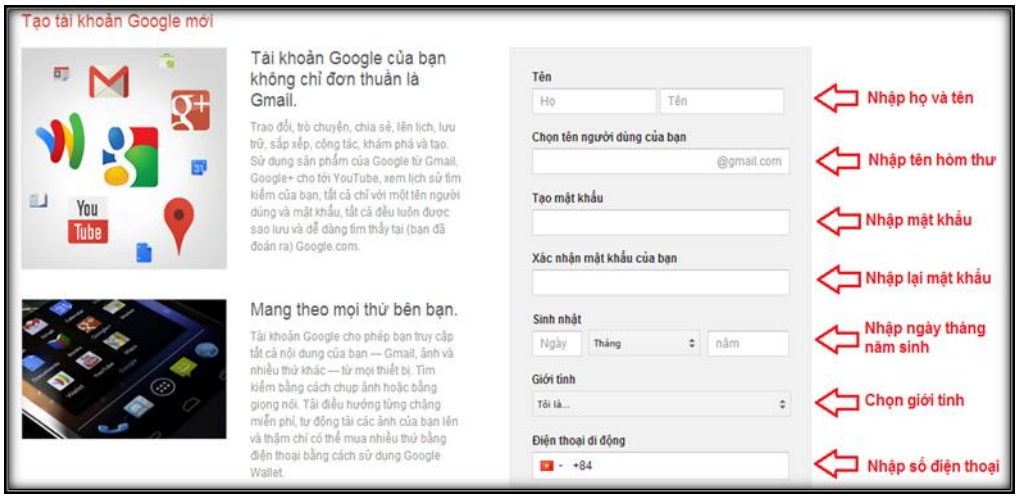
Cách tạo và sử dụng gmail

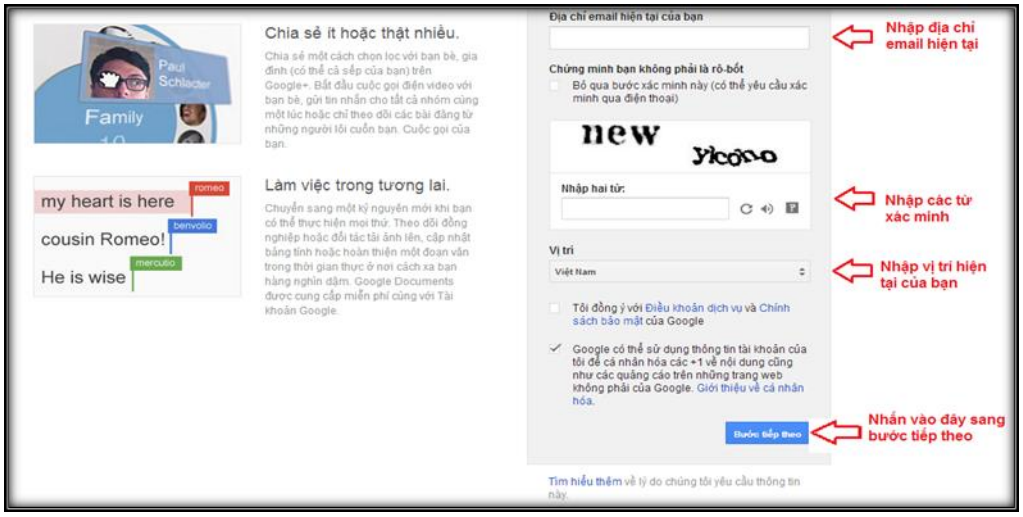
Phần này giới thiệu với các bạn về một trong những dịch vụ thư điện tử miễn phí tốt và phổ biến nhất hiện nay của Google. Để sử dụng gmail, người dùng phải đăng kí tài khoản với dịch vụ gmail. Từ trình duyệt nhập vào địa chỉ https://mail.google.com. Từ trang chủ gmail, kích vào "Tạo một tài khoản" (hình 4.10)



Hình 4.10. Trang chủ gmail

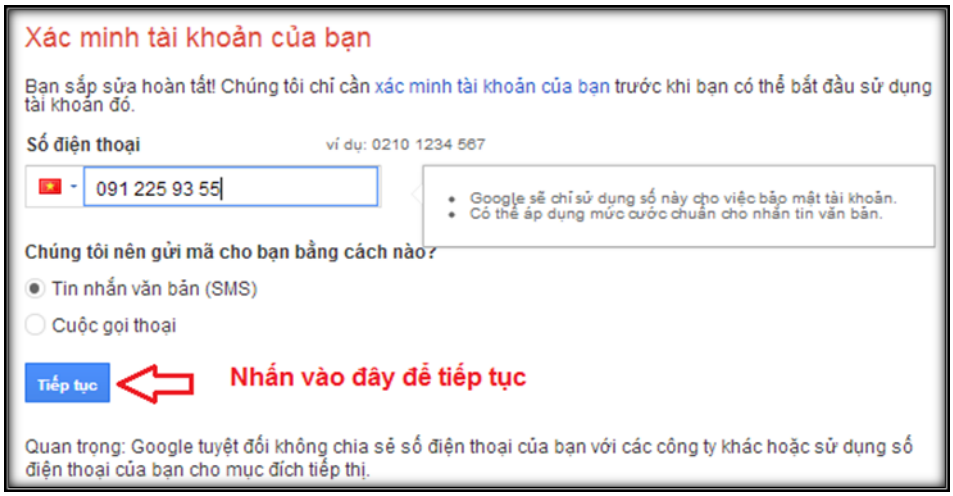
Sau đó điền đầy đủ thông tin như chỉ dẫn vào form (hình 4.11)





Hình 4.11. Form điền thông tin đăng kí tài khoản gmail

Kích vào "Bước tiếp theo" (hình 4.12) để tiếp tục thực hiện đăng kí. Tại bước này, người dùng cần nhập số điện thoại cá nhân để xác minh tài khoản.



Hình 4.12. Xác minh tài khoản qua điện thoại

Kích vào "Tiếp tục" để tiếp tục thực hiện việc đăng kí. Bạn đợi Google gửi lại mã số xác minh tài khoản qua số điện thoại đã cung cấp ở bước trước, sau đó nhập mã xác minh. Kết thúc bước này bạn đã hoàn tất quá trình đăng kí.

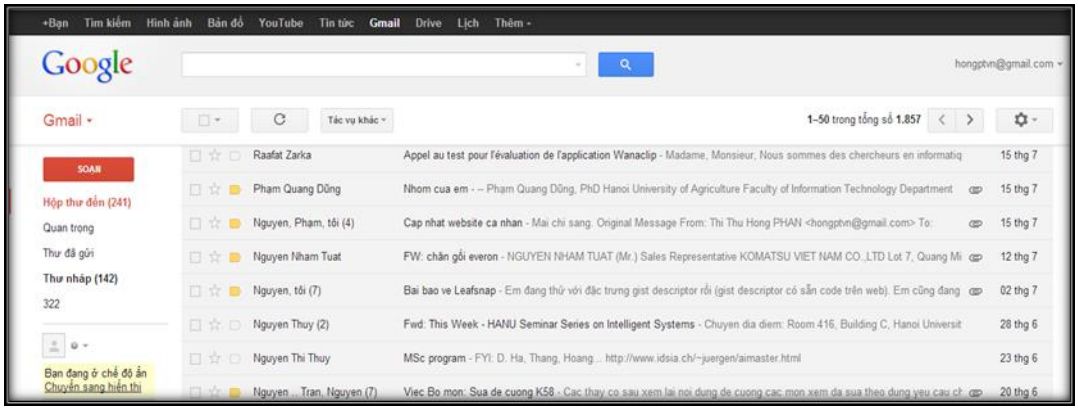
Hình 4.13. Nhập mã xác minh qua tin nhắn SMS

Sử dụng gmail

Gõ địa chỉ https://mail.google.com từ trình duyệt web. Tại mục đăng nhập:

- Nhập tên người dùng (Username).
- Mật khẩu (Password).
- Kích vào nút Đăng nhập.

Kết quả sau khi đăng nhập (hình 4.14):



Hình 4.14. Kết quả sau khi đăng nhập

- Chọn **Hộp thư đến** để xem thư đến
- Chọn **Thư đã gửi** để xem thư đã gửi đi
- Soạn thảo email mới

Kích vào nút "**Soạn**" để soạn thảo một email mới, xuất hiện trang soạn thảo. Nhập các yêu cầu như hình 4.15.

Kích vào "**Gửi**" để gửi thư đi sau khi đã hoàn thành thao tác soạn thảo.

Hình 4.15. Cửa sổ soạn thư mới

4.3.4. Lưu trữ dữ liệu đám mây

Lưu trữ dữ liệu đám mây là một dịch vụ lưu trữ (hay sao lưu - backup) dữ liệu ở các thiết bị mà người dùng không biết được địa chỉ thực. Về bản chất đó chính là những dịch vụ trực tuyến cung cấp giải pháp giúp người dùng cất giữ các loại dữ liệu của họ lên "đám mây", tức hệ thống máy chủ của nhà cung cấp dịch vụ. Người dùng có thể truy cập (explore), tải lên (upload),

tải xuống (download), đồng bộ hóa (sync) dữ liệu của họ từ bất cứ đâu có Internet, thông qua nhiều thiết bị có khả năng kết nối Internet.

Với dịch vụ lưu trữ dữ liệu đám mây, ta thấy:

Ưu điểm: Không còn lệ thuộc vào các thiết bị lưu trữ vật lý như đĩa nhớ, CD... và có thể truy cập ở mọi nơi có Internet. Giảm thiểu rủi ro mất mát, hư hỏng dữ liệu có thể xảy ra khi sử dụng các biện pháp lưu trữ truyền thống, như: sử dụng ổ cứng, ổ cứng di động USB (thuật ngữ được dùng để chỉ các ổ cứng hay thanh nhớ di động, hay được gọi tắt là ổ USB), hay đĩa CD, DVD.

Nhược điểm: Phụ thuộc vào Internet và nhà cung cấp dịch vụ lưu trữ đám mây.

Có hai đặc điểm rất quan trọng tạo nên sự khác biệt của lưu trữ dữ liệu đám mây so với các loại lưu trữ khác:

- Với các loại lưu trữ khác, người dùng hoàn toàn biết chính xác nơi lưu trữ dữ liệu như trong ổ USB hay ổ cứng, đĩa DVD... Còn với lưu trữ dữ liệu đám mây người dùng không thể biết chính xác được địa chỉ thực của các máy dịch vụ.

- Người dùng dịch vụ lưu trữ dữ liệu đám mây có thể đồng bộ hóa (sync) tất cả các dữ liệu ở bất kỳ nơi đâu có Internet.

Hiện nay có rất nhiều nhà cung cấp dịch vụ lưu trữ dữ liệu đám mây với các chính sách về giá cả và chế độ bảo mật khác nhau. Về giá, phần lớn cung cấp các gói miễn phí bên cạnh các dịch vụ có phí đi kèm như tăng dung lượng lưu trữ, chương trình truy cập dữ liệu trên máy tính hay thiết bị cầm tay... Về bảo mật, có rất nhiều dịch vụ không đảm bảo (hoặc không qui định rõ) chế độ bảo mật thông tin cho người dùng miễn phí. Một câu hỏi đặt ra là làm thế nào để *chọn được một dịch vụ lưu trữ dữ liệu đám mây phù hợp*? Một số tiêu chí được gợi ý sau đây giúp người dùng dễ dàng hơn trong việc ra quyết định lựa chọn của mình: Dung lượng cho phép lưu trữ miễn phí (Free Space); dung lượng tối đa cho phép đối với mỗi tệp tin được tải lên (Max file size); băng thông giới hạn tải lên và tải xuống trong mỗi tháng/ngày (Bandwidth limit/month/day); cho phép mã hóa dữ liệu để đảm bảo tính an toàn (Encrypt)…

Một số dịch vụ lưu trữ dữ liệu đám mây điển hình hiện nay:

a) Mediafire

MediaFire là một trang web chia sẻ dữ liệu miễn phí và không giới hạn. Tất cả các thành viên đăng ký đều có thể sử dụng đầy đủ chức năng của MediaFire.

Website bao gồm 4 loại dịch vụ gồm Free, Persernal, Pro và Business. Người dùng dịch vụ Free (miễn phí) được tải lên các tệp tin có dung lượng không quá 200 MB, sau đó được cung cấp một URL để tải xuống tệp tin và có thể chia sẻ cho bất kỳ ai. Thêm vào đó, ảnh số cũng có thể được tải lên và hiển thị dưới dạng thư viện. MediaFire phiên bản 3.0 với nhiều chức năng mới như: Không yêu cầu đăng ký, truy cập tệp tin không cần chờ đợi, không giới hạn thời gian lưu trữ, không giới hạn số tệp tin tải cùng một lúc, cho phép trực tiếp tải tệp tin lên một thư mục, cho phép đổi tên tệp tin, không phân biệt quốc gia, vùng lãnh thổ của người dùng.

b) Google Drive

Google Drive là dịch vụ lưu trữ trực tuyến được Google ra mắt vào đầu tháng 5 năm 2012, cho phép người dùng dễ dàng tải lên, chia sẻ và đồng bộ hóa dữ liệu lên dịch vụ này. Người dùng có thể sử dụng Google Drive để lưu trữ tất cả các loại tệp, bao gồm tài liệu, bản trình bày, nhạc, ảnh và video. Người dùng có thể mở nhiều loại tệp ngay trong trình duyệt, bao gồm tệp PDF, tệp Microsoft Office, video có độ phân giải cao và nhiều loại tệp hình ảnh.

Google Drive tự động cập nhật mọi nội dung, do đó người dùng có thể thực hiện các chỉnh sửa và truy cập phiên bản mới nhất từ bất kỳ nơi nào. Google Drive cung cấp nhiều cách để xem, tìm và sắp xếp các tệp; có các tùy chọn tìm kiếm mạnh mẽ - thậm chí là khả năng tìm kiếm văn bản trong hình ảnh, giúp người dùng nhanh chóng tìm thấy nội dung đang tìm kiếm.

Google Drive cho phép:

- Lưu trữ miễn phí 15GB nội dung thư, tài liệu, hình ảnh, video.
- Dung lượng tối đa cho phép đối với mỗi tệp tin được tải lên là 1GB.
- Truy cập mọi thứ trong Google Drive từ nhiều thiết bị khác nhau như: máy tính, điện thoại thông minh hoặc máy tính bảng.
- Đồng bộ hóa các tệp. Khả năng đồng bộ hóa cài sẵn của Google Drive đảm bảo các tệp, thư mục và tài liệu Google giống nhau trên tất cả thiết bị, chỉ cần kết nối web.
- Ngừng gửi tệp đính kèm qua email. Google Drive cho phép người dùng chọn lựa chính xác những người - bạn bè, gia đình, đồng nghiệp - nhận được tệp cần chia sẻ.

c) SkyDrive

SkyDrive là dịch vụ lưu trữ dữ liệu trực tuyến miễn phí của Microsoft. Ứng dụng này có sẵn cho hệ điều hành Windows và Mac, Windows Phone và iPhone.

Về cơ bản, dịch vụ này hoạt động tương tự như các dịch vụ lưu trữ dữ liệu trực tuyến khác. Người dùng Windows sau khi cài đặt ứng dụng này vào máy tính của mình chỉ cần đăng nhập vào tài khoản Windows Live của họ để liên kết máy tính đang sử dụng với dịch vụ lưu trữ trực tuyến này.

SkyDrive cho phép:

- Lưu trữ miễn phí 7GB nội dung.
- Dung lượng tối đa cho phép đối với mỗi tệp tin được tải lên là 2GB.
- Tạo ra một thư mục riêng của nó trong máy tính của người dùng.
- Đồng bộ hóa ngay lập tức và hoàn toàn tự động tất cả các tệp tin và thư mục được lưu trong SkyDrive.
- Lưu mọi tệp tin có trong SkyDrive của người dùng trong thư mục SkyDrive trên máy tính cũng như được lưu trên đám mây.

Phiên bản dành cho Windows của dịch vụ này cho phép người dùng truy cập vào tất cả các tệp tin trên máy tính (của người dùng) từ giao diện web của SkyDrive. Chỉ cần chọn một

máy tính đang bật có trong danh sách các máy tính được liên kết với tài khoản SkyDrive của người dùng trên giao diện web, sau đó điền đoạn mã bí mật được hệ thống gửi đến tài khoản email của người dùng; người dùng sẽ có toàn quyền xem tất cả các tệp tin có trên máy tính vừa chọn ngay trên giao diện web của dịch vụ lưu trữ dữ liệu trực tuyến này.

d) Dropbox

Dropbox là dịch vụ sao lưu, lưu trữ dữ liệu trực tuyến với khả năng đồng bộ theo thời gian thực và tự động thực hiện sao lưu. Phần mềm này chạy tốt trên Windows, Mac, Linux, đồng thời có thể sử dụng DropBox trên điện thoại như iPhone, Blackberry, Android, iPad...

Dropbox cho phép:

- Lưu trữ miễn phí 2GB dữ liệu.
- Dung lượng tối đa cho phép đối với mỗi tệp tin được tải lên là 300MB.
- Không giới hạn băng thông tải về và khả năng làm việc trên nhiều nền tảng từ máy tính đến thiết bị di động.
- Tự động đồng bộ hóa các tệp tin (tất cả các tệp tin khi được lưu trữ vào thư mục My Dropbox trên máy tính cá nhân sẽ được tự động đồng bộ hóa với các máy tính khác có cài đặt sẵn Dropbox).
- Chia sẻ tệp tin một cách dễ dàng. Người dùng đặt các tệp tin/thư mục muốn chia sẻ vào trong Dropbox rồi sau đó, gửi e-mail mời những người muốn chia sẻ thư mục đó.
- Sao lưu trực tuyến là tự động. Bất kì các tệp tin mà người dùng đặt vào trong thư mục Dropbox sẽ được tự động sao lưu vào các máy chủ.
- Dropbox cho phép người dùng quay trở lại quá khứ để khôi phục những dữ liệu bị xóa hoặc bị thay đổi. Dropbox giữ 30 ngày lịch sử của những lần thay đổi do người dùng thực hiện để người dùng có thể hoàn tác lại bất kỳ sai lầm và thậm chí lấy lại được các tệp tin đã bị xóa.

CÂU HỎI ÔN TẬP

1: Hãy trình bày khái niệm mạng máy tính, cách phân loại mạng máy tính.

2: Dựa theo kĩ thuật truyền tải thông tin có mấy kiểu mô hình kết nối? Trình bày những mô hình kết nối đó.

3: Thế nào là giao thức mạng máy tính?

4: Mục đích của dịch vụ DNS – phân giải tên miền. Cấu trúc của tên miền?

5: Điều kiện để kết nối Internet là gì?

6: Thế nào là mạng Internet? Các dịch vụ cơ bản của Internet là gì?

7: Hãy trình bày dịch vụ tìm kiếm trên Internet.

8: Hãy trình bày cấu trúc của một địa chỉ email; cấu trúc của một email?

9: Thế nào là dịch vụ lưu trữ đám mây? Trình bày ưu nhược điểm của dịch vụ này?

10: Hãy so sánh một số dịch vụ lưu trữ đám mây điển hình (Mediafire, Google Drive, SkyDrive, Dropbox).

Chương 5

CƠ SỞ DỮ LIỆU

Chương này giới thiệu những kiến thức cơ bản liên quan đến cơ sở dữ liệu, hệ quản trị cơ sở dữ liệu, ngôn ngữ truy vấn dữ liệu SQL. Từ đó, giúp sinh viên hiểu được: cơ sở dữ liệu là gì, sự cần thiết của việc tổ chức dữ liệu dưới dạng cơ sở dữ liệu, phần mềm dùng để tạo lập và quản trị cơ cở dữ liệu, ngôn ngữ truy vấn dữ liệu SQL.

5.1. CƠ SỞ DỮ LIỆU

5.1.1. Khái niệm cơ sở dữ liệu

Cơ sở dữ liệu (CSDL, thuật ngữ tiếng Anh là *database*) là một tập hợp các dữ liệu có liên quan với nhau chứa thông tin về một tổ chức nào đó (như một trường đại học, một ngân hàng, một bệnh viện, một công ty…) được lưu trữ trên các thiết bị nhớ thứ cấp (như băng từ, đĩa từ…) để đáp ứng nhu cầu khai thác thông tin của nhiều người sử dụng với nhiều mục đích khác nhau.

Ví dụ 5.1: Bài toán Quản lý sinh viên đơn giản, ta có thể dùng một cơ sở dữ liệu lưu trữ thông tin về sinh viên và kết quả học tập của họ bao gồm 5 bảng (ví dụ được tạo bởi phần mềm Microsoft Access) như sau:

Bảng **KHOA**:

Khoa

MaKhoa	TenKhoa	SoDT
CNTT	Công nghệ thông tin	0462617701
TY	Thú ý	0436782978
KE	Kế toán	
CNSH	Công nghệ sinh học	
NH	Nông học	
CNTP	Công nghệ thực phẩm	
QLDD	Quản lý đất đai	0438674739

Bảng **LOP**:

Lop

MaLop	TenLop	MaK
K56CNSHA	Công nghệ sinh học A K56	CNSH
K57CNSHA	Công nghệ sinh học A K57	CNSH
K57CNTTA	Công nghệ thông tin A K57	CNTT
K57QLTT	Quản lý thông tin K57	CNTT
K58CNTTA	Công nghệ thông tin A K58	CNTT
K58CNTTB	Công nghệ thông tin B K58	CNTT
K58TYA	Thú A K58	TY
K58TYB	Thú Y B K58	TY
K59TYC	Thú Y C- K59	TY

Bảng **SINHVIEN**:

Sinhvien

Masv	Hodem	Ten	Ngaysinh	Gioitinh	Tinh	MaLop
561000	Lê Thị Minh	Thùy	04/06/1992	Nữ	Thanh Hóa	K56CNSHA
571233	Trần Trung	Hải	04/06/1993	Nam	Nam Định	K57CNSHA
573435	Lê Thị	Linh	23/06/1993	Nữ	Hà Nội	K57CNSHA
582004	Trần Trung	Kiên	05/07/1994	Nam	Hà Nội	K57CNTTA
583456	Lê Thu	Hoài	03/05/1994	Nữ	Hà Nội	K58CNTTA
586786	Nguyễn Thị	Huyền	04/06/1994	Nữ	Thanh Hóa	K58CNTTA

Bảng **MONHOC**:

MonHoc

MaMH	TenMH	DVHT	DieuKien
TH1009	Tin học đại cương	2	
TH1010	Tin học ứng dụng	2	Tin học đại cương
TH2010	Cơ sở dữ liệu	3	Tin học đại cương
TH2021	Hệ quản trị CSDL	3	Cơ sở dữ liệu
TH3020	Lập trình mạng	2	Tin học đại cương

Bảng **KETQUA**:

Ketqua

MaSV	MaMH	Diem
561000	TH1009	7
561000	TH2010	3
561000	TH2021	5
561000	TH3020	7
571233	TH1010	6
573435	TH1009	7
573435	TH1010	8
573435	TH2021	8
573435	TH3020	9

5.1.2. Các mức thể hiện của cơ sở dữ liệu

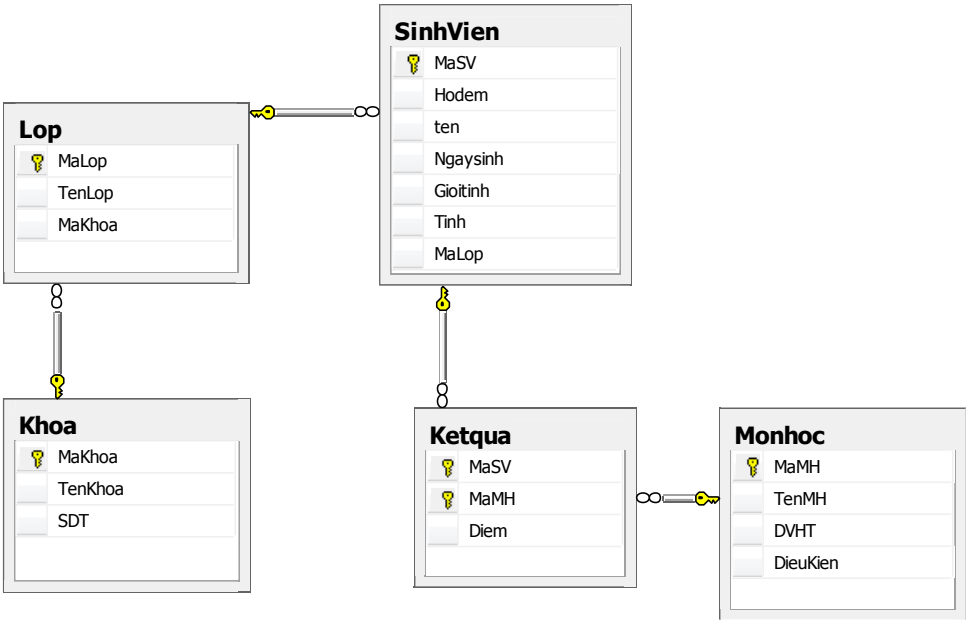
Vì mỗi nhóm người dùng có vai trò, nhu cầu hiểu và khai thác thông tin khác nhau nên để quản lý thông tin một cách hiệu quả, các hệ CSDL phải có các mức thể hiện khác nhau. Có 3 mức thể hiện CSDL là mức vật lý, mức khái niệm, mức khung nhìn.

a. Mức vật lý

Những chuyên gia tin học cần hiểu chi tiết về cách lưu trữ dữ liệu trong bộ nhớ, chẳng hạn: các tệp dữ liệu được lưu trữ trong vùng nhớ nào? mỗi bản ghi chiếm bao nhiêu byte? Mức hiểu biết chi tiết về một CSDL như vậy gọi là mức vật lý của hệ CSDL đó.

b. Mức khái niệm

Nhóm phát triển các ứng dụng không cần hiểu chi tiết ở mức vật lý, nhưng họ cần phải biết những dữ liệu nào được lưu giữ trong CSDL, giữa các dữ liệu có mối quan hệ như thế nào. Mức hiểu CSDL như vậy được gọi là mức khái niệm.

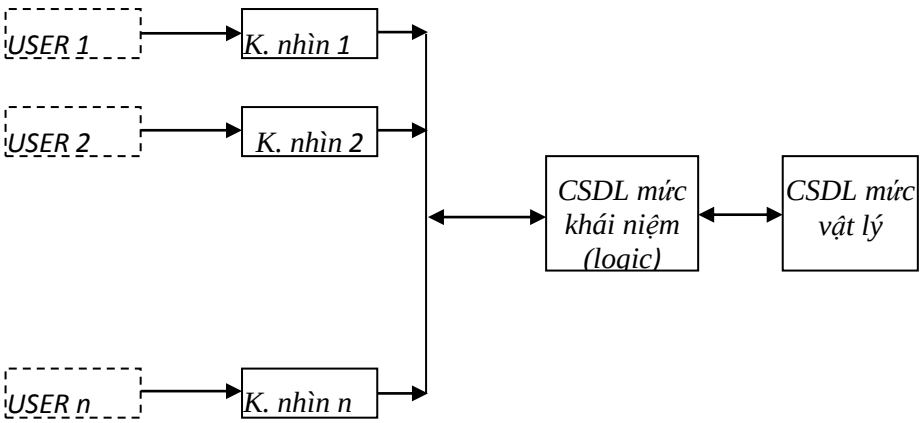


Hình 5.1. Ví dụ về mức khái niệm của CSDL

c. Mức khung nhìn

Mỗi nhóm người dùng chỉ cần biết phần thông tin nào đó của CSDL phù hợp với nghiệp vụ hay mục đích sử dụng của mình. Ví dụ, sinh viên thông qua khung nhìn biết được những thông tin liên quan đến bản thân họ. Người quản trị CSDL cần biết được toàn bộ thông tin về CSDL... Vì vậy, mức khung nhìn là mức hiểu CSDL của người dùng thông qua khung nhìn.

Ba mức hiểu về CSDL như trên chính là ba mức mô tả và làm việc với CSDL, phù hợp với nhu cầu khác nhau của những người liên quan đến CSDL.



Hình 5.2. Ba mức thể hiện của CSDL

5.1.3. Mô hình dữ liệu quan hệ

Mô hình dữ liệu (data model) là một tập các khái niệm và kí pháp dùng để mô tả dữ liệu, các mối quan hệ của dữ liệu, các ràng buộc trên dữ liệu của một tổ chức.

Hiện nay, có khá nhiều mô hình dữ liệu như:

- Mô hình dữ liệu quan hệ (Relational Data Model).
- Mô hình dữ liệu phân cấp (Hierarchical Data Model).
- Mô hình dữ liệu mạng (Network Data Model).
- Mô hình dữ liệu thực thể - liên kết (Entity-Relationship Data Model).
- Mô hình dữ liệu hướng đối tượng (Object Data Model).

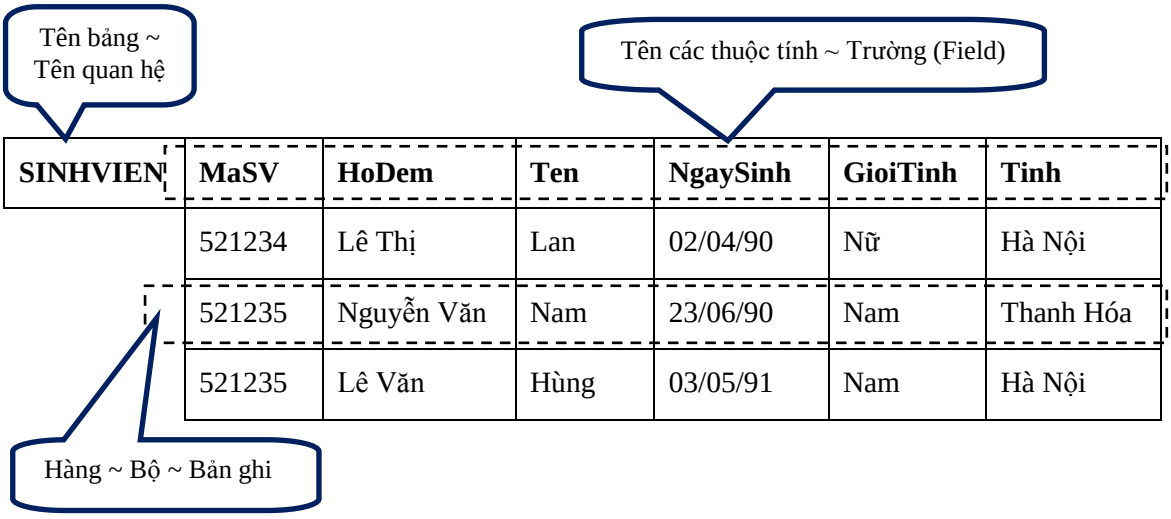
Trong các mô hình trên thì mô hình dữ liệu quan hệ được sử dụng khá phổ biến. Mô hình này được đề xuất bởi E. F. Codd vào những năm 1970 - 1972. Nó cung cấp một cấu trúc dữ liệu đơn giản đó là quan hệ (bảng).

Cơ sở dữ liệu được xây dựng trên mô hình dữ liệu quan hệ được gọi là CSDL quan hệ. Một CSDL quan hệ thông thường chứa nhiều bảng. Mỗi bảng chứa dữ liệu của một tập thực thể, bao gồm các hàng và các cột. Mỗi hàng là một bản ghi (Record), mỗi cột là một trường (Field).

a. Một số khái niệm cơ bản trong mô hình dữ liệu quan hệ

* *Quan hệ:* Dữ liệu lưu trữ trong CSDL được tổ chức thành bảng 2 chiều. Mỗi bảng 2 chiều được gọi là một quan hệ.

  Dưới đây là ví dụ của một quan hệ:



SINHVIEN	MaSV	HoDem	Ten	NgaySinh	GioiTinh	Tinh
	521234	Lê Thị	Lan	02/04/90	Nữ	Hà Nội
	521235	Nguyễn Văn	Nam	23/06/90	Nam	Thanh Hóa
	521235	Lê Văn	Hùng	03/05/91	Nam	Hà Nội

Hình 5.3. Ví dụ về quan hệ SINHVIEN

* *Lược đồ (schema)*

Tên của một quan hệ và tập các thuộc tính của nó được gọi là một lược đồ đối với quan hệ đó. Ta biểu diễn lược đồ cho một quan hệ bởi **Tên** của quan hệ và theo sau là danh sách các thuộc tính của nó. Vậy lược đồ của quan hệ SINHVIEN trong hình 5.3 là:

SINHVIEN(MaSV, HoDem, Ten, NgaySinh, GioiTinh, Tinh)

Lược đồ cơ sở dữ liệu quan hệ (relational database schema) là tập các lược đồ quan hệ của bài toán.

Ví dụ 5.2: Bài toán quản lý sinh viên trong ví dụ 5.1 có lược đồ CSDL bao gồm 5 lược đồ quan hệ sau:

KHOA(MaKhoa, TenKhoa, SoDT)

LOP(MaLop, TenLop, MaKhoa)

SINHVIEN(MaSV, HoDem, Ten, NgaySinh, GioiTinh, Tinh, MaLop)

MONHOC(MaMH, TenMH, DVHT, Dieukien)

KETQUA(MaSV, MaMH, Diem)

* *Bộ* (*tuble)*

Bộ là dòng của một quan hệ, trừ dòng tiêu đề (tên của các thuộc tính). Bộ còn có cách gọi khác là bản ghi (record). Trong một quan hệ các bộ không được trùng nhau.

* *Miền (domain)*

Miền là tập các giá trị mà thuộc tính có thể nhận.

Ví dụ, miền của thuộc tính Gioitinh (giới tính) trong ví dụ 5.2 gồm hai giá trị {Nam, Nữ}.

* *Khóa (key*, còn gọi là *khóa chính)*

Khóa của một quan hệ là một hoặc nhiều thuộc tính tối thiểu để xác định tính duy nhất của mỗi bộ trong quan hệ đó

Ví dụ 5.3: Trong quan hệ SINHVIEN ở trên, dễ hiểu là MaSV của mỗi sinh viên là duy nhất, không thể có 2 mã sinh viên trùng nhau. Vậy MaSV được thiết lập là khóa.

Ví dụ 5.4: Trong quan hệ KETQUA(MaSV, MaMH, Diem) ở trên, vì một sinh viên có thể học nhiều môn học nên để ghi điểm các môn của sinh viên đó vào bảng thì các bộ có cùng MaSV, khác nhau MaMH. Tương tự, một môn học có thể được học bởi nhiều sinh viên, nên các bộ có thể trùng MaMH, khác nhau MaSV (xem bảng KETQUA ở trên). Nhưng cặp MaSV, MaMH không thể trùng nhau để xác định duy nhất một điểm môn học của sinh viên.

Chú ý: Một quan hệ có thể có nhiều khóa, khi đó mỗi một khóa được gọi là một khóa dự tuyển. Thông thường có một khóa dự tuyển được chỉ định làm khóa chính. Việc lựa chọn một khóa dự tuyển làm khóa chính là tùy ý, nhưng nên chọn khóa dự tuyển đặc trưng cho bộ và chỉ gồm một thuộc tính hoặc có ít thuộc tính nhất làm khóa chính.

Ví dụ 5.5: Lược đồ quan hệ KHOA(MaKhoa, TenKhoa, SoDT) có hai khoá ứng cử là: K_1= {MaKhoa}, K_2 ={TenKhoa}, tuy nhiên ta chọn MaKhoa làm khoá chính vì nó đặc trưng cho Khoa hơn và các giá trị của thuộc tính này ngắn, không có dấu và không có khoảng trống.

* *Khóa ngoại (foreign key)*

Khóa ngoại (khóa ngoài) của một lược đồ quan hệ là một tập gồm một hay nhiều thuộc tính không phải là khóa chính của lược đồ quan hệ này nhưng lại là khóa chính của một lược đồ quan hệ khác.

Khoá ngoại dùng để biểu thị liên kết giữa quan hệ này và quan hệ khác trong mô hình quan hệ.

Ví dụ 5.6: Xét lược đồ CSDL trong ví dụ 5.2. Ta thấy, trong lược đồ quan hệ LOP có MaKhoa là khoá ngoại (vì nó là khoá chính trong lược đồ quan hệ KHOA nhưng không phải là khoá chính của lược đồ quan hệ LOP). Khi đã tạo mối quan hệ (Relationship) hợp lý giữa trường MaKhoa trong bảng KHOA với trường MaKhoa trong bảng LOP thì ta có thể truy vấn các thông tin liên quan giữa 2 quan hệ này như lớp này thuộc khoa nào, có tên khoa là gì. Những nội dung như này sẽ được minh họa rõ ràng hơn ở phần sau và khi thực hành trên máy tính.

Một ví dụ tương tự trong lược đồ quan hệ SINHVIEN có MaLop là khoá ngoại (vì nó là khoá chính trong lược đồ quan hệ LOP nhưng không phải là khoá chính của lược đồ quan hệ SINHVIEN).

Lưu ý là khóa ngoại không xác định tính duy nhất của bộ dữ liệu như khóa chính. Với ví dụ này:

- Trong quan hệ LOP, thuộc tính MaKhoa của các bộ có thể trùng nhau (tương đương một khoa có nhiều lớp); nhưng trong quan hệ KHOA thì thuộc tính MaKhoa của mỗi bộ phải là duy nhất.
- Trong quan hệ SINHVIEN, thuộc tính MaLop của các bộ có thể trùng nhau (tương đương một lớp có nhiều sinh viên); nhưng trong quan hệ LOP thì thuộc tính MaLop của mỗi bộ phải là duy nhất.

5.1.4. Hệ cơ sở dữ liệu

Hệ CSDL (Database system) là một hệ thống gồm 4 thành phần:

- Cơ sở dữ liệu.
- Người sử dụng cơ sở dữ liệu: Là bất cứ một người nào đó có thể truy nhập (hợp pháp) vào CSDL, có nghĩa là bao gồm tất cả những người sử dụng cuối, những người viết chương trình ứng dụng và những người điều khiển toàn bộ hệ thống (còn gọi là người quản trị CSDL).
- Hệ quản trị CSDL (*Phần 5.2*).
- Phần cứng: Bao gồm các thiết bị nhớ thứ cấp được sử dụng để lưu trữ CSDL.

5.1.5. Lợi ích của hệ cơ sở dữ liệu

Trước khi các hệ CSDL ra đời (khoảng đầu những năm 60) là giai đoạn tiền xử lý cơ sở dữ liệu: Dữ liệu được tổ chức và xử lý bởi các tệp ghi trên các băng từ. Các ngôn ngữ lập trình như COBOL, BASIC được sử dụng để lập trình xử lý dữ liệu. Mỗi chương trình ứng dụng đều có một tệp dữ liệu tương ứng và mỗi khi chương trình ứng dụng cần được sửa đổi hoặc mở rộng thì tệp dữ liệu tương ứng cũng phải thay đổi theo. Cách tổ chức lưu trữ như vậy sẽ bị *dư thừa dữ liệu, dữ liệu không nhất quán, khó khăn trong việc truy cập và chia sẻ dữ liệu, dữ liệu không được bảo mật cao*... Việc sử dụng hệ CSDL để lưu trữ dữ liệu theo lý thuyết cơ sở dữ liệu sẽ khắc phục được những hạn chế của cách lưu trữ trên, cụ thể có những ưu điểm sau:

- *Giảm bớt dư thừa dữ liệu trong lưu trữ:* Trong các ứng dụng lập trình truyền thống, phương pháp tổ chức lưu trữ dữ liệu vừa tốn kém, dư thừa thông tin và lãng phí bộ nhớ.

Nhiều chương trình ứng dụng khác nhau cùng xử lý trên các dữ liệu như nhau nhưng lại không dùng chung dữ liệu, dẫn đến sự dư thừa đáng kể về dữ liệu.

- *Tránh được sự không nhất quán trong lưu trữ dữ liệu và bảo đảm được tính toàn vẹn của dữ liệu:* Nếu một thuộc tính được mô tả trong nhiều tệp dữ liệu khác nhau và các bản ghi bị lặp lại nhiều lần thì khi thực hiện việc cập nhật, sửa đổi, bổ sung sẽ không sửa hết nội dung các mục đó. Nếu dữ liệu càng nhiều thì sự sai sót khi cập nhật, bổ sung càng lớn. Khả năng xuất hiện mâu thuẫn, không nhất quán thông tin càng nhiều, dẫn đến không nhất quán dữ liệu trong lưu trữ. Tất yếu kéo theo sự dị thường thông tin, thừa, thiếu và mâu thuẫn thông tin. Nếu một thuộc tính của một đối tượng chỉ được lưu trữ một lần trong một CSDL thì sẽ đảm bảo được tính toàn vẹn và nhất quán của dữ liệu.
- *Có thể triển khai đồng thời nhiều ứng dụng trên cùng một CSDL:* Điều này có nghĩa là trên cùng một CSDL có thể triển khai đồng thời nhiều ứng dụng khác nhau tại các thiết bị đầu cuối khác nhau.
- *Thống nhất các tiêu chuẩn, thủ tục và các biện pháp bảo vệ, an toàn dữ liệu:* Các CSDL sẽ được quản lý tập trung bởi một người hay một nhóm người quản trị CSDL. Người quản trị CSDL có thể áp dụng thống nhất các tiêu chuẩn, quy định, thủ tục chung như quy định thống nhất về mẫu biểu báo cáo, thời gian bổ sung, cập nhật dữ liệu. Nhờ đó công việc bảo trì dữ liệu trở nên dễ dàng. Người quản trị CSDL có thể bảo đảm việc truy nhập tới CSDL, có thể kiểm tra, kiểm soát các quyền truy nhập của người sử dụng, có thể cho phép nhiều người truy nhập đồng thời mà vẫn đảm bảo tính đúng đắn của dữ liệu. Người quản trị CSDL có thể cho phép mỗi người dùng của hệ CSDL chỉ được phép truy cập một phần CSDL, điều đó cũng là một biện pháp giữ cho dữ liệu trong CSDL được an toàn. Ví dụ, trong hệ thống quản lý học tập theo tín chỉ trên mạng, các sinh viên của trường chỉ nhìn thấy một phần CSDL chứa thông tin về sinh viên đó thông qua tài khoản họ được cấp chứ không nhìn thấy các thông tin của sinh viên khác.

Như vậy, việc tổ chức lưu trữ dữ liệu trong CSDL giúp người dùng quản lý dữ liệu tốt hơn và thông qua các hệ quản trị CSDL ta có thể thực hiện các nhiệm vụ quan trọng như: tổng hợp, sắp xếp, tìm kiếm, thêm, xóa, sửa... và khai thác dữ liệu.

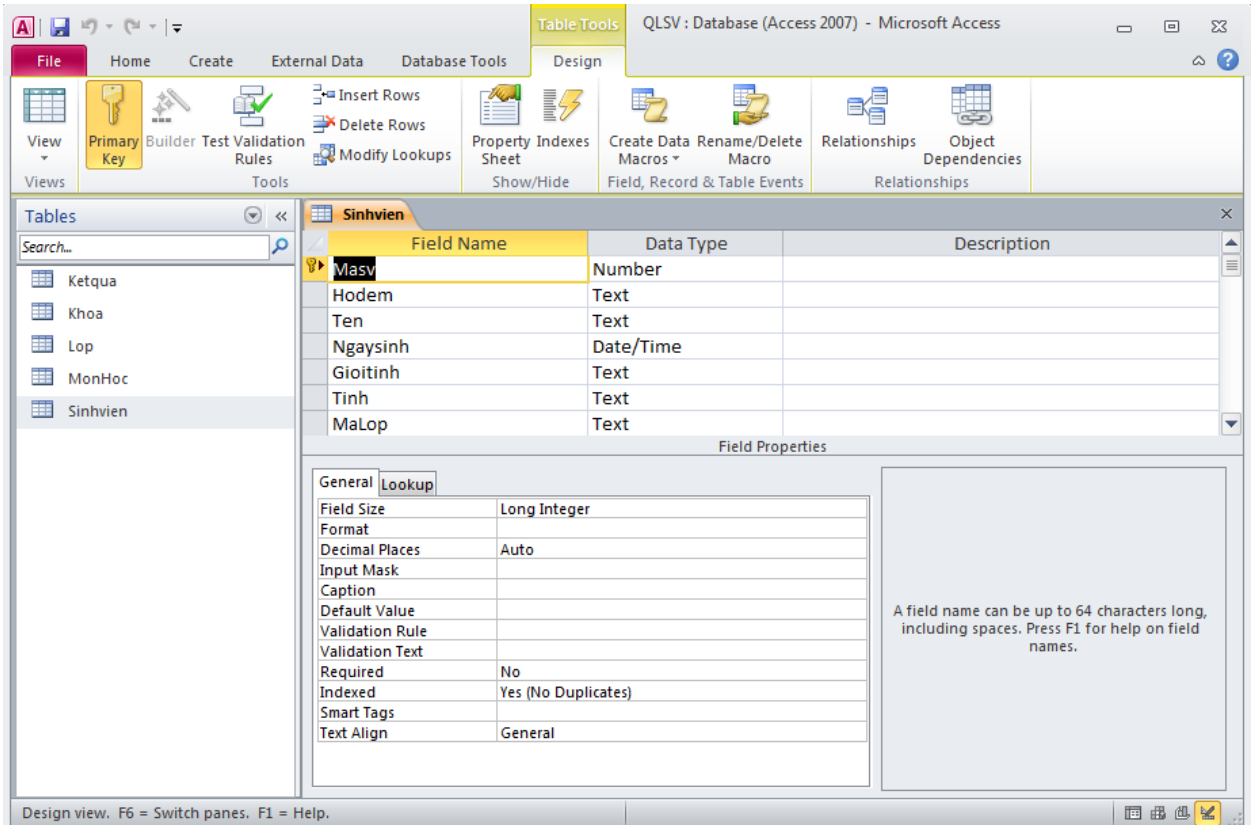
5.2. HỆ QUẢN TRỊ CƠ SỞ DỮ LIỆU

5.2.1. Khái niệm

Hệ quản trị cơ sở dữ liệu (HQTCSDL, thuật ngữ tiếng Anh là Database Management System - DBMS) là phần mềm được thiết kế để thuận lợi việc tạo lập, lưu trữ và khai thác thông tin của CSDL.

Như vậy, HQTCSDL cung cấp một môi trường thuận lợi, đơn giản và hiệu quả để người sử dụng có thể tạo lập, lưu trữ và thao tác trên CSDL mà không cần quan tâm nhiều đến thuật toán chi tiết và cách biểu diễn dữ liệu trong bộ nhớ.

Dưới đây là giao diện của việc dùng HQTCSDL Microsoft Access để tạo lập cơ sở dữ liệu:



Hình 5.4. Giao diện thiết kế các bảng trong HQTCSDL Microsoft Access

5.2.2. Phân loại hệ quản trị cơ sở dữ liệu

a. Phân loại

Các hệ quản trị cơ sở dữ liệu được phân thành 3 loại: XML DBMS, ODBMS và RDBMS.

- ✓ *XML DBMS:* Phù hợp cho dữ liệu đã được định dạng XML (eXtensible Markup Language).
- ✓ *ODBMS (object database management system):* Phù hợp cho mô hình CSDL hướng đối tượng.
- ✓ *RDBMS (relational database management system):* Phù hợp cho mô hình cơ sở dữ liệu quan hệ. Ngày nay, các HQTCSDL này đều có tính năng thao tác trên dữ liệu XML và các lớp đối tượng.

Thông thường, mỗi một HQTCSDL được xây dựng để phục vụ cho một mô hình dữ liệu nhất định, nhưng cũng có một số HQTCSDL có thể phục vụ cho nhiều mô hình dữ liệu khác nhau.

b. Một số hệ quản trị cơ sở dữ liệu phổ biến

Ngày nay, có một số HQTCSDL phổ biến như: Microsoft Access, Microsoft SQL Server, MySQL, SQLite, Oracle.

- DB2 là sản phẩm của IBM thuộc dòng HQTCSDL quan hệ, phiên bản đầu tiên ra đời năm 1982. Đây là HQTCSDL được dùng cho tất cả các máy tính, từ máy tính cá nhân đến những dòng máy tính lớn.
- Microsoft SQL Server và Microsoft Access là sản phẩm của Microsoft, thuộc loại HQTCSDL quan hệ.
- MySQL là HQTCSDL nguồn mở đa luồng, đa người dùng, được hãng MySQL sản xuất. MySQL rất phổ biến với những ứng dụng web và có thể làm việc với các CSDL trên nền Linux/Mac/Windows.
- SQLite là một hệ quản trị cơ sở dữ liệu có đặc điểm của gọn, nhẹ, đơn giản. Chương trình gồm 1 file duy nhất có dung lượng chưa đến 500kB, không cần cài đặt mà có thể sử dụng ngay. CSDL được lưu ở một file duy nhất, chạy tốt trên platform Mac OS-X, Android, iOS.

5.2.3. Chức năng cơ bản của hệ quản trị cơ sở dữ liệu

Một hệ quản trị cơ sở dữ liệu có các chức năng cơ bản sau đây:

* Cung cấp môi trường tạo lập CSDL
* Cung cấp môi trường cập nhật và khai thác dữ liệu
    - *Cập nhật*: Thêm, xóa, sửa dữ liệu.
    - *Khai thác*: Sắp xếp, tìm kiếm, kết xuất báo cáo...
* Cung cấp công cụ kiểm soát, điều khiển CSDL
    - Phát hiện và ngăn chặn sự truy cập không được phép.
    - Duy trì tính nhất quán của dữ liệu.
    - Tổ chức và điều khiển các truy nhập đồng thời.
    - Khôi phục CSDL khi có sự cố ở phần cứng hay phần mềm.
    - Quản lý các mô tả dữ liệu.

Nói chung các HQTCSDL đều có các chức năng trên, nhưng các HQTCSDL khác nhau có chất lượng và khả năng khác nhau khi đáp ứng các nhu cầu thực tế. Các HQTCSDL luôn phát triển theo hướng đáp ứng đòi hỏi ngày càng cao của người dùng, bởi vậy các chức năng của chúng ngày càng được mở rộng.

5.3. NGÔN NGỮ TRUY VẤN SQL

SQL (*Structured Query Language* – ngôn ngữ truy vấn có cấu trúc) là ngôn ngữ truy vấn dựa trên đại số quan hệ được xác nhận là rất mạnh, phổ dụng và dễ sử dụng. SQL được sử dụng hầu hết trong các thao tác: truy vấn, thêm, xóa, sửa trên các hệ quản trị CSDL quan hệ. SQL được thiết lập như một ngôn ngữ chuẩn đối với cơ sở dữ liệu.

Ngôn ngữ SQL gồm các thành phần:

- Ngôn ngữ định nghĩa dữ liệu (Data Definition Language - DDL): Cung cấp các câu lệnh cho phép định nghĩa các lược đồ quan hệ, các ràng buộc toàn vẹn mà dữ liệu được lưu trữ trong CSDL phải thỏa mãn, cho phép xóa, sửa cấu trúc các quan hệ.

- Ngôn ngữ thao tác dữ liệu (Data Manipulation Language - DML): Đây là nhóm câu lệnh cho phép thao tác trên các dữ liệu của quan hệ. Nó dùng để tìm kiếm, trích rút, tổng hợp dữ liệu từ các quan hệ, đồng thời cho phép thêm, xóa, sửa dữ liệu trong các quan hệ

- Nhóm ngôn ngữ kiểm soát dữ liệu (Data Control Language - DCL): Bao gồm các câu lệnh đảm bảo tính an toàn và toàn vẹn dữ liệu, cấp phát quyền truy cập vào dữ liệu.

Trong nội dung giáo trình Tin học đại cương này chủ yếu giới thiệu nhóm ngôn ngữ thao tác dữ liệu cơ bản.

Nhóm ngôn ngữ SQL thao tác dữ liệu bao gồm các câu lệnh cho phép thao tác trên dữ liệu của CSDL. Nó dùng để thực hiện các truy vấn như: tìm kiếm, thêm, xóa, sửa các bản ghi.

- SELECT – trích dữ liệu từ một cơ sở dữ liệu
- INSERT – chèn dữ liệu mới vào trong một cơ sở dữ liệu
- DELETE – xóa dữ liệu từ một cơ sở dữ liệu
- UPDATE – cập nhật dữ liệu trong một cơ sở dữ liệu

5.3.1. Câu lệnh truy vấn dữ liệu

Loại câu lệnh này cho phép ta tìm kiếm, trích rút dữ liệu từ cơ sở dữ liệu. Kết quả của câu lệnh được hiển thị dưới dạng bảng 2 chiều.

Cú pháp:

```
SELECT [DISTINCT] <danh sách các cột>| *| <biểu thức>
FROM <bảng 1> [,bảng 2...]
[WHERE <điều kiện>]
[GROUP BY <danh sách tên cột> [HAVING <biểu thức điều kiện>]]
[ORDER BY <danh sách tên cột|<biểu thức> [ASC|DESC]]
```

Giải thích:

- DISTINCT là từ khoá để được một danh sách không có các bản ghi trùng nhau.

- Các thành phần mà người dùng phải điền cụ thể vào khi viết lệnh được viết trong cặp < >.

- Các thành phần tùy chọn (những thành phần có thể có hoặc không) được viết trong cặp [ ].

- Danh sách các cột là tên các cột cần truy vấn trong các bảng (chú ý nếu tên cột xuất hiện trên nhiều bảng thì phải chỉ rõ cột đó được tham chiếu qua bảng nào bằng cú pháp: **Tenbang.Tencot**)

- SELECT *: dùng để hiển thị tất cả các cột trong bảng.

- Bảng 1, bảng 2... là tên các bảng hoặc tên các khung nhìn dùng để truy vấn dữ liệu.

- Mệnh đề WHERE dùng để chỉ định một tiêu chuẩn chọn các bản ghi.

- <điều kiện> là một biểu thức logic có kết quả trả về là TRUE hoặc FALSE.

- GROUP BY <tên các cột> dùng để nhóm kết quả hiển thị theo từng loại giá trị của cột.

- Having là tiêu chuẩn chọn trên từng nhóm, được đặt sau GROUP BY.

- ORDER BY <danh sách tên cột> dùng để sắp xếp kết quả vừa chọn ở trên theo cột nào (ASC là tăng, DESC là giảm).

Biểu thức <điều kiện> có các toán tử sau có thể được dùng:

Bảng 5.1. Các toán tử trong SQL

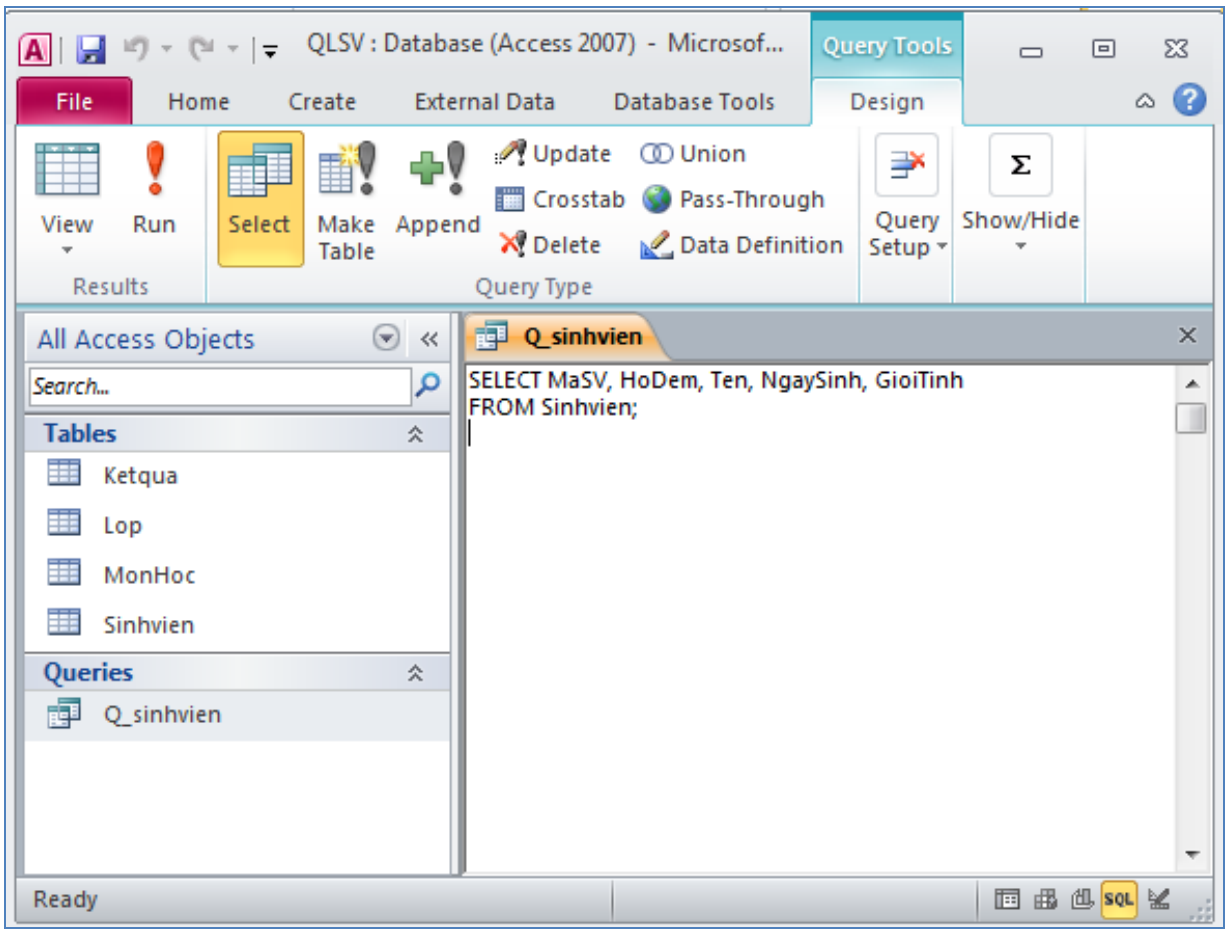
Phép toán	**Giải thích**
=	Bằng
<>	Khác
>	Lớn hơn
<	Nhỏ hơn
>=	Lớn hơn hoặc bằng
<=	Nhỏ hơn hoặc bằng
[NOT] Like	Phép toán tìm một mẫu kí tự
AND	Và
OR	Hoặc
BETWEEN <giá trị 1> AND <giá trị 2>	Chọn tất cả các trị trong khoảng giới hạn giữa hai giá trị. Các trị này có thể là các số, chuỗi kí tự, hay ngày tháng.
<biểu thức> [NOT] IN (danh sách|câu truy vấn])	Kiểm tra giá trị của biểu thức có trong tập danh sách các giá trị không
[NOT] EXISTS (<câu truy vấn>)	Kết quả trả về TRUE nếu câu truy vấn khác rỗng.
<biểu thức><phép so sánh> SOME|ALL|ANY <câu truy vấn>	Kết quả trả TRUE nếu phép so sánh thỏa mãn.

Ví dụ 5.7: Sử dụng CSDL trong ví dụ 5.1 để thực hiện các yêu cầu sau:

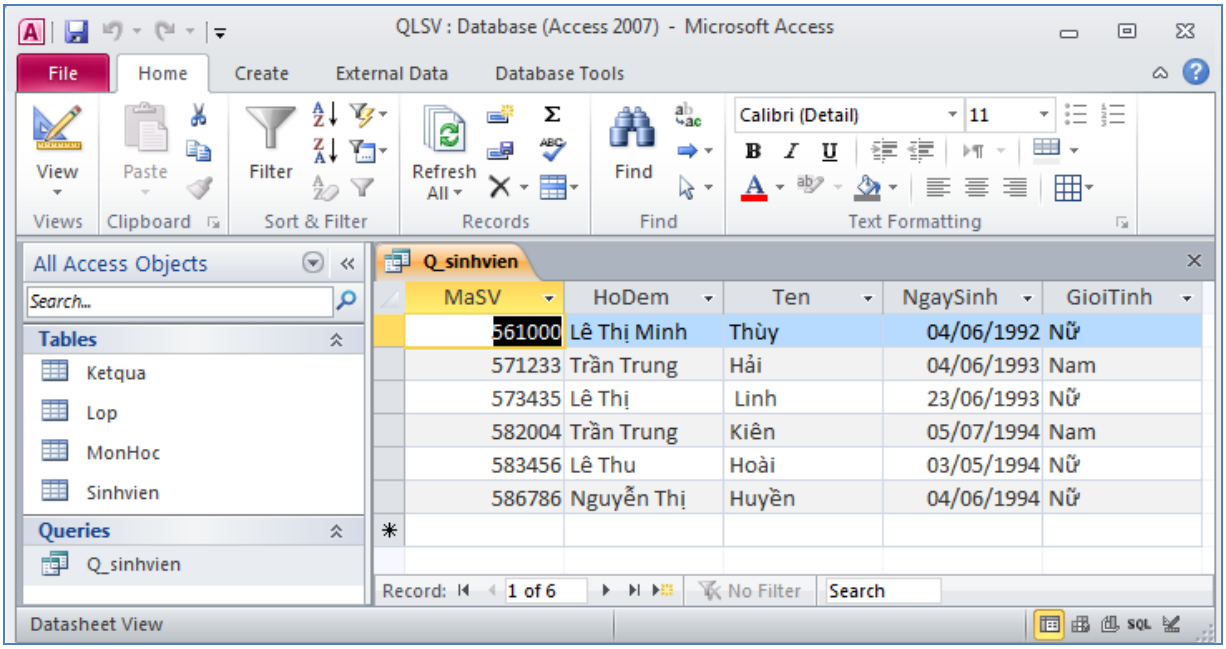
Câu lệnh SQL hiển thị thông tin về tất cả sinh viên, thông tin hiển thị cần: mã sinh viên, họ tên, ngày sinh, giới tính:

```
SELECT MaSV, HoDem, Ten, NgaySinh, GioiTinh
FROM Sinhvien;
```

Để chạy thử câu lệnh SQL, ta mở CSDL của bài toán trên đã được cài đặt trong một HQTCSDL, sau đó gõ câu lệnh SQL và chạy thử, màn hình sẽ hiển trị kết quả truy vấn.



Hình 5.5. Giao diện soạn thảo câu lệnh SQL trong HQTCSDL Microsoft Access



Hình 5.6. Kết quả thực hiện câu lệnh SQL hình 5.5 trong HQTCSDL Microsoft Access

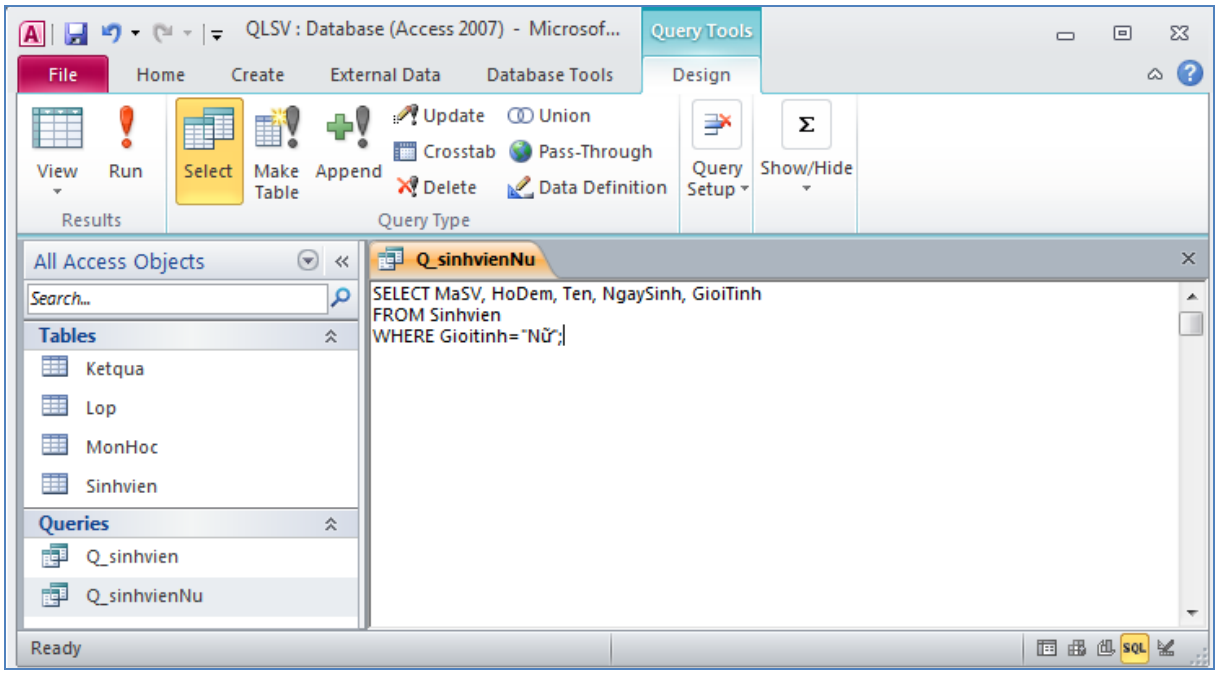
Câu lệnh SQL hiển thị thông tin về các sinh viên nữ, thông tin hiển thị cần: mã sinh viên, họ tên, ngày sinh, giới tính.

SELECT MaSV, HoDem, Ten, NgaySinh, GioiTinh

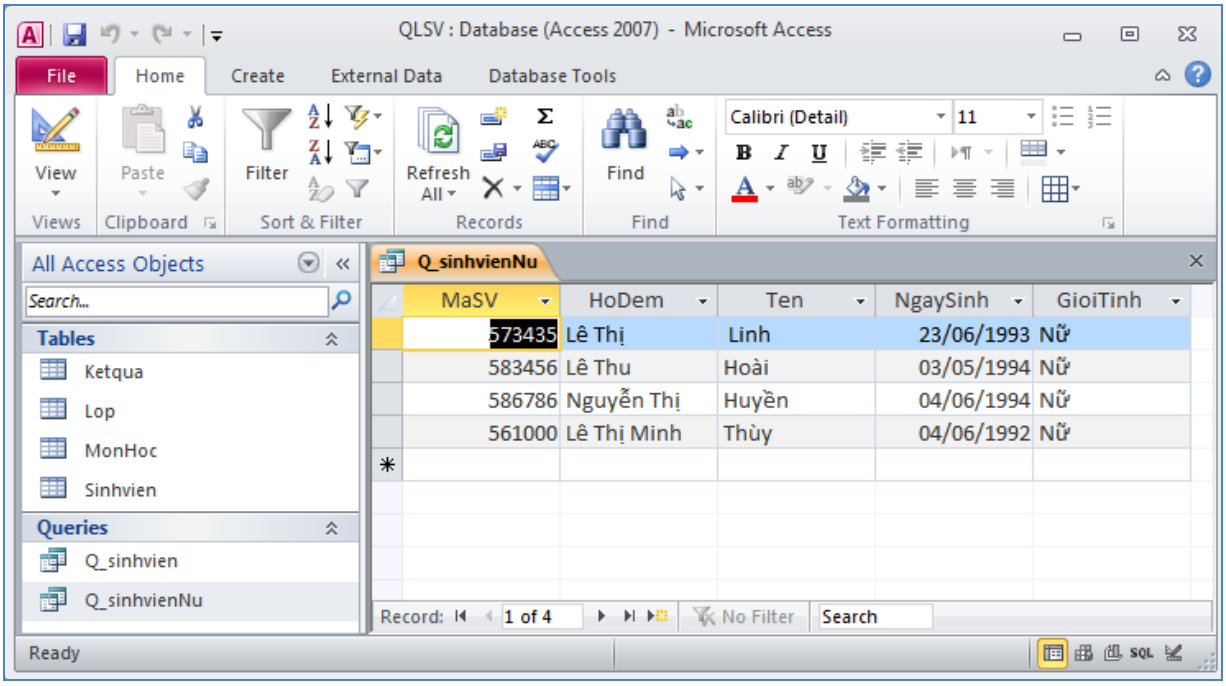
FROM Sinhvien

WHERE Gioitinh=“Nữ”;

Chạy thử câu lệnh SQL trong HQTCSDL Microsoft Access:



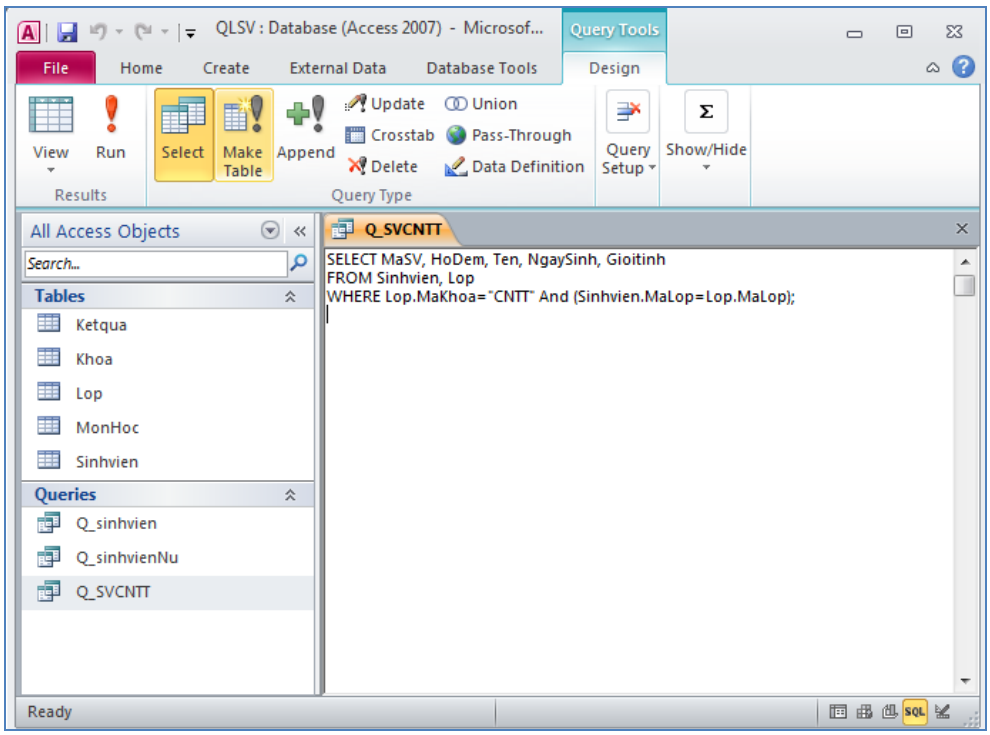
Hình 5.7. Giao diện soạn thảo câu lệnh SQL trong HQTCSDL Microsoft Access



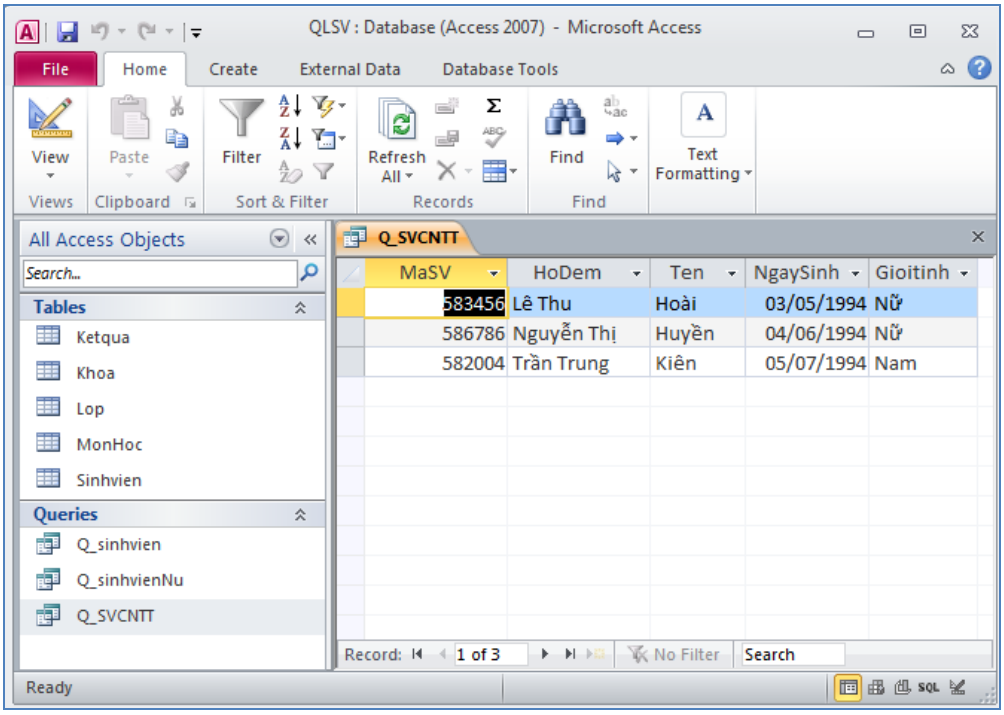
Hình 5.8. Kết quả thực hiện câu lệnh SQL hình 5.7 trong HQTCSDL Microsoft Access

Câu lệnh SQL hiển thị thông tin về các sinh viên khoa CNTT (có MaKhoa="CNTT"). Thông tin hiển thị cần: mã sinh viên, họ tên, ngày sinh, giới tính:

```
SELECT MaSV, HoDem, Ten, NgaySinh, GioiTinh
FROM Sinhvien, Lop
WHERE Lop.MaKhoa="CNTT" AND (Sinhvien.MaLop=Lop.MaLop);
```



Hình 5.9. Giao diện soạn thảo câu lệnh SQL trong HQTCSDL Microsoft Access

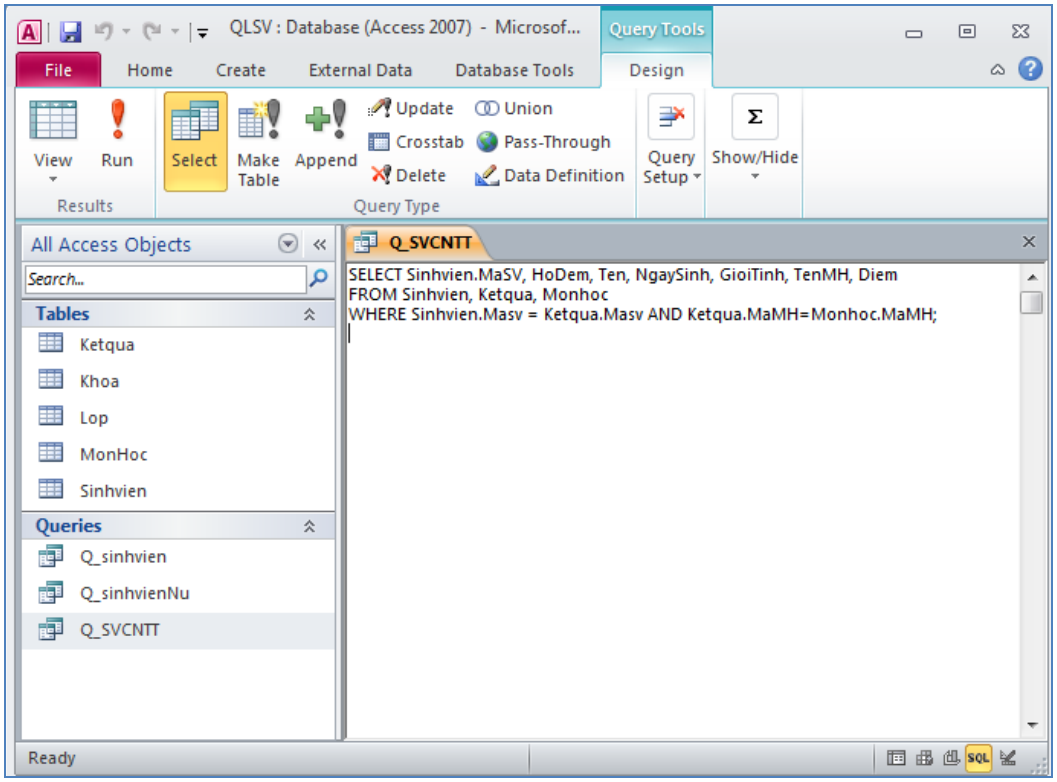


Hình 5.10. Kết quả thực hiện câu lệnh SQL hình 5.9 trong HQTCSDL Microsoft Access

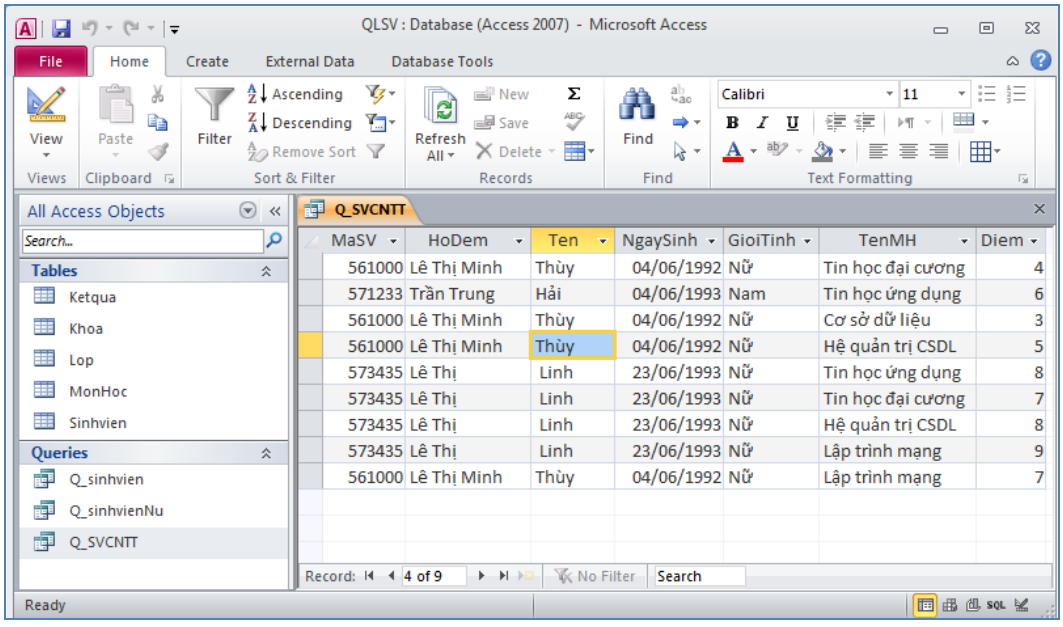
Câu lệnh SQL hiển thị thông tin về các sinh viên với các kết quả học tập của họ. Thông tin hiển thị cần: mã sinh viên, họ tên, ngày sinh, giới tính, tên môn học, điểm.

```
SELECT Sinhvien.MaSV, HoDem, Ten, NgaySinh, GioiTinh, TenMH, Diem
FROM Sinhvien, Ketqua, Monhoc
WHERE Sinhvien.MaSV = Ketqua.MaSV AND Ketqua.MaMH=Monhoc.MaMH;
```

Chạy câu lệnh SQL trong HQTCSDL Microsoft Access:



Hình 5.11. Giao diện soạn thảo câu lệnh SQL trong HQTCSDL Microsoft Access



Hình 5.12. Kết quả thực hiện câu lệnh SQL hình 5.11 trong HQTCSDL Microsoft Access

Câu lệnh SQL hiển thị thông tin về các sinh viên đạt điểm A học phần Tin học đại cương (MaMH="TH01009"). Thông tin hiển thị cần (mã sinh viên, họ tên, ngày sinh, tên môn học, điểm) và được sắp xếp theo vần alphabet của tên và họ (nếu trùng tên thì sắp xếp theo họ đệm):

```
SELECT Sinhvien.MaSV, HoDem, Ten, NgaySinh, TenMH, Diem
FROM Sinhvien, Ketqua, Monhoc
WHERE Sinhvien.MaSV=Ketqua.MaSV AND Ketqua.MaMH="TH1009" AND
Ketqua.MaMH=Monhoc.MaMH AND Diem>=8.5
ORDER BY Ten, Hodem;
```

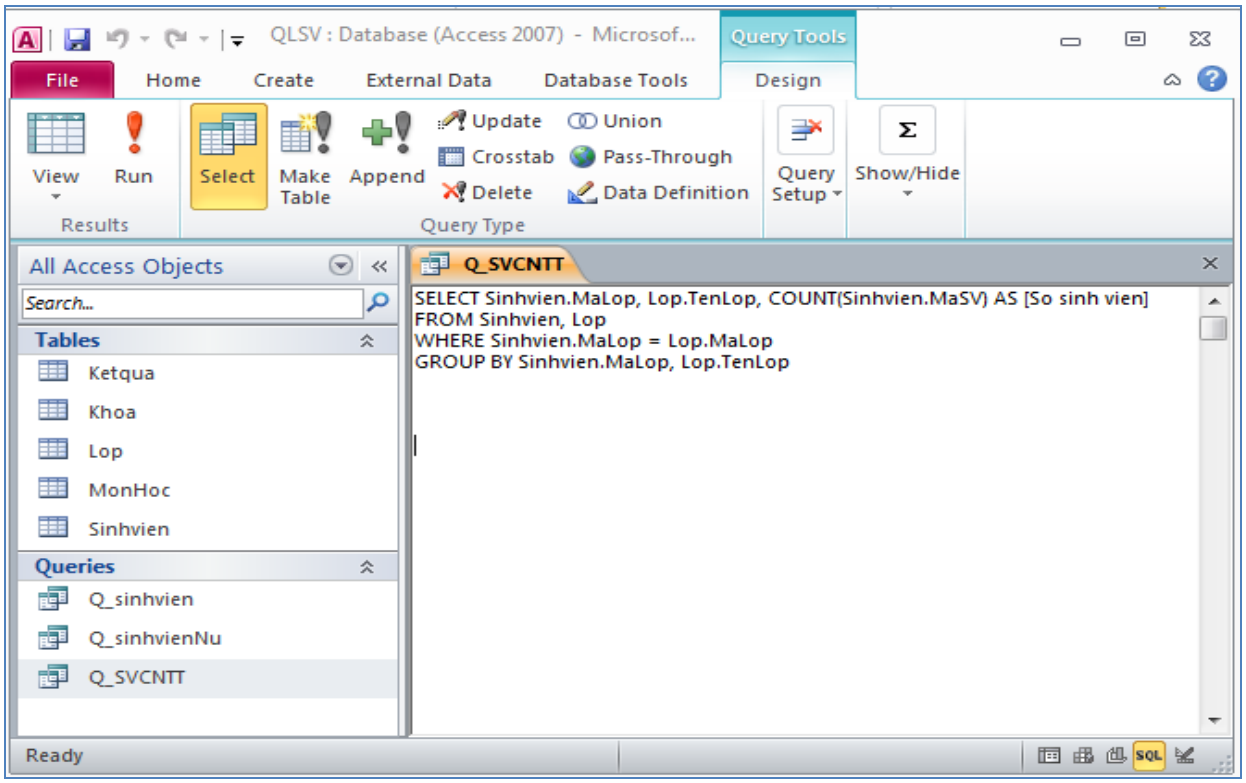
* Toán tử **GROUP BY**

Có thể phân hoạch các bộ của một quan hệ thành các nhóm tách biệt nhau và áp dụng các phép toán gộp cho các nhóm. Trong câu lệnh SELECT – FROM – WHERE, mệnh đề GROUP BY nhóm lại bởi một danh sách các thuộc tính của quan hệ cần nhóm.

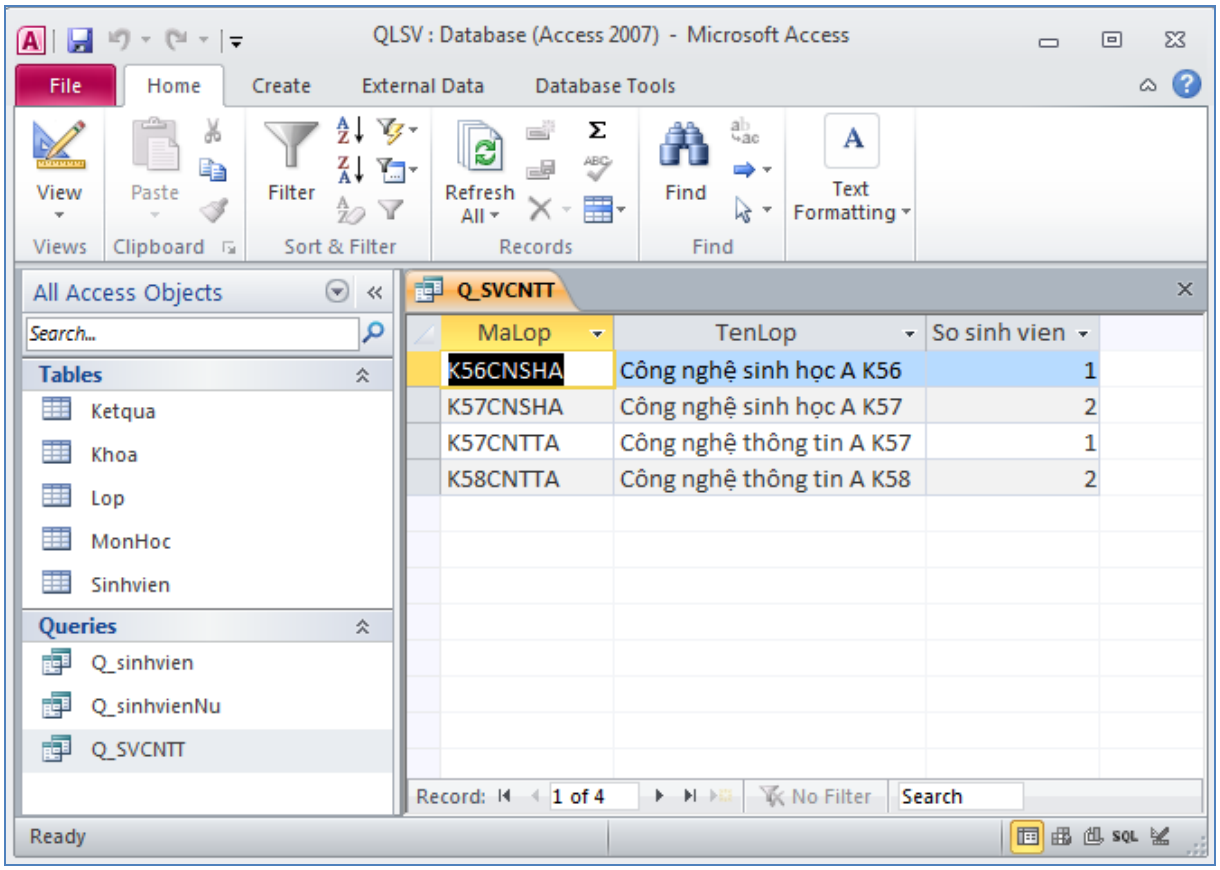
Ví dụ 5.8: In ra danh sách các lớp và số sinh viên trong mỗi lớp:

```
SELECT Sinhvien.MaLop, Lop.TenLop, COUNT(Sinhvien.MaSV) AS [So sinh vien]
FROM Sinhvien, Lop
WHERE Sinhvien.MaLop = Lop.MaLop
GROUP BY Sinhvien.MaLop, Lop.TenLop
```

Chạy câu lệnh SQL trong HQTCSDL Microsoft Access:



Hình 5.13. Giao diện soạn thảo câu lệnh SQL trong HQTCSDL Microsoft Access



Hình 5.14. Kết quả thực hiện câu lệnh SQL hình 5.13 trong HQTCSDL Microsoft Access

* Toán tử **GROUP BY … HAVING …**

Phân hoạch các bộ của một quan hệ thành các nhóm tách biệt nhau và áp dụng các phép toán gộp cho các nhóm. Trong câu lệnh SELECT – FROM – WHERE, mệnh đề GROUP BY nhóm lại bởi một danh sách các thuộc tính của quan hệ cần nhóm và thoả mãn một điều kiện nhóm HAVING:

GROUP BY $A_1, A_2, .., A_k$

HAVING E

Phân hoạch quan hệ thành các nhóm sao cho hai bộ cùng trong một nhóm khi và chỉ khi chúng giống nhau ở mọi thuộc tính $A_1, A_2 ... , A_k$. Để cho kết quả của câu vấn tin có nghĩa, các thuộc tính $A_1, A_2 ... , A_k$ cũng phải xuất hiện trong mệnh đề SELECT mặc dù chúng có thể có những bí danh để in ra nếu cần.

Ví dụ 5.9: In ra danh sách các lớp có số sinh viên ≥ 2. Thông tin hiển thị cần: Mã lớp, Tên lớp, Số sinh viên.

```
SELECT Sinhvien.MaLop, Lop.TenLop, COUNT(Sinhvien.MaSV) AS 'So sinh vien'
FROM Sinhvien, Lop
WHERE Sinhvien.MaLop = Lop.MaLop
GROUP BY Sinhvien.MaLop, Lop.TenLop
        HAVING COUNT(Sinhvien.MaSV)>=2;
```

∗ Toán tử LIKE: dùng chỉ định việc tìm gần đúng một xâu kí tự trong một cột.

Cú pháp:

SELECT <tên cột> FROM <tên bảng> WHERE <tên cột> LIKE <“xâu kí tự”>;

Một dấu "%" có thể dùng như ký tự đại diện cho một số kí tự

Ví dụ 5.10: Hiển thị thông tin về những sinh viên có tên bắt đầu bằng chữ “N”.

SELECT * FROM Sinhvien WHERE Ten LIKE “N%”;

∗ Toán tử BETWEEN <giá trị 1> AND <giá trị 2>: chọn tất cả các trị trong khoảng giới hạn giữa hai giá trị. Các giá trị này có thể là các số, chuỗi kí tự, hay ngày tháng.

Ví dụ 5.11: Hiển thị thông tin về những sinh viên có ngày sinh trong khoảng 01/01/93 đến 31/12/94. Lưu ý là cần chuyển chuỗi ngày tháng trong câu lệnh thành dạng tháng trước ngày sau.

SELECT * FROM Sinhvien

WHERE Ngaysinh BETWEEN #01/01/93# AND #12/31/94#;

∗ Từ khóa DISTINCT dùng để trả về chỉ các giá trị khác biệt (distinct).

Ví dụ 5.12: Hiển thị tên các tỉnh của sinh viên.

SELECT DISTINCT Tinh

FROM Sinhvien;

❖ ***Truy vấn trên nhiều bảng dùng kết nối Join***

∗ Kết nối bằng trên các thuộc tính cùng tên

Cú pháp:

SELECT <danh sách các cột>
FROM Bảng1 **INNER JOIN** Bảng2 **ON** <Bảng1.khóachính = Bảng2.khóangoại>
[**WHERE** <điều kiện>];

Ví dụ 5.13: Hiển thị thông tin về các sinh viên cùng với tên lớp của họ.

SELECT Sinhvien.*, Lop.TenLop

FROM Sinhvien INNER JOIN Lop ON Sinhvien. MaLop=Lop.MaLop;

∗ Kết nối ngoài trên các thuộc tính cùng tên

Cú pháp:

SELECT <danh sách các cột>
FROM Bảng1 **LEFT|RIGHT JOIN** Bảng2 **ON** <Bảng1.khóachính= Bảng2.khóangoại>
[**WHERE** <điều kiện>];

- LEFT JOIN trả về tất cả các hàng từ bảng thứ nhất, cho dù nó không được so trùng trong bảng thứ hai. Nếu các hàng trong bảng R1 không so trùng trong bảng R2, những hàng này **cũng được** liệt kê.
- RIGHT JOIN trả về tất cả các hàng từ bảng thứ hai, cho dù nó không được so trùng trong bảng thứ nhất. Nếu có bất kỳ hàng nào trong bảng R1 không được so trùng trong bảng R2, các hàng này **cũng được** liệt kê.

Ví dụ 5.14: Hiển thị thông tin về các lớp của các khoa, kể cả những khoa chưa có lớp trong bảng lớp. Thông tin cần hiển thị gồm: mã khoa, tên khoa, tên lớp.

```
SELECT  Khoa.MaKhoa, TenKhoa, TenLop
FROM Khoa LEFT JOIN Lop ON Khoa.MaKhoa=Lop.MaK;
```

❖ ***Câu lệnh truy vấn lồng***

Cú pháp:

```
SELECT <danh sách các cột>
FROM <danh sách các bảng>                          } Câu truy vấn cha
WHERE <so sánh tập hợp> (
        SELECT <danh sách các cột>
        FROM <danh sách các bảng>                  } Câu truy vấn con
        WHERE <điều kiện>);
```

Trong đó, phép so sánh tập hợp thường đi cùng với một số toán tử: IN, NOT IN, ALL, ANY, SOME, EXISTS, NOT EXISTS.

Ví dụ 5.15: Hiển thị thông tin về những SV đã có kết quả điểm ít nhất một học phần:

```
SELECT *
FROM Sinhvien
WHERE MaSV IN  (
        SELECT MaSV
        FROM Ketqua);
```

Ví dụ 5.16: Hiển thị thông tin về những sinh viên đã đăng kí học và không phải học lại học phần nào.

```
SELECT *
FROM Sinhvien, Ketqua
WHERE (Sinhvien.MaSV=Ketqua.MaSV) AND Sinhvien.MaSV NOT IN
                (SELECT MaSV
                FROM Ketqua
                WHERE Diem<4);
```

5.3.2. Câu lệnh cập nhật dữ liệu

Dùng để thay đổi giá trị của thuộc tính cho các dòng của bảng.

Cú pháp:

UPDATE <tên bảng> **SET**
<tên cột 1> = <giá trị| biểu thức mới>
[, <tên cột 2> = <giá trị| biểu thức mới>...]
[WHERE <điều kiện>];

Ví dụ 5.17: Sửa tỉnh của sinh viên có mã sinh viên là 531236 từ Nam Định về Hà Nội.

UPDATE Sinhvien SET Tinh="Hà Nội"

WHERE MaSV="531236";

Lưu ý:

- Những dòng thỏa mãn điều kiện tại mệnh đề WHERE sẽ được cập nhật giá trị mới.
- Nếu không chỉ định điều kiện ở mệnh đề WHERE, tất cả các dòng trong bảng sẽ được cập nhật.
- Lệnh UPDATE có thể gây ra vi phạm ràng buộc tham chiếu như sau:
  - Không cho sửa.
  - Sửa luôn những dòng có giá trị đang tham chiếu đến (ví dụ với HQTCSDL Microsoft Access, trong Relationship nếu ta chọn Cascade Update Related Fields).

5.3.3. Thêm dữ liệu

Khi muốn thêm các dòng mới vào một bảng ta sử dụng cú pháp sau:

Cú pháp 1: Thêm 1 dòng mới vào bảng với các giá trị cụ thể:

INSERT INTO <tên bảng>[<danh sách các thuộc tính>]
VALUES (danh sách các giá trị);

Ví dụ 5.18: Thêm sinh viên có MaSV="536780", Hodem="Lê Thị", Ten="Hà", Ngaysinh=#25/5/90#, Gioitinh="Nữ", Tinh="Hà Nội", MaLop="K52THA" vào bảng Sinhvien.

INSERT INTO Sinhvien

VALUES ("536780", "Lê Thị","Hà", #25/5/90#, "Nữ", "Hà Nội", "K52THA");

Chú ý:

- Thứ tự các giá trị chèn vào phải trùng với thứ tự các cột trong bảng cần chèn.
- Có thể chèn giá trị NULL ở những thuộc tính không là khóa chính.

- Câu lệnh INSERT sẽ gặp lỗi nếu vi phạm các ràng buộc toàn vẹn sau:
    - Khóa chính
    - Tham chiếu
    - NOT NULL - các thuộc tính có ràng buộc NOT NULL bắt buộc phải có giá trị

Cú pháp 2: Thêm nhiều dòng vào bảng

INSERT INTO <tên bảng>[<danh sách các thuộc tính>]
< Câu lệnh truy vấn con >;

Ví dụ 5.19: Sao lưu những sinh viên có quê ở Hà Nội sang bảng Sinhvien_HN.

```
INSERT INTO Sinhvien_HN
SELECT * FROM Sinhvien
WHERE Tinh="Hà Nội";
```

5.3.4. Xóa dữ liệu

Dùng để xóa các dòng của một bảng.

Cú pháp:

DELETE FROM <tên bảng>
[**WHERE** <điều kiện>];

Ví dụ 5.20: Xóa sinh viên có mã sinh viên là 536780 ra khỏi bảng Sinhvien.

```
DELETE FROM Sinhvien
WHERE Masv="536780";
```

Chú ý: - Số lượng dòng bị xóa phụ thuộc vào điều kiện ở mệnh đề WHERE

- Nếu không chỉ định điều kiện ở mệnh đề WHERE, tất cả các dòng trong bảng sẽ bị xóa.
- Lệnh DELETE có thể gây ra vi phạm ràng buộc tham chiếu.
    - Không cho xóa.
    - Xóa luôn những dòng có giá trị đang tham chiếu đến (ví dụ trong Microsoft Access, nếu trong Relationship ta chọn Cascade Delete Related Records).

5.3.5. Các hàm của SQL

SQL xây dựng sẵn một số hàm để tính toán.

a. Hàm AVG

Hàm AVG cho trị trung bình cộng của dữ liệu trong một cột dạng dữ liệu số. Các trị NULL sẽ không được tính toán.

Ví dụ 5.21: Hãy trả về điểm trung bình lần 1 của những sinh viên trong bảng "Ketqua".

```
SELECT AVG(DiemL1) AS [Điểm trung bình]
FROM Ketqua;
```

Ví dụ 5.22: Hãy tính điểm trung bình lần 1 của các sinh viên theo từng lớp.

```
SELECT MaLop, AVG(DiemL1) AS [Điểm trung bình]
FROM Sinhvien INNERJOIN Ketqua ON Sinhvien.Masv=Ketqua.Masv
GROUPBY MaLop;
```

b. Hàm SUM

Hàm SUM tính tổng của dữ liệu trong một cột có kiểu dữ liệu số. Các trị NULL sẽ không được tính toán.

Ví dụ 5.23: Hãy trả về tổng điểm lần 1 của những sinh viên trong bảng "Ketqua".

```
SELECT SUM(DiemL1) AS [Tổng điểm]
FROM Ketqua;
```

c. Hàm MAX

Hàm MAX trả về giá trị lớn nhất trong một cột. Các trị NULL sẽ không được tính toán.

Ví dụ 5.24: Hãy trả về giá trị điểm lớn nhất lần 1 của những sinh viên trong bảng "Ketqua".

```
SELECT MAX(DiemL1) AS [Điểm lớn nhất]
FROM Ketqua;
```

Ví dụ 5.25: Hãy trả về giá trị điểm lớn nhất lần 1 của các sinh viên theo môn học:

```
SELECT Ketqua.MaMH, TenMH, MAX(DiemL1) AS [Điểm lớn nhất]
FROM Ketqua INNER JOIN Monhoc ON Ketqua.MaMH=Monhoc.MaMH
GROUP BY Ketqua.MaMH, TenMH;
```

d. Hàm MIN

Hàm MIN trả về giá trị lớn nhất trong một cột, các trị NULL sẽ không được tính toán.

Ví dụ 5.26: Hãy trả về giá trị điểm lần 1 nhỏ nhất của các sinh viên trong bảng "Ketqua".

```
SELECT MIN(DiemL1) AS [Điểm nhỏ nhất] FROM Ketqua;
```

CÂU HỎI VÀ BÀI TẬP

1. Nêu khái niệm cơ sở dữ liệu.
2. Phần mềm tốt nhất để tạo lập và quản lý CSDL là gì?
3. Trình bày ưu điểm của việc sử dụng CSDL.

4. Nêu các thành phần của hệ CSDL.
5. Hãy phân biệt HQTCSDL và CSDL.
6. Chức năng cơ bản của HQTCSDL là gì?
7. Nhóm ngôn ngữ thao tác dữ liệu SQL bao gồm các lệnh cho phép làm gì trên CSDL?
8. Hãy phân biệt các thuật ngữ sau: Quan hệ, lược đồ quan hệ, lược đồ CSDL quan hệ.
9. Bộ của quan hệ (bản ghi) là gì? Trong một quan hệ có cho phép tồn tại hai bộ giống nhau không?
10. Tại sao cần phải có Khóa trong quan hệ?
11. Hãy phân biệt khái niệm *Khóa* và *Khoá ngoại* trong CSDL.

Chương 6

THUẬT TOÁN VÀ NGÔN NGỮ LẬP TRÌNH

Chương này đề cập đến phương pháp giải quyết vấn đề bằng máy tính, sau đó đi sâu vào hai nội dung chính: 1) Thuật toán (khái niệm thuật toán, các tính chất, các cách diễn đạt thuật toán, phương pháp thiết kế thuật toán và vấn đề đánh giá thuật toán dựa trên độ phức tạp tính toán); 2) Ngôn ngữ lập trình (khái niệm, lịch sử phát triển của ngôn ngữ lập trình, trình biên dịch, trình thông dịch và các bước cơ bản khi lập trình).

6.1. PHƯƠNG PHÁP GIẢI QUYẾT VẤN ĐỀ BẰNG MÁY TÍNH

Như ta đã biết, một trong những chức năng cơ bản nhất của máy tính là xử lý thông tin. Tất cả các quá trình xử lý thông tin bằng máy tính đều được thực hiện theo trình tự:

VÀO → XỬ LÝ → RA

Đầu tiên máy tính tiếp nhận dữ liệu đầu vào, sau đó thực hiện các thao tác xử lý dữ liệu, rồi trả về kết quả sau xử lý dưới dạng thông tin/dữ liệu ra.

Câu hỏi đặt ra là máy tính thực hiện quá trình xử lý thông tin như thế nào? Theo nguyên lý Von Neumann, máy tính hoạt động theo các chương trình (phần mềm) được lập sẵn, việc xử lý thông tin trong máy tính luôn tuân theo nguyên lý này. Với mỗi vấn đề/bài toán đặt ra, để có thể giải quyết được bằng máy tính thì cần phải xây dựng một chương trình máy tính tương ứng. Mỗi chương trình là thể hiện của thuật toán dưới dạng một ngôn ngữ lập trình xác định, thuật toán là thể hiện của phương pháp giải quyết, hướng dẫn các thao tác cụ thể cho máy tính thực hiện để giải quyết vấn đề/bài toán đó.

Nhìn chung, phương pháp chung để giải quyết vấn đề/bài toán bằng máy tính được thể hiện theo sơ đồ sau:

BÀI TOÁN → THUẬT TOÁN → CHƯƠNG TRÌNH → NGÔN NGỮ MÁY→ MÁY THỰC HIỆN

Từ bài toán đặt ra, cần xác định được những dữ liệu nào cần nhập vào máy tính, những dữ liệu/thông tin nào cần phải đưa ra khi kết thúc quá trình xử lý. Sau đó, cần xây dựng thuật toán hay chính là phương pháp xử lý dữ liệu đầu vào để có được dữ liệu /thông tin ra. Khi đã có thuật toán, cần sử dụng một ngôn ngữ lập trình để xây dựng chương trình máy tính tương ứng. Vì máy tính chỉ có thể hiểu được một ngôn ngữ duy nhất là ngôn ngữ máy nên chương trình muốn thực thi được thì cần phải được dịch sang ngôn ngữ máy. Cuối cùng, máy tính sẽ thực hiện các thao tác xử lý theo chương trình lập sẵn và đưa ra các dữ liệu/thông tin ra theo yêu cầu của bài toán.

6.2. THUẬT TOÁN

6.2.1. Khái niệm thuật toán

Thuật toán (thuật giải, algorithm) là một khái niệm quan trọng trong lĩnh vực toán học và tin học. Thuật ngữ *algorithm* được đưa ra từ khá sớm vào khoảng năm 825, xuất phát từ chữ algoritmi – phiên âm La tinh tên của nhà toán học người Trung Á Al-Khwarizmi (780-850) (ông là tác giả của một cuốn sách về số học, trong đó ông đã dùng phương pháp mô tả rất rõ ràng, mạch lạc cách giải những bài toán; sau này, phương pháp mô tả cách giải toán của ông được xem là một chuẩn mực và đã được nhiều nhà toán học khác tuân theo).

Đối với việc giải quyết một bài toán, thuật toán có thể hiểu đơn giản là một dãy hữu hạn các thao tác thích hợp để có thể giải quyết bài toán đó.

Hình 6.1. Al-Khwarizmi

Con tem phát hành vào ngày 06/09/1983 tại Liên Xô, nhân dịp kỷ niệm 1200 năm ngày sinh của ông.

Trong lĩnh vực tin học, thuật toán được xem là một dãy hữu hạn các thao tác, các phép toán có thể thực hiện được theo một trình tự xác định trên một số đối tượng dữ liệu nào đó để đạt được kết quả mong muốn. Lưu ý rằng trong phương pháp giải quyết vấn đề/bài toán bằng máy tính thì đối tượng thực hiện thuật toán là máy tính. Bởi vậy, thuật toán được xây dựng phải bao gồm các thao tác được xác định rõ ràng, đơn giản và thực hiện được hay nói một cách khác là phải "giao cho máy làm được".

Với mỗi một bài toán có thể có nhiều cách giải quyết khác nhau. Một thuật toán đơn giản, có độ chính xác cao, được đảm bảo về mặt toán học, lại dễ triển khai thực hiện trên máy tính với thời gian thực hiện nhanh được coi là một thuật toán hiệu quả.

Khi xây dựng một thuật toán cần xác định rõ thuật toán đó tác động lên dữ liệu nào, bởi xét cho cùng, thuật toán chỉ phản ánh các phép xử lý, còn đối tượng được xử lý chính là dữ liệu. Căn cứ vào các yêu cầu của bài toán đặt ra cần xác định rõ dữ liệu vào, dữ liệu ra và cả dữ liệu trung gian trong quá trình thực hiện các thao tác xử lý.

Việc lựa chọn cấu trúc dữ liệu phù hợp cùng với việc xây dựng được các thuật toán đúng đắn và hiệu quả là những vấn đề mấu chốt khi xây dựng phần mềm. Niklaus Wirth - người sáng lập ra ngôn ngữ lập trình PASCAL đã tổng kết: Thuật toán + Cấu trúc dữ liệu = Chương trình.

Ví dụ: Xét bài toán tìm ước số chung lớn nhất của 2 số nguyên dương a và b:

Input: 2 số nguyên dương a, b

Output: ước số chung lớn nhất của a và b (ký hiệu là (a,b))

Một thuật toán điển hình để giải bài toán này là thuật toán Euclid nguyên bản - thuật toán nổi tiếng được biết đến từ thời Hy Lạp cổ đại (khoảng năm 300 trước công nguyên, trong cuốn Euclid's Elements). Thuật toán được xây dựng dựa trên tính chất: Nếu a=b thì (a,b) = b; ngược lại nếu a>b thì (a,b) = (a-b,b), nếu a<b thì (a,b) = (a,b-a).

<table>
<tr><th>Thuật toán Euclid</th><th colspan="4">Minh họa thuật toán Euclid với a = 20, b = 32</th></tr>
<tr><td rowspan="11">- Bước 1: So sánh a và b, nếu a=b thì dừng thuật toán và thông báo (a,b) = b. Nếu a≠b thì chuyển sang bước 2.
- Bước 2: Nếu a>b thì thay thế a bởi a-b, nếu a<b thì thay thế b bởi b-a. Quay lại thực hiện bước 1.</td><td>Bước thực hiện</td><td>a</td><td>b</td><td>Kiểm tra điều kiện a=b</td></tr>
<tr><td>Bước 1</td><td>20</td><td>32</td><td>Sai</td></tr>
<tr><td>Bước 2</td><td>20</td><td>12</td><td></td></tr>
<tr><td>Bước 1</td><td>20</td><td>12</td><td>Sai</td></tr>
<tr><td>Bước 2</td><td>8</td><td>12</td><td></td></tr>
<tr><td>Bước 1</td><td>8</td><td>12</td><td>Sai</td></tr>
<tr><td>Bước 2</td><td>8</td><td>4</td><td></td></tr>
<tr><td>Bước 1</td><td>8</td><td>4</td><td>Sai</td></tr>
<tr><td>Bước 2</td><td>4</td><td>4</td><td></td></tr>
<tr><td>Bước 1</td><td>4</td><td>4</td><td>Đúng</td></tr>
<tr><td colspan="4">Kết quả: (20,32) = 4</td></tr>
</table>

6.2.2. Các tính chất của thuật toán

Trong cuốn "Những thuật toán cơ bản" (tập 1 của bộ sách Nghệ thuật lập trình máy tính – The Art of Computer Programming), tác giả Donald Ervin Knuth đã chỉ ra rằng, các thuật toán có 5 đặc trưng cơ bản: Đầu vào, Đầu ra, Tính hữu hạn, Tính xác định, Tính hiệu quả.

Hình 6.2. Donald Ervin Knuth

Sinh ngày 10/01/1938, Donald Ervin Knuth là một nhà khoa học máy tính nổi tiếng hiện đang là giáo sư danh dự tại đại học Stanford; ông là tác giả của bộ sách "Nghệ thuật lập trình máy tính" (The Art of Computer Programming) - một trong những bộ sách tham khảo được coi trọng nhất trong ngành khoa học máy tính; ông cũng chính là người đã tạo ra ngành phân tích thuật toán và đã đem lại nhiều cống hiến nền tảng cho ngành khoa học máy tính lý thuyết.

a. Đầu vào (Input)

Một thuật toán có thể không có hoặc có nhiều dữ liệu đầu vào. Các dữ liệu đầu vào sẽ được xác định ngay tại thời điểm ban đầu trước khi thuật toán được bắt đầu/thực thi. Các dữ liệu này được lấy từ các tập hợp/đối tượng quy ước.

Trong thuật toán Euclid nguyên bản, đầu vào là 2 số a, b được lấy từ tập hợp các số nguyên dương.

b. Đầu ra (Output)

Mỗi thuật toán có thể có một hoặc nhiều dữ liệu đầu ra, các dữ liệu đầu ra này có mối liên hệ ràng buộc với dữ liệu đầu vào và chính là kết quả cần đạt được theo yêu cầu của bài toán.

Trong thuật toán Euclid nguyên bản, có duy nhất một đầu ra là giá trị b tại thời điểm kết thúc thuật toán, đó chính là ước chung lớn nhất của 2 số a, b ở dữ liệu đầu vào.

c. Tính hữu hạn (hay còn gọi là Tính kết thúc/Tính dừng - Finiteness)

Thuật toán phải kết thúc sau một số hữu hạn bước thực hiện.

Trong thuật toán Euclid nguyên bản, tổng a+b giảm thực sự qua mỗi lần thực hiện bước 2 và bởi vì ước chung nhỏ nhất của 2 số nguyên dương luôn là 1 nên không bao giờ tổng a+b nhỏ hơn 2. Chính vì vậy, thuật toán bắt buộc phải kết thúc sau một số hữu hạn bước thực hiện.

d. Tính xác định (Definiteness)

Mỗi bước của thuật toán phải được xác định một cách chính xác; các thao tác đưa ra phải được quy định chặt chẽ, rõ ràng cho từng trường hợp.

Tính chất này sẽ đảm bảo cho thuật toán luôn trả về kết quả đúng theo yêu cầu mà bài toán đặt ra. Kết quả thực hiện thuật toán chỉ phụ thuộc vào dữ liệu đầu vào, tức là với cùng một dữ liệu đầu vào thì dù là những người hay những máy tính khác nhau thực hiện thì cũng chỉ cho một kết quả đúng duy nhất.

Trong thuật toán Euclid nguyên bản, các bước thực hiện 1 và 2 được xác định một cách chặt chẽ, rõ ràng dựa theo tính chất đã có để tìm ước chung lớn nhất của 2 số. Chính vì vậy, kết quả trả về của thuật toán là đúng đắn.

e. Tính hiệu quả (Effectiveness)

Một thuật toán luôn được mong đợi là có hiệu quả, trong đó các thao tác phải đủ cơ bản mà ngay cả bản thân con người cũng có thể thực hiện chúng một cách chính xác trong một khoảng thời gian hữu hạn.

Rõ ràng một thuật toán đơn giản với các thao tác cơ bản và không phải sử dụng quá nhiều dữ liệu trung gian sẽ dễ dàng cho việc cài đặt chương trình, đồng thời đảm bảo cho máy tính có

thể thực thi một cách chính xác mà không phải tốn quá nhiều bộ nhớ dùng để lưu trữ và giúp giảm thiểu thời gian thực hiện.

Với thuật toán Euclid, thao tác thực hiện chỉ bao gồm các phép so sánh, phép trừ và phép gán giá trị; đây là những thao tác cơ bản mà dù là con người hay máy tính cũng đều có thể dễ dàng thực hiện trong một khoảng thời gian ngắn.

Ngoài các đặc trưng trên, trong một số tài liệu còn đề cập đến Tính tổng quát (Generality) của thuật toán, tức là thuật toán có thể áp dụng cho cả một lớp các bài toán đồng dạng chứ không phải chỉ áp dụng cho một trường hợp riêng lẻ với dữ liệu đầu vào cụ thể. Ví dụ, thuật toán Euclid áp dụng được với mọi cặp số a, b nguyên dương chứ không phải chỉ áp dụng cho trường hợp a = 20, b = 32 trong ví dụ minh họa. Tuy nhiên, không phải thuật toán nào cũng có tính tổng quát bởi vì trong thực tế, có nhiều bài toán được đặt ra với các dữ liệu đầu vào hoàn toàn xác định mà không tồn tại một lớp bài toán tương tự.

6.2.3. Cách diễn đạt thuật toán

Về cơ bản, có thể diễn đạt thuật toán theo một trong 3 cách sau:

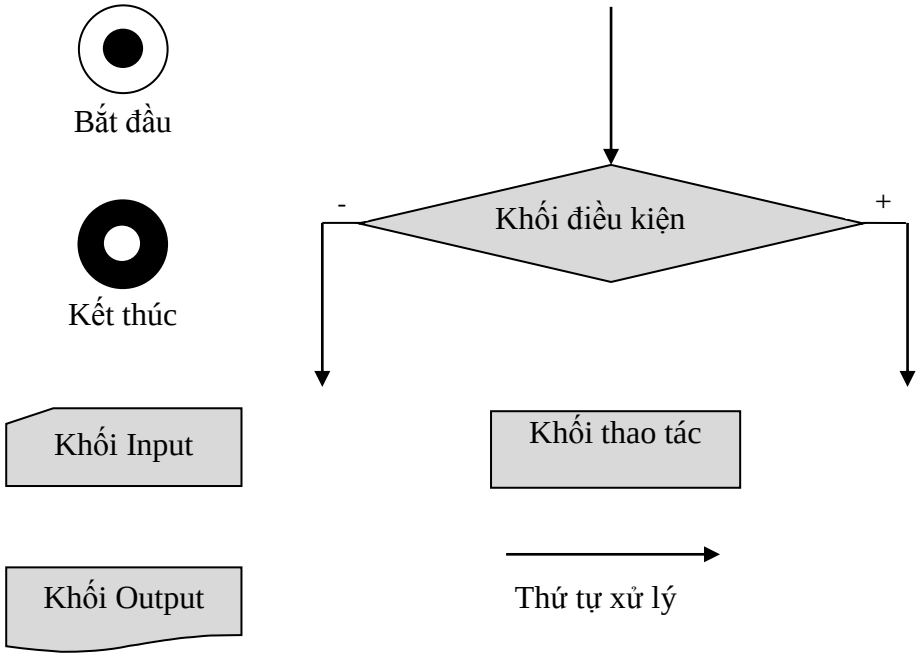
a. Liệt kê từng bước bằng ngôn ngữ tự nhiên

Đây là phương pháp diễn đạt sử dụng ngôn ngữ tự nhiên liệt kê từng bước thực hiện của thuật toán với các quy tắc, các thao tác cụ thể.

Ví dụ: Xem lại thuật toán Euclid tìm ước số chung lớn nhất của 2 số nguyên dương đã đề cập đến ở mục 6.2.1. Khái niệm thuật toán.

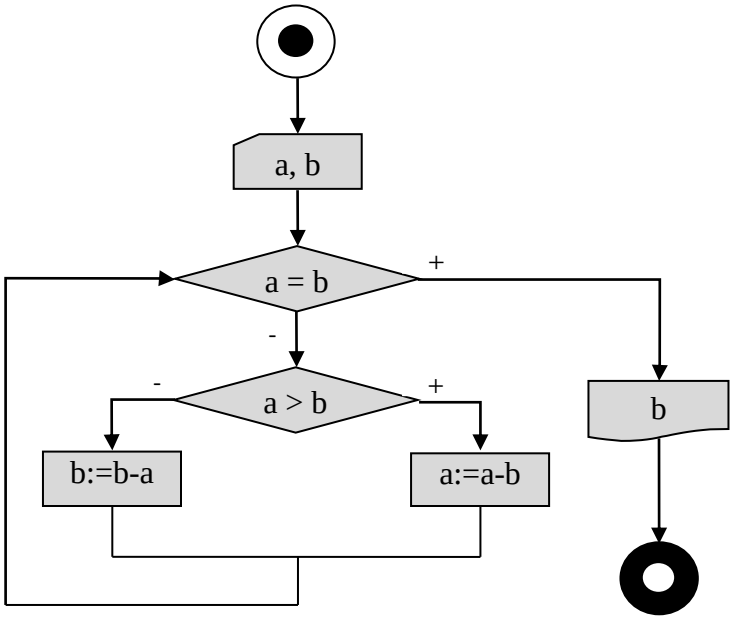
b. Dùng lưu đồ (sơ đồ khối)

Sử dụng các hình khối cơ bản (Bắt đầu, Kết thúc, Khối Input, Khối Output, Khối điều kiện, Khối thao tác) và các cung để thể hiện các thao tác và trình tự thực hiện các thao tác của thuật toán:



Hình 6.3. Các hình khối và cung để biểu diễn thuật toán

Ví dụ: Lưu đồ thuật toán Euclid tìm ước số chung lớn nhất của 2 số nguyên dương được xác định như sau:



Hình 6.4. Lưu đồ thuật toán Euclid tìm ước chung lớn nhất của 2 số nguyên dương a, b

c. Sử dụng giả mã (pseudo code)

Giả mã (hay giả ngôn ngữ lập trình) là một bản mô tả thuật toán ngắn gọn, trong đó sử dụng các cấu trúc điều khiển của một ngôn ngữ lập trình kết hợp linh hoạt với ngôn ngữ tự nhiên và các ký hiệu toán học đơn giản, nhằm diễn tả thuật toán theo cách dễ hiểu nhất đối với người đọc nhưng cũng gần gũi với các ngôn ngữ lập trình để có thể dễ dàng chuyển sang ngôn ngữ lập trình khi cài đặt chương trình.

Nhìn chung, không có bất cứ tiêu chuẩn nào cho cú pháp của giả mã, vì một chương trình viết bằng giả mã không phải là một chương trình có thể thực thi được. Thông thường, khi xây dựng giả mã dựa theo cấu trúc của một ngôn ngữ lập trình, ta có thể bỏ đi những chi tiết không cần thiết như các khai báo, các chương trình con và thay thế những đoạn mã đặc biệt bằng ngôn ngữ tự nhiên để thuật toán trở nên dễ hiểu hơn.

Ví dụ: Giả mã cho thuật toán Euclid tìm ước số chung lớn nhất của 2 số nguyên dương a, b được viết tựa theo cấu trúc của ngôn ngữ lập trình PASCAL:

Nhập a,b
While a≠b do
    If a>b then thay a bởi a-b
    else thay b bởi b-a
Thông báo ước chung lớn nhất là b

Đoạn mã tương ứng được viết bằng ngôn ngữ PASCAL là:

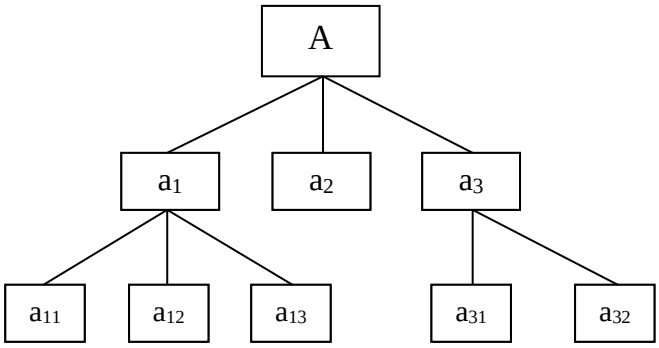
```
Writeln('Nhap 2 so nguyen duong a, b:');
Write('a = '); Readln(a);
Write('b = '); Readln(b);
While a<>b do
    If a>b then a:=a-b
```

```
            else b:=b-a;
        Writeln('Uoc chung lon nhat la ',b);
```

6.2.4. Thiết kế thuật toán

a. Mô-đun hoá và việc giải quyết bài toán

Các bài toán trong thực tế thường khá phức tạp, yêu cầu phải thực hiện nhiều công việc khác nhau mới có thể đạt được mục tiêu đề ra. Để giải quyết một bài toán như vậy, người ta thường sử dụng chiến thuật "chia để trị" (divide and conquer), tức là chia nhỏ bài toán thành các bài toán nhỏ hơn, dễ giải quyết hơn rồi đi giải quyết từng bài toán nhỏ đó. Nếu coi bài toán ban đầu là mô-đun chính, ta chia nó thành các mô-đun con nhỏ hơn, mỗi mô-đun con này lại có thể được tiếp tục chia thành các mô-đun con nhỏ hơn nữa. Quá trình chia nhỏ bài toán như vậy được gọi là quá trình mô-đun hóa bài toán, theo đó bài toán sẽ được thể hiện theo một mô hình phân cấp dạng như mô hình của bài toán A dưới đây:



Hình 6.5. Mô hình phân cấp của bài toán A

Việc giải quyết bài toán tuân theo phương pháp thiết kế top-down (top-down design hay phương pháp thiết kế từ đỉnh xuống). Trước tiên cần phân tích tổng quát bài toán: xuất phát từ các dữ liệu đầu vào và đầu ra, xác định các công việc chính cần thực hiện, sau đó đi sâu phân tích từng công việc và giải quyết từng bước một cách cụ thể, chi tiết hơn. Các thuật toán được thiết kế theo phương pháp top-down cho phép giải quyết các bài toán theo định hướng rõ ràng và là nền tảng cho phương pháp lập trình có cấu trúc.

b. Tinh chỉnh từng bước thuật toán

Tinh chỉnh từng bước là phương pháp thiết kế thuật toán gắn liền với lập trình. Nó cũng thể hiện quá trình mô-đun hóa và phương pháp thiết kế top-down. Bước đầu thuật toán được minh họa bằng ngôn ngữ tự nhiên thể hiện các công việc chính cần thực hiện, càng ở các bước sau, việc minh họa càng trở nên chi tiết hơn với các thao tác xử lý, các phép toán cần thực hiện được chỉ ra một cách cụ thể, đồng thời ngôn ngữ tự nhiên dùng để minh họa được thay thế dần bởi giả ngôn ngữ và ngày càng tiến gần đến ngôn ngữ lập trình. Trong quá trình thiết kế thuật toán từ ngôn ngữ tự nhiên rồi phát triển thành chương trình máy tính, ngôn ngữ thể hiện dần được chuyển đổi theo sơ đồ: Ngôn ngữ tự nhiên $\rightarrow$ Giả ngôn ngữ $\rightarrow$ Ngôn ngữ lập trình.

Xét bài toán sắp xếp phần tử - một bài toán cơ bản trong xử lý thông tin: Cho một dãy gồm n phần tử thuộc kiểu có thứ tự: a_1, a_2..., a_n. Hãy đổi chỗ các phần tử trong dãy sao cho dãy sau khi đổi chỗ là có thứ tự (tăng hoặc giảm dần).

Với bài toán này, hiện đã có nhiều thuật toán được đưa ra. Ở đây ta xét một thuật toán tương đối đơn giản: thuật toán sắp xếp theo kiểu lựa chọn và xét trong trường hợp sắp xếp tăng.

Ý tưởng ban đầu của thuật toán như sau:

- Chọn phần tử nhỏ nhất trong dãy nguồn rồi xếp vào vị trí đầu tiên trong dãy đích;
- Chọn phần tử nhỏ nhất trong dãy nguồn còn lại (tức phần tử nhỏ thứ hai trong dãy nguồn ban đầu) rồi xếp vào vị trí thứ hai trong dãy đích;
- …
- Lặp lại quá trình này cho đến khi hết dãy nguồn.

Một cách tổng quát, với thuật toán này thì tại bước thứ i, ta chọn ra phần tử nhỏ nhất trong dãy nguồn còn lại (tức phần tử nhỏ thứ i trong dãy nguồn ban đầu) rồi xếp vào vị trí thứ i trong dãy đích.

Lưu ý rằng, ở đây ta sẽ sử dụng cấu trúc mảng một chiều để lưu trữ dữ liệu, theo đó các phần tử của dãy sẽ được lưu trữ tại các từ máy kế tiếp trong bộ nhớ. Để tiết kiệm bộ nhớ, ta chỉ sử dụng chung một mảng để lưu trữ cả dãy nguồn và dãy đích, tức là thao tác "chọn ra phần tử nhỏ nhất trong dãy nguồn còn lại rồi xếp vào vị trí thứ i trong dãy đích" thực chất là thao tác "chọn ra phần tử nhỏ nhất trong dãy nguồn còn lại rồi đổi chỗ cho phần tử a[i]".

Giả sử định hướng chương trình sau này sẽ được viết bằng ngôn ngữ lập trình PASCAL, khi đó thuật toán trên sẽ được viết bằng giả mã như sau:

```
For i:=1 to n do
    Begin
        - Chọn phần tử nhỏ nhất aj trong số các phần tử ai..., an
        - Đổi chỗ aj và ai cho nhau
    End;
```

Các công việc trong khối Begin … End sẽ được làm rõ hơn như sau:

- Việc "chọn phần tử nhỏ nhất a_j trong số các phần tử a_i..., a_n" có thể thực hiện bằng cách: Đầu tiên, coi a_i là phần tử nhỏ nhất (j:=i), sau đó lần lượt so sánh phần tử nhỏ nhất với các phần tử a_{i+1}..., a_n; nếu thấy phần tử nào nhỏ hơn thì coi phần tử đó là phần tử nhỏ nhất mới (j:=chỉ số của phần tử nhỏ nhất mới):

```
j:=i;
For k:=i+1 to n do
    If ak<aj then j:=k;
```

- Việc "đổi chỗ a_j và a_i cho nhau" muốn thực hiện được cần sử dụng thêm một phần tử trung gian min:

```
min:=aj;
aj:= ai;
ai:=min;
```

Khi chuyển hoàn toàn thuật toán sang ngôn ngữ lập trình PASCAL, ta có đoạn mã tương ứng như sau:

```
For i:=1 to n-1 do
    Begin
        j:=i;
```

```
For k:=i+1 to n do
        If a[k]<a[j] then j:=k;
If j<>i then
        Begin
                min:=a[j];
                a[j]:=a[i];
                a[i]:=min;
        End;
End;
```

Ví dụ: Cho dãy số ban đầu:

3 6 -2 7 5

Dãy mới được sắp sau từng bước thực hiện thuật toán sắp xếp lựa chọn (i = 1..4):

i=1:	-2	6	3	7	5
i=2:	-2	3	6	7	5
i=3:	-2	3	5	7	6
i=4:	-2	3	5	6	7

6.2.5. Độ phức tạp của thuật toán và vấn đề đánh giá thuật toán

a. Sơ lược về đánh giá thuật toán

Trong thực tế, khi một bài toán được đưa ra, vấn đề không phải chỉ là xây dựng được một thuật toán để giải quyết bài toán đó. Thông thường, với cùng một bài toán, có thể có nhiều thuật toán khác nhau, khi đó vấn đề đặt ra là cần phải xác định được đâu là thuật toán tốt nhất dựa theo một số tiêu chí nào đó. Điều đó đưa chúng ta tiếp cận với một bài toán hết sức quan trọng và thú vị của phân tích thuật toán: xác định các đặc trưng hiệu suất thực thi của thuật toán. Một trong những tiêu chí để đánh giá một thuật toán có phải là tốt hay không, đó là thời gian thực hiện thuật toán được thể hiện qua số lần thực hiện các thao tác. Các tiêu chí khác như khả năng thích ứng của thuật toán với các loại máy tính khác nhau, tính đúng đắn, mức độ đơn giản, hình thức của thuật toán, dung lượng bộ nhớ sử dụng để lưu trữ dữ liệu. Tùy theo từng trường hợp, cần áp dụng các tiêu chí đánh giá khác nhau.

Xét yêu cầu phân tích tính đúng đắn của thuật toán, một thuật toán hiển nhiên chỉ được chấp nhận khi đưa ra được kết quả đúng với yêu cầu của bài toán. Thông thường, người ta có thể cài đặt chương trình thể hiện thuật toán, sau đó chạy chương trình thử nghiệm với một số bộ dữ liệu đầu vào và so sánh các kết quả thử nghiệm với các kết quả đã biết. Tuy nhiên, phương pháp thử nghiệm này chỉ cho phép khẳng định tính sai chứ chưa đủ để đảm bảo được tính đúng đắn của thuật toán. Một phương pháp khác là sử dụng các công cụ toán học để chứng minh tính đúng đắn của thuật toán nhưng phương pháp này không phải lúc nào cũng dễ dàng.

Với yêu cầu phân tích mức độ đơn giản của thuật toán, thông thường chúng ta vẫn mong muốn xây dựng được một thuật toán đơn giản, dễ hiểu, dễ lập trình và chỉnh sửa; nhưng nhiều khi những thuật toán đơn giản lại gây ra sự lãng phí về thời gian và bộ nhớ. Đối với các bài toán cụ thể có số lượng dữ liệu đầu vào nhỏ hoặc thuật toán chỉ được sử dụng một vài lần thì tính đơn giản của thuật toán sẽ được chú trọng.

Tuy nhiên với các bài toán phức tạp, có số lượng dữ liệu đầu vào lớn hoặc các bài toán phổ biến, thuật toán được sử dụng nhiều lần thì cần quan tâm chủ yếu đến hai vấn đề: thời gian thực hiện thuật toán và dung lượng bộ nhớ dùng để lưu trữ dữ liệu. Trong đó, tiêu chí về thời gian thực hiện thuật toán được coi là một tiêu chí quan trọng hàng đầu thường được sử dụng khi phân tích, đánh giá thuật toán. Mặc dù tốc độ xử lý dữ liệu của máy tính hiện nay là rất lớn, có thể lên đến hàng tỷ phép tính trên 1 giây (ví dụ: siêu máy tính Blue Gene/L dạng nhỏ của IBM, ra đời năm 2005, có tốc độ xử lý dữ liệu đạt 100 teraflop – tương đương với 100 nghìn tỷ phép tính trên giây; hay siêu máy tính K của tập đoàn Fujitsu – Nhật Bản, ra đời năm 2012, có tốc độ xử lý dữ liệu đạt 10 petaflop – tương đương với 10 triệu tỷ phép tính trên giây) nhưng với cùng một bài toán, ngay cả khi mức độ chênh lệch về số phép toán thực hiện giữa hai thuật toán được đưa ra là ít thì khi thực hiện chúng lặp đi lặp lại hàng triệu lần, mức độ chênh lệch về thời gian thực hiện cũng không phải là nhỏ.

Có thể thấy, khi cài đặt thành chương trình máy tính thì thời gian thực hiện của một thuật toán phụ thuộc vào rất nhiều yếu tố:

- Số lượng các phép toán sơ cấp: các phép tính số học, các phép tính logic, các phép gán giá trị, chuyển chỗ. Số lượng các phép toán sơ cấp này hiển nhiên sẽ phụ thuộc vào kích thước dữ liệu đầu vào của bài toán (ví dụ, khi dùng cùng 1 thuật toán tính giai thừa, việc tính 4! sẽ cần ít phép toán hơn việc tính 10!). Chính vì vậy, có thể coi thời gian thực hiện của một thuật toán là phụ thuộc vào kích thước dữ liệu đầu vào.

- Ngôn ngữ lập trình, chương trình dịch, hệ điều hành, tốc độ xử lý của máy tính. Tuy nhiên, những yếu tố này không đồng đều với mỗi loại máy tính, vì vậy không thể dùng chúng làm căn cứ để đánh giá thời gian thực hiện của thuật toán và thời gian thực hiện của thuật toán không thể biểu diễn bằng các đơn vị thời gian thông thường như giờ, phút, giây.

Để đánh giá thời gian thực hiện của một thuật toán, người ta sử dụng "Độ phức tạp tính toán của thuật toán" (gọi tắt là độ phức tạp của thuật toán), đây chính là phương pháp đánh giá chỉ phụ thuộc vào kích thước của dữ liệu đầu vào mà không phụ thuộc vào máy tính và các yếu tố liên quan.

b. Độ phức tạp tính toán của thuật toán

Thuật toán T sử dụng để giải một bài toán có kích thước dữ liệu đầu vào n sẽ cần thực hiện T(n) các phép toán sơ cấp. T(n) là một hàm của tham số n, chính là đặc trưng cho độ phức tạp tính toán của thuật toán. Độ phức tạp được ký hiệu bởi chữ "O" lớn.

Xét định nghĩa tổng quát: Giả sử f(n), g(n) là hai hàm số không âm, đồng biến theo n. Hàm f(n) được xác định là có độ phức tạp tính toán cấp g(n), kí hiệu là O(g(n)), f(n) = O(g(n)), khi và chỉ khi tồn tại các hằng số c và n_0 sao cho $f(n) \leq cg(n)$ khi $n \geq n_0$. Khi đó, ta nói f(n) có cấp g(n) (thực chất là cấp lớn không vượt quá g(n)).

Ví dụ: với $f(n) = n^2 + 2n + 3$, rõ ràng $f(n) \leq n^2 + 2n^2 + 3n^2 = 6n^2$ với $\forall n \geq 1$. Do đó, ta có $f(n) = O(n^2)$.

Khi viết T(n) = O(g(n)) nghĩa là tốc độ tăng của T(n) khi $n \to \infty$ không vượt quá tốc độ tăng của g(n). Khi n lớn, g(n) cho ta hình dung độ lớn của T(n) hay nói cách khác, g(n) chính là thước đo độ lớn của T(n). Người ta thường cố gắng ước lượng g(n) sao cho sát với T(n) nhất và có dạng đơn giản nhất.

Độ phức tạp tính toán của thuật toán có thể thuộc các dạng dưới đây (được sắp xếp theo mức độ tăng dần):

- $T(n) = O(1)$: độ phức tạp cấp hằng số (độ phức tạp bằng với một hằng số xác định, không phụ thuộc vào kích thước dữ liệu đầu vào; ví dụ: các thao tác đơn giản như gán, đọc, viết, so sánh giá trị có độ phức tạp cấp O(1));
- $T(n) = O(\log_2 n)$: độ phức tạp cấp hàm logarit;
- $T(n) = O(n)$: độ phức tạp cấp hàm tuyến tính;
- $T(n) = O(n\log_2 n)$: độ phức tạp cấp hàm $n\log_2 n$;
- $T(n) = O(n^2), O(n^3)..., O(n^k)$: độ phức tạp cấp hàm đa thức;
- $T(n) = O(2^n), O(n!), O(n^n)$: độ phức tạp cấp hàm mũ.

Một thuật toán có độ phức tạp tính toán cấp hàm mũ ($O(2^n), O(n!), O(n^n)$) thì tốc độ thực hiện rất chậm.

Một thuật toán có độ phức tạp tính toán từ cấp hàm đa thức trở xuống ($O(1), O(\log_2 n), O(n), O(n\log_2 n), O(n^2), O(n^3)..., O(n^k)$) thì thường chấp nhận được.

Những bài toán chưa tìm được thuật toán với độ phức tạp tính toán từ cấp đa thức trở xuống sẽ được xếp vào dạng bài toán khó.

c. Xác định độ phức tạp tính toán của thuật toán

Đối với các thuật toán phức tạp, việc xác định độ phức tạp tính toán không phải lúc nào cũng dễ dàng. Tuy nhiên, với các thuật toán đơn giản, ta hoàn toàn có thể xác định độ phức tạp của chúng thông qua một số quy tắc sau:

Quy tắc cộng:

Nếu $T_1(n) = O(f(n))$, $T_2(n) = O(g(n))$, thì $T_1(n) + T_2(n) = O(\max\{f(n),g(n)\})$.

Ví dụ: Trong một thuật toán có 3 bước, mỗi bước có độ phức tạp tính toán lần lượt là $T_1(n) = O(n^3)$, $T_2(n) = O(n)$, $T_3(n) = O(n\log_2 n)$ thì thời gian thực hiện 3 bước là:

$$T_1(n) + T_2(n) + T_3(n) = O(\max\{n^3,n,n\log_2 n\}) = O(n^3)$$

Quy tắc nhân:

Nếu $T_1(n) = O(f(n))$, $T_2(n) = O(g(n))$ thì: $T_1(n) . T_2(n) = O(f(n).g(n))$.

Ví dụ: Xét 2 câu lệnh viết bằng ngôn ngữ PASCAL sau:

- Câu lệnh 1: *For j:=1 to n do x:=x+1;*

  Đây là một câu lệnh lặp với số lần lặp là n, tại mỗi bước lặp, chỉ thực hiện thao tác gán giá trị của biểu thức x+1 cho x, do đó thời gian thực hiện là:

  $$T(n) = O(n.1) = O(n)$$

- Câu lệnh 2:

  For i:=1 to n do

  For j:=1 to n do x:=x+1;

  Câu lệnh này gồm 2 vòng lặp lồng nhau, thời gian thực hiện được đánh giá là:

  $$T(n) = O(n.n) = O(n^2)$$

Quy tắc bỏ hằng số:

O(c.f(n)) = O(f(n)) trong đó c là một hằng số.

Ví dụ: $O(n^2/2) = O(n^2)$

Căn cứ vào các quy tắc trên, khi đánh giá độ phức tạp tính toán của thuật toán ta có cũng có thể đánh giá theo một cách đơn giản là chỉ cần quan tâm đến số lần thực hiện phép toán tích cực (active operation - phép toán mà số lần thực hiện nó không ít hơn số lần thực hiện của bất kỳ phép toán nào khác trong thuật toán).

Xét thuật toán sắp xếp theo kiểu lựa chọn cho dãy n phần tử a_1, a_2..., a_n đã được trình bày ở mục 6.2.4, phần b. Phép toán tích cực là phép toán so sánh a[k] < a[j]. Số lần thực hiện phép so sánh này là: $\sum_{i=1}^{n-1}(n-i) = \frac{n(n-1)}{2}$. Do đó, độ phức tạp của thuật toán này là $O(n^2)$.

Trong thực tế, nhiều khi không chỉ kích thước dữ liệu đầu vào mà ngay cả tình trạng dữ liệu cũng là một yếu tố gây ảnh hưởng đến thời gian thực hiện thuật toán. Ví dụ, với bài toán sắp xếp dãy số, rõ ràng nếu dãy số đầu vào đã có sẵn thứ tự giống hoặc gần giống với thứ tự mong muốn thì thời gian thực hiện thuật toán sẽ nhỏ hơn thời gian thực hiện thông thường. Nhìn chung, tùy theo tình trạng dữ liệu đầu vào mà ta sẽ có các độ phức tạp khác nhau ứng với từng trường hợp:

- T_{max}: ứng với trường hợp tình trạng dữ liệu bất lợi nhất cho thuật toán.
- T_{min}: ứng với trường hợp tình trạng dữ liệu thuận lợi nhất cho thuật toán.
- T_{avg}: ứng với trường hợp tình trạng dữ liệu ở mức độ trung bình.

Thông thường, T_{avg} được dùng để so sánh, đánh giá các thuật toán. Tuy nhiên, trong trường hợp việc xác định thời gian thực hiện trung bình quá khó khăn, có thể đánh giá căn cứ vào trường hợp xấu nhất tức là dùng T_{max}. Đặc biệt, với các bài toán *thời gian thực,* đòi hỏi thời gian trả lời không được vượt quá một giới hạn cho trước thì chỉ có thể dùng ước lượng trong trường hợp xấu nhất T_{max}.

6.3. NGÔN NGỮ LẬP TRÌNH

6.3.1. Khái niệm về ngôn ngữ lập trình

Ngôn ngữ lập trình (Programming language) hiểu một cách đơn giản là ngôn ngữ dùng để viết các chương trình máy tính. Mỗi ngôn ngữ lập trình bao gồm một hệ thống các ký hiệu, các từ khóa, các từ dành riêng (hay từ vựng) và các quy tắc để viết chương trình (hay cú pháp). Người lập trình sử dụng ngôn ngữ lập trình để viết chương trình thể hiện thuật toán bao gồm một tập hợp các lệnh được viết theo đúng cú pháp, trong đó, mỗi lệnh mang một ý nghĩa nhất định (còn gọi là ngữ nghĩa), chỉ dẫn cho máy tính thực hiện một công việc cụ thể.

6.3.2. Lịch sử phát triển của ngôn ngữ lập trình

Trong lịch sử phát triển, ngôn ngữ lập trình có thể chia ra làm 3 loại chính:

a. Ngôn ngữ máy (Mã máy - Machine language hay Machine code)

Cùng với sự ra đời của máy tính điện tử, ngôn ngữ máy được xem như là ngôn ngữ nền tảng của bộ vi xử lý. Đây là ngôn ngữ duy nhất mà bộ vi xử lý có thể nhận biết và thực hiện một cách trực tiếp, tất cả các chương trình máy tính được viết bằng các ngôn ngữ khác đều phải được dịch sang ngôn ngữ máy trước khi thực thi. Các lệnh của ngôn ngữ máy được viết ở dạng nhị phân hoặc biến thể của chúng trong hệ 16. Ưu điểm của ngôn ngữ máy là cho phép người lập

trình viết các chương trình điều khiển trực tiếp máy tính thông qua các lệnh máy và các chương trình được thực hiện nhanh chóng do không phải thực hiện bước dịch chương trình. Tuy nhiên, nhược điểm của nó là các lệnh máy dài và khó nhớ, chương trình được viết thường cồng kềnh, vừa mất thời gian khi viết vừa khó khăn cho việc đọc, phát hiện lỗi và hiệu chỉnh chương trình. Ngoài ra, vì tập lệnh của ngôn ngữ máy phụ thuộc vào loại bộ vi xử lý nên một chương trình chỉ chạy được trên những máy tính có cùng loại bộ vi xử lý mà thôi. Trong số các ngôn ngữ lập trình, ngôn ngữ máy được xem là một ngôn ngữ lập trình bậc thấp (thế hệ thứ nhất).

b. Hợp ngữ (Assembly)

Ra đời từ đầu những năm 1950, hợp ngữ được đưa ra nhằm khắc phục các nhược điểm của ngôn ngữ máy. Về cơ bản, hợp ngữ có các cấu trúc lệnh rất giống với ngôn ngữ máy nhưng điểm khác biệt lớn nhất là việc cho phép viết lệnh dưới dạng mã chữ thay vì mã nhị phân. Các mã lệnh ở dạng chữ thường là những từ tiếng Anh viết tắt có ý nghĩa rõ ràng, dễ nhớ. Ngoài ra, hợp ngữ cũng cho phép định địa chỉ hình thức, nghĩa là một vị trí bộ nhớ trong máy tính có thể được tham chiếu tới thông qua một cái tên hoặc ký hiệu, thay vì phải sử dụng địa chỉ thực sự của nó dưới dạng mã nhị phân như trong ngôn ngữ máy. Ví dụ, lệnh *ADD AX, BX* cho phép cộng (addition) số liệu trong các thanh ghi AX, BX với nhau, kết quả để trong thanh ghi AX. Các chương trình hợp ngữ được chuyển sang mã máy thông qua một chương trình đặc biệt gọi là trình hợp dịch (assembler). Mặc dù tương đối dễ dùng hơn mã máy nhưng hợp ngữ vẫn được xem là một ngôn ngữ lập trình bậc thấp (thế hệ thứ hai) bởi vì nó vẫn còn rất gần với tầng thiết kế máy tính, các chương trình được viết bằng ngôn ngữ này luôn có sự liên quan chặt chẽ đến kiến trúc máy tính. Hiện nay, ngôn ngữ này có phạm vi sử dụng khá hẹp, chủ yếu chỉ dùng khi cần lập trình thao tác trực tiếp với phần cứng máy tính hoặc làm các công việc không thường xuyên, thường là trong các trình điều khiển (driver), các hệ nhúng bậc thấp (low-level embedded system) và các hệ thống thời gian thực (real-time system).

c. Ngôn ngữ lập trình bậc cao (Ngôn ngữ thuật toán – High level programming language)

Năm 1957, sự ra đời của ngôn ngữ lập trình bậc cao FORTRAN đã đánh dấu sự khởi đầu cho cuộc cách mạng của ngôn ngữ máy tính, kể từ đó cho đến nay đã có hàng trăm ngôn ngữ lập trình bậc cao ra đời, à ngôn ngữ rất gần gũi với ngôn ngữ tự nhiên và ngôn ngữ toán học, các ngôn ngữ lập trình bậc cao thường sử dụng hệ thống ký hiệu phong phú với các ký hiệu số, các ký hiệu chữ, các ký hiệu toán học và nhiều ký hiệu thông dụng khác, cùng với các từ khóa tiếng Anh đơn giản, các cấu trúc lệnh chặt chẽ, rõ ràng và mang ý nghĩa thực tế. Chính vì vậy, các ngôn ngữ lập trình bậc cao thường dễ học, dễ đọc, dễ viết và hiệu chỉnh chương trình, vừa cho phép thể hiện chính xác các thuật toán lại vừa có tính độc lập cao, ít phụ thuộc vào phần cứng máy tính. Người ta còn gọi ngôn ngữ lập trình bậc cao là ngôn ngữ thuật toán. Cũng giống như hợp ngữ, các chương trình viết bằng các ngôn ngữ lập trình bậc cao muốn máy tính thực thi được thì cần phải được dịch sang ngôn ngữ máy nhờ các chương trình dịch. Ví dụ về một số ngôn ngữ lập trình bậc cao như: FORTRAN, PASCAL, C, C++, JAVA, PHP...

Hiện nay, với hàng loạt các ngôn ngữ lập trình được đưa ra, việc phân loại ngôn ngữ lập trình chỉ mang tính tương đối. Tùy theo từng mục đích mà chúng ta có thể phân loại ngôn ngữ lập trình theo những cách khác nhau. Ví dụ: phân loại theo mức trừu tượng, chúng ta có nhóm ngôn ngữ lập trình bậc thấp và nhóm ngôn ngữ lập trình bậc cao; phân loại theo hình thức lập trình, có nhóm ngôn ngữ khai báo (LIST, PROLOG...) và nhóm ngôn ngữ mệnh lệnh (PASCAL, C...); phân loại theo các họ, có họ ngôn ngữ máy và hợp ngữ, họ ngôn ngữ cổ điển (ALGOL, PASCAL, C...), họ ngôn ngữ hàm (LISP...), họ ngôn ngữ logic (PROLOG...), họ ngôn ngữ hướng đối tượng (C++, JAVA...), họ ngôn ngữ truy vấn (SQL...).

Dưới đây là một số ngôn ngữ điển hình trong lịch sử phát triển của ngôn ngữ lập trình:

- Giai đoạn từ năm 1957 đến những năm đầu 1960:

+ Ngôn ngữ FORTRAN (FORmula TRANslator): được công bố vào năm 1957 bởi công ty IBM, FORTRAN được thiết kế như là một ngôn ngữ lập trình dành cho các nhà khoa học, các kỹ sư và các nhà toán học; ngôn ngữ này được xem như là ngôn ngữ lập trình cấp cao đầu tiên và được chú ý bởi khả năng diễn đạt và tính toán các phương trình toán học một cách dễ dàng.

+ Ngôn ngữ ALGOL (ALGOrithmetic Language): được công bố bởi một ủy ban quốc tế vào những năm cuối 1950 trong một báo cáo có tựa đề ALGOL 58, sau đó được phát triển tiếp thành ALGOL 60, ALGOL 68; với cấu trúc điều khiển hiện đại, ALGOL được sử dụng phổ biến trong các ứng dụng khoa học và toán học.

+ Ngôn ngữ LISP (LISt Processing): được John McCarthy đề xuất vào năm 1958 tại viện công nghệ Massachusetts (MIT) - Mỹ; LISP là ngôn ngữ lập trình hàm đầu tiên, được xem như một ngôn ngữ xử lý danh sách.

+ Ngôn ngữ COBOL (COmmon Business Oriented Language): được phát triển bởi một hội đồng bao gồm các đại diện từ các tổ chức chính phủ, quốc phòng và doanh nghiệp nước Mỹ, trong đó Grace Hopper - làm việc trong Hải quân Mỹ - được mệnh danh là "mẹ đẻ của COBOL"; với khả năng xử lý các tệp tin lớn và thực hiện những phép tính thương mại tương đối đơn giản, COBOL đã từng là một trong những ngôn ngữ được sử dụng rộng rãi nhất cho các ứng dụng thương mại.

- Năm 1963, ngôn ngữ BASIC**,** viết tắt của cụm từ Beginner's All-purpose Symbolic Instruction Code, được phát triển bởi John Kermeny và Thomas Kurtz tại trường đại học Dartmouth; ban đầu, BASIC được thiết kế là một ngôn ngữ lập trình đơn giản, có tính tương tác để các sinh viên học tập và sử dụng, sau đó ngôn ngữ này đã nhanh chóng trở thành một trong những ngôn ngữ lập trình thông dụng.

- Năm 1970, ngôn ngữ PASCAL (lấy theo tên của nhà toán học/vật lý học người Pháp Blaise Pascal)**,** được phát triển bởi Niklaus Wirth, một nhà khoa học máy tính tại Zurich, Thụy Sĩ; ban đầu, PASCAL được phát triển cho mục đích giảng dạy về lập trình cấu trúc, sau đó phiên bản thương mại của nó đã được phát triển rộng rãi trong những năm 80.

- Năm 1972, ngôn ngữ C, được phát triển bởi Dennis Ritchie tại phòng thí nghiệm Bell, Mỹ; ban đầu, C được thiết kế như là một ngôn ngữ dùng để viết các phần mềm hệ thống phục vụ cho hệ thống Unix, nhưng sau đó, nhu cầu dùng C để phát triển nhiều loại phần mềm, kể cả các ứng dụng thương mại đã tăng lên nhanh chóng; nhiều ngôn ngữ lập trình hiện nay được phát triển từ C như: JAVA, JAVASCRIPT, PERL, PHP, PYTHON.

- Năm 1983, ngôn ngữ C++, được phát triển bởi Bjarne Stroustrup tại phòng thí nghiệm Bell, Mỹ; C++ được nâng cao từ ngôn ngữ C, với sự cải tiến về các lớp, hàm ảo và template; ngôn ngữ C++ được sử dụng trong nhiều ứng dụng thương mại, phần mềm nhúng, phần mềm client/server...

Cũng trong năm 1983, ngôn ngữ OBJECTIVE C (lập trình hướng đối tượng mở rộng từ C), được phát triển bởi Bradcox và Tomlove tại công ty Stepstone; đây là ngôn ngữ mở rộng từ C, bổ sung thêm chức năng message-passing cho phép truyền dữ liệu từ tiến trình này sang tiến trình khác trên máy tính thông qua ngôn ngữ SMALLTALK, ngôn ngữ này thường được dùng để lập trình trong hệ điều hành iOS và OS X của Apple.

- Năm 1987, ngôn ngữ PERL (Practical Extraction and Report Language), được phát triển bởi Larry wall, tại công ty Unisys. PERL được tạo ra để xử lý các báo cáo trong hệ thống Unix; hiện nay, ngôn ngữ này được biết đến như là một ngôn ngữ lập trình mạnh mẽ và có tính

linh hoạt cao, được sử dụng trong nhiều ứng dụng cơ sở dữ liệu, quản lý hệ thống, lập trình mạng, lập trình đồ họa...

- Năm 1991, ngôn ngữ PYTHON (đặt tên theo đoàn hài kịch Anh MONTY PYTHON), được phát triển bởi Guido Van Rossum, làm việc tại công ty CWI; PYTHON được tạo ra để hỗ trợ các dạng ngôn ngữ khác và khá thú vị khi sử dụng, thường được dùng trong lập trình ứng dụng web, phát triển phần mềm, bảo mật thông tin.

- Năm 1993, ngôn ngữ RUBBY, được phát triển bởi Yukihiro Matsumoto; đây là ngôn ngữ được đưa ra với mục đích dùng trong giảng dạy, chịu ảnh hưởng của nhiều ngôn ngữ khác như PERL, ADA, LISP, SMALLTALK; hiện nay, ngôn ngữ này thường được sử dụng trong phát triển ứng dụng web.

- Năm 1995, ngôn ngữ JAVA, được phát triển bởi Jame Gosling, làm việc tại công ty Sun Microsystems; JAVA được tạo ra cho một dự án truyền hình tương tác và hiện đã trở thành ngôn ngữ lập trình được sử dụng phổ biến nhất trên thế giới, thường dùng trong lập trình mạng, phát triển ứng dụng web, phát triển phần mềm, phát triển giao diện đồ họa người dùng.

Cũng trong năm 1995, ngôn ngữ PHP (trước đây có nghĩa là Personal HomePage – trang chủ cá nhân, hiện nay được hiểu theo nghĩa là Hypertext PreProcessor – Bộ tiền xử lý siêu văn bản), được phát triển bởi Rasmus Lerdorf; PHP là ngôn ngữ nguồn mở hiện được sử dụng rất rộng rãi để xây dựng, bảo trì các trang web động, phát triển các server...

Cũng trong năm 1995, ngôn ngữ JAVASCRIPT, được phát triển bởi Brendan Eich, làm việc tại công ty NetScape; là ngôn ngữ được tạo ra để mở rộng các chức năng của trang web, ứng dụng trong phát triển web động, xử lý tài liệu dạng pdf, công cụ màn hình...

Ngoài ra còn có nhiều ngôn ngữ lập trình khác như:

- APL (AProgramming Language), một ngôn ngữ khá mạnh, dễ dùng, rất tốt trong việc xử lý dữ liệu được lưu dưới dạng bảng (ma trận).

- FORTH, tương tự như ngôn ngữ C, cho phép tạo mã chương trình nhanh và hiệu quả, ban đầu được phát triển để điều khiển kính viễn vọng không gian.

- LOGO, chủ yếu được biết đến như là một công cụ trong giảng dạy khả năng giải quyết vấn đề.

- MODULA-3, tương tự như PASCAL, sử dụng chủ yếu để phát triển các phần mềm hệ thống.

- PILOT (Programmed Inquiry Learning Or Teaching), được sử dụng bởi những người công tác trong lĩnh vực giảng dạy để viết các chương trình hướng dẫn CAD.

- PL/I (Programming Language/One), ngôn ngữ thương mại và khoa học phối hợp nhiều chức năng của FORTRAN và COBOL.

- PROLOG (PROgramming LOGic), được sử dụng trong trí tuệ nhân tạo.

- RPG (Report Program Generator), cho phép sử dụng các mẫu đặc biệt để giúp người dùng xác định dữ liệu vào, dữ liệu ra và các yêu cầu tính toán của một chương trình.

- ADA **(**lấy theo tên của Augusta Ada Bryon, người được xem là đã viết chương trình đầu tiên), được thiết kế để phục vụ cho việc viết, bảo trì các chương trình lớn trong một khoảng thời gian dài.

6.3.3. Trình biên dịch và trình thông dịch

Máy tính chỉ hiểu được một ngôn ngữ duy nhất là ngôn ngữ máy. Bởi vậy, trước khi được thực thi, các chương trình viết bằng các ngôn ngữ lập trình không phải là ngôn ngữ máy (chương trình nguồn) phải được dịch sang ngôn ngữ máy nhờ các chương trình dịch. Các chương trình dịch có thể chia làm hai loại: trình thông dịch và trình biên dịch.

Trình thông dịch (Interpreter): Sử dụng kỹ thuật thông dịch, dịch từng câu lệnh trong chương trình nguồn được viết bằng ngôn ngữ lập trình bậc cao sang ngôn ngữ máy để máy tính "hiểu" và thực thi ngay câu lệnh đó mà không lưu lại đoạn mã máy tương ứng, sau đó chuyển sang dịch câu lệnh tiếp theo. Với kỹ thuật thông dịch, không có bất kỳ tệp mã đối tượng (tệp mã máy tương ứng với chương trình nguồn) nào được tạo ra. Mỗi lần thực hiện chương trình là một lần chương trình nguồn được thông dịch lại sang ngôn ngữ máy. Thậm chí, nếu một câu lệnh trong chương trình được thực hiện lặp đi lặp lại nhiều lần thì mỗi lần thực hiện lệnh là một lần phải dịch lại câu lệnh đó. Lợi thế của trình thông dịch là cho phép dịch và thực hiện ngay câu lệnh mà không cần phải đợi dịch xong toàn bộ chương trình, ngoài ra trình thông dịch cũng giúp cho việc dò tìm lỗi dễ dàng hơn vì nó chỉ ra chính xác câu lệnh nào chứa lỗi. Nhìn chung, với việc dịch và thực hiện từng câu lệnh, trình thông dịch là thích hợp trong môi trường cần có sự đối thoại giữa con người và hệ thống. Một số ngôn ngữ lập trình có sử dụng trình thông dịch như: BASIC, VISUAL BASIC, PERL, PYTHON...

Trình biên dịch (Compiler): Sử dụng kỹ thuật biên dịch, dịch toàn bộ chương trình nguồn được viết bằng ngôn ngữ lập trình bậc cao sang ngôn ngữ máy và tạo ra tệp mã đối tượng tương ứng, sau đó bộ liên kết sẽ liên kết các đối tượng thành phần với nhau và tạo ra tệp thực thi, cả tệp đối tượng và tệp thực thi đều được nạp vào máy tính để sử dụng khi cần. Trong quá trình biên dịch, trình biên dịch sẽ phân tích từ vựng và cú pháp của các câu lệnh, nếu trong chương trình nguồn có lỗi về mặt cú pháp thì trình biên dịch sẽ thông báo danh sách tất cả các lỗi để lập trình viên chỉnh sửa, tệp mã đối tượng chỉ được tạo ra khi chương trình nguồn không còn bất kỳ lỗi cú pháp nào. Mỗi lần thực hiện chương trình chỉ cần sử dụng chương trình thực thi đã được tạo trước đó mà không cần phải tiến hành biên dịch lại chương trình nguồn. Vì vậy, việc sử dụng trình biên dịch là thích hợp với các chương trình có tính ổn định và được thực hiện nhiều lần. Thông thường, mỗi ngôn ngữ lập trình bậc cao đều có một trình biên dịch tương ứng, ví dụ: PASCAL, C, C++...

6.3.4. Các công việc của người lập trình

Về cơ bản, để tạo ra một chương trình bằng ngôn ngữ lập trình bậc cao, người lập trình cần thực hiện các công việc theo trình tự sau:

Bước 1: Soạn thảo chương trình

Dựa vào thuật toán và ngôn ngữ lập trình để viết chương trình, sau đó sử dụng một trình soạn thảo chuyên dụng để nhập nội dung chương trình, lưu tệp chương trình với phần mở rộng tên tệp phù hợp với ngôn ngữ lập trình sử dụng, ví dụ: phần mở rộng tên tệp là .pas cho tệp chương trình viết bằng ngôn ngữ PASCAL, .c cho tệp chương trình viết bằng ngôn ngữ C hay .cpp cho tệp chương trình viết bằng ngôn ngữ C++... Tệp chương trình này được gọi là tệp mã nguồn (source code).

Bước 2: Biên dịch chương trình

Sử dụng trình biên dịch (compiler) thích hợp để biên dịch tệp chương trình nguồn sang tệp mã máy tương ứng (tệp đối tượng hay object code). Nếu chương trình nguồn có một số lỗi nào đó về mặt cú pháp thì trình biên dịch sẽ thông báo danh sách tất cả các lỗi, khi đó cần quay lại bước 1, sử dụng trình soạn thảo để chỉnh sửa chương trình nguồn. Khi tệp đối tượng đã được

tạo, bộ liên kết (linker) sẽ thực hiện việc liên kết các đối tượng thành phần với nhau và tạo ra tệp thực thi (executable code) cho chương trình.

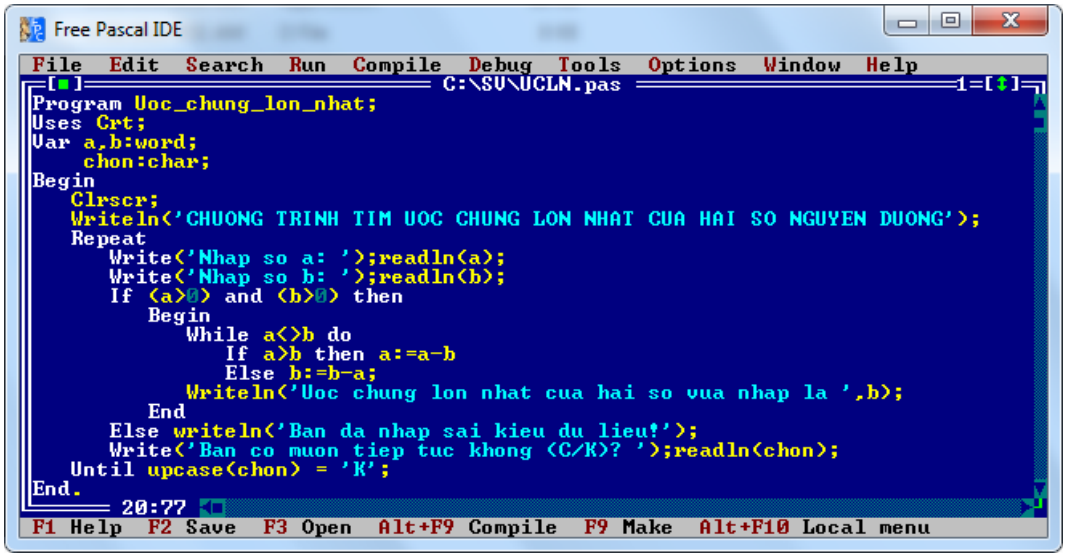
Bước 3: Chạy thử chương trình

Chạy chương trình (kích hoạt tệp thực thi), nhập các dữ liệu đầu vào (các dữ liệu mẫu dùng để kiểm tra) và kiểm tra các kết quả được đưa ra. Nếu kết quả thu được không đúng hoặc có lỗi khi thực thi chương trình thì cần kiểm tra, chỉnh sửa lại thuật toán, rồi quay lại bước 1 để chỉnh sửa lại chương trình.

Thông thường, trong các môi trường phát triển tích hợp (IDE - Integrated Development Environment) có tích hợp sẵn trình soạn thảo, trình biên dịch, bộ liên kết, trình gỡ rối... và cho phép chạy thử chương trình. Tuy nhiên, người lập trình cũng có thể sử dụng một trình soạn thảo chuyên dụng, độc lập để soạn thảo chương trình nguồn; sau đó sử dụng một trình biên dịch thích hợp để biên dịch rồi chạy chương trình bằng cách kích hoạt tệp thực thi đã được tạo. Một trình soạn thảo được sử dụng khá phổ biến hiện nay là Notepad++, đây là một phần mềm miễn phí cho phép soạn thảo với nhiều ngôn ngữ lập trình khác nhau, phù hợp với các cá nhân thường xuyên phải làm việc cùng lúc với nhiều ngôn ngữ lập trình.

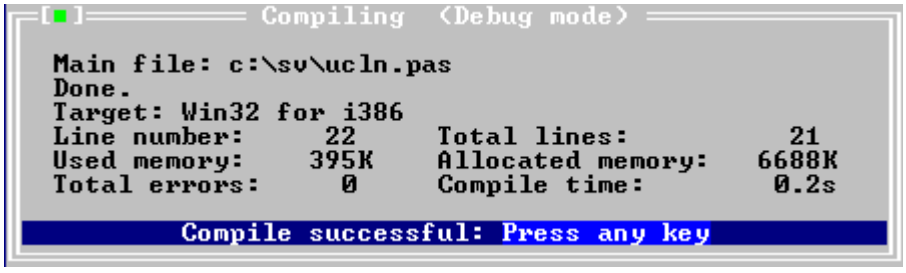
Ví dụ 1: Viết chương trình tìm ước số chung lớn nhất của 2 số nguyên dương, chương trình viết bằng ngôn ngữ PASCAL, sử dụng phần mềm Free Pascal (IDE, version 2.6.2).

Bước 1: Khởi động phần mềm Free Pascal, sử dụng trình soạn thảo nhập nội dung chương trình nguồn, sau đó lưu tệp mã nguồn dưới dạng .pas:



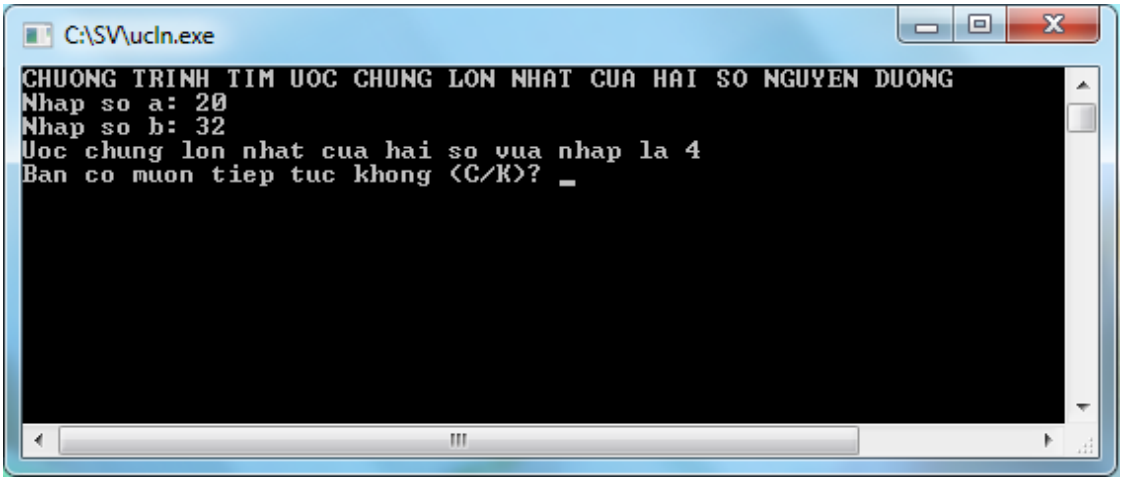
Hình 6.6. Soạn thảo chương trình tìm ước chung lớn nhất với Free Pascal

Bước 2: Nhấn tổ hợp phím Alt+F9 để biên dịch chương trình. Khi chương trình nguồn không có lỗi cú pháp, hệ thống sẽ đưa ra thông báo quá trình biên dịch đã thành công:



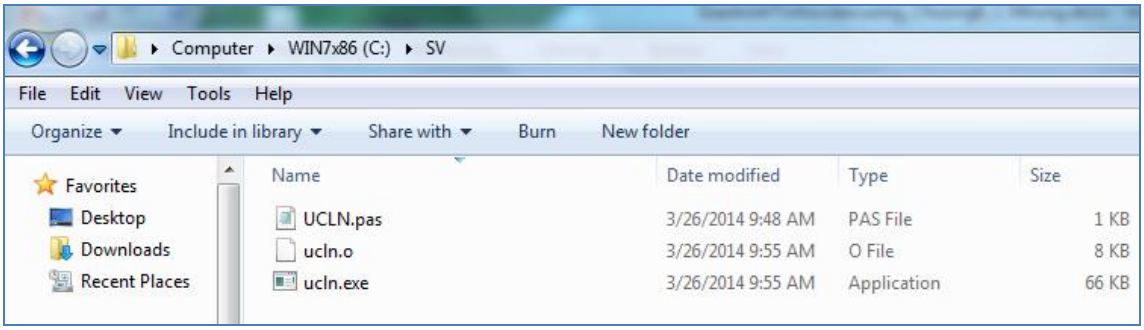
Hình 6.7. Biên dịch chương trình tìm ước chung lớn nhất

Bước 3: Nhấn tổ hợp phím Ctrl+F9 để chạy thử chương trình:



Hình 6.8. Chạy thử chương trình tìm ước chung lớn nhất

Sử dụng trình quản lý ổ đĩa Windows Explorer, kiểm tra trong thư mục lưu tệp mã nguồn, ta sẽ thấy bên cạnh tệp mã nguồn UCLN.pas được tạo ở bước 1, sẽ có thêm các tệp đối tượng UCLN.o và tệp thực thi UCLN.exe được tạo ở bước 2.

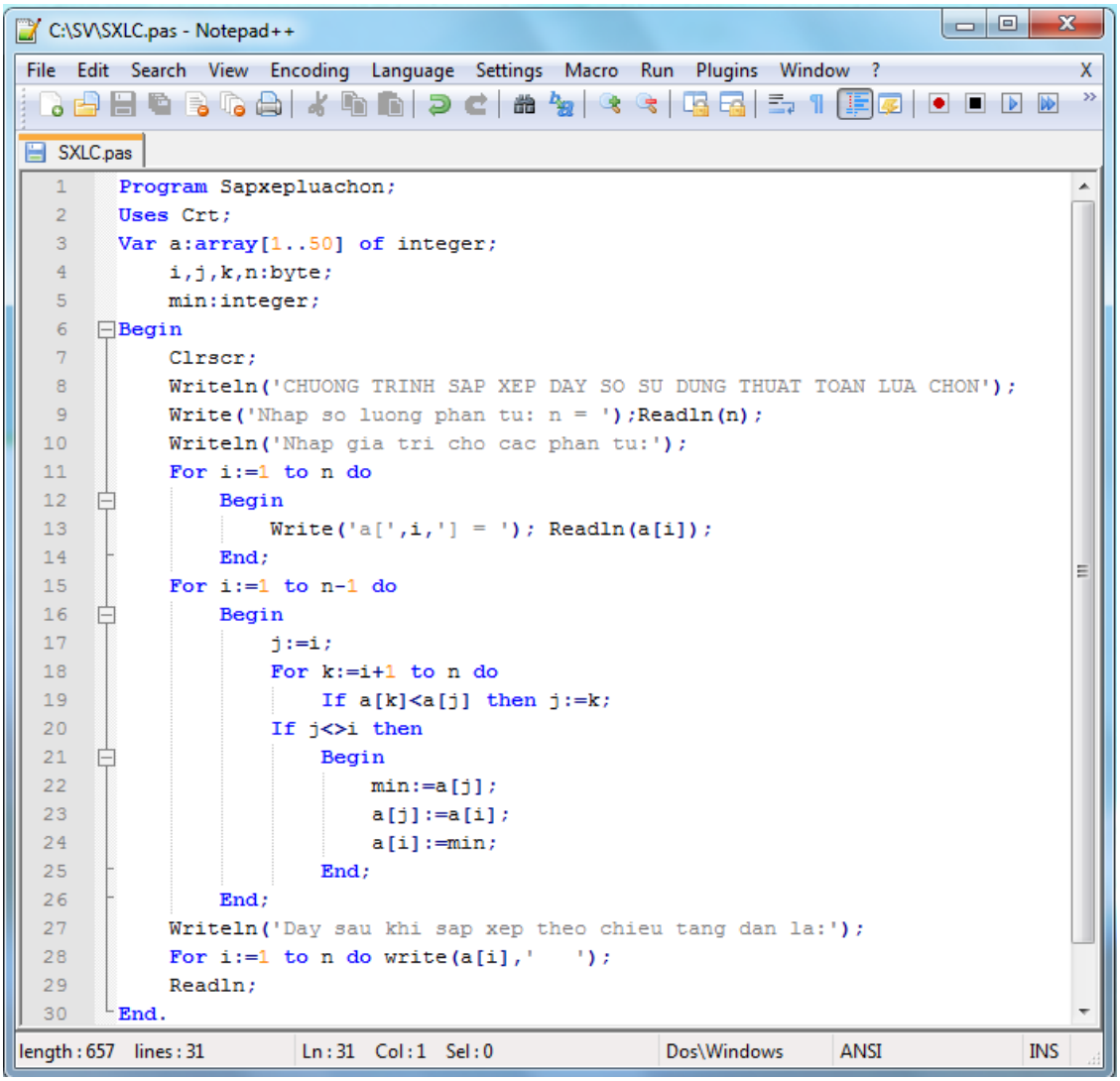


Hình 6.9. Các tệp được tạo sau bước biên dịch chương trình tìm ước chung lớn nhất

Những lần tiếp theo khi muốn chạy chương trình tìm ước chung lớn nhất của hai số nguyên dương, ta chỉ cần kích hoạt tệp UCLN.exe đã được lưu trữ.

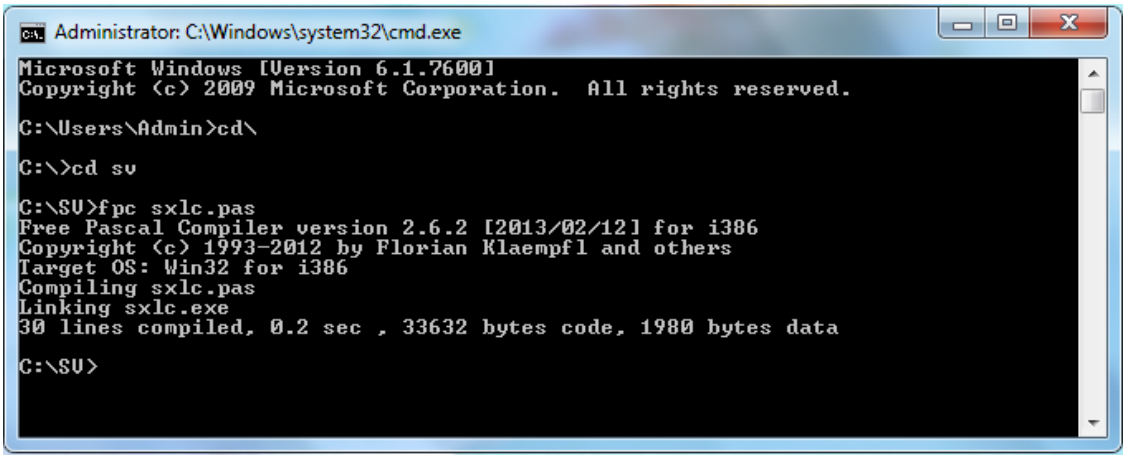
Ví dụ 2: Viết chương trình cho bài toán sắp xếp dãy số nguyên a_1, a_2..., a_n theo chiều tăng dần, sử dụng thuật toán lựa chọn (chương trình viết bằng ngôn ngữ lập trình PASCAL, sử dụng phần mềm Notepad++ để soạn thảo và trình biên dịch Free Pascal Compiler để biên dịch chương trình).

Bước 1: Sử dụng phần mềm Notepad++ soạn thảo nội dung chương trình nguồn, sau đó lưu tệp mã nguồn dưới dạng .pas:



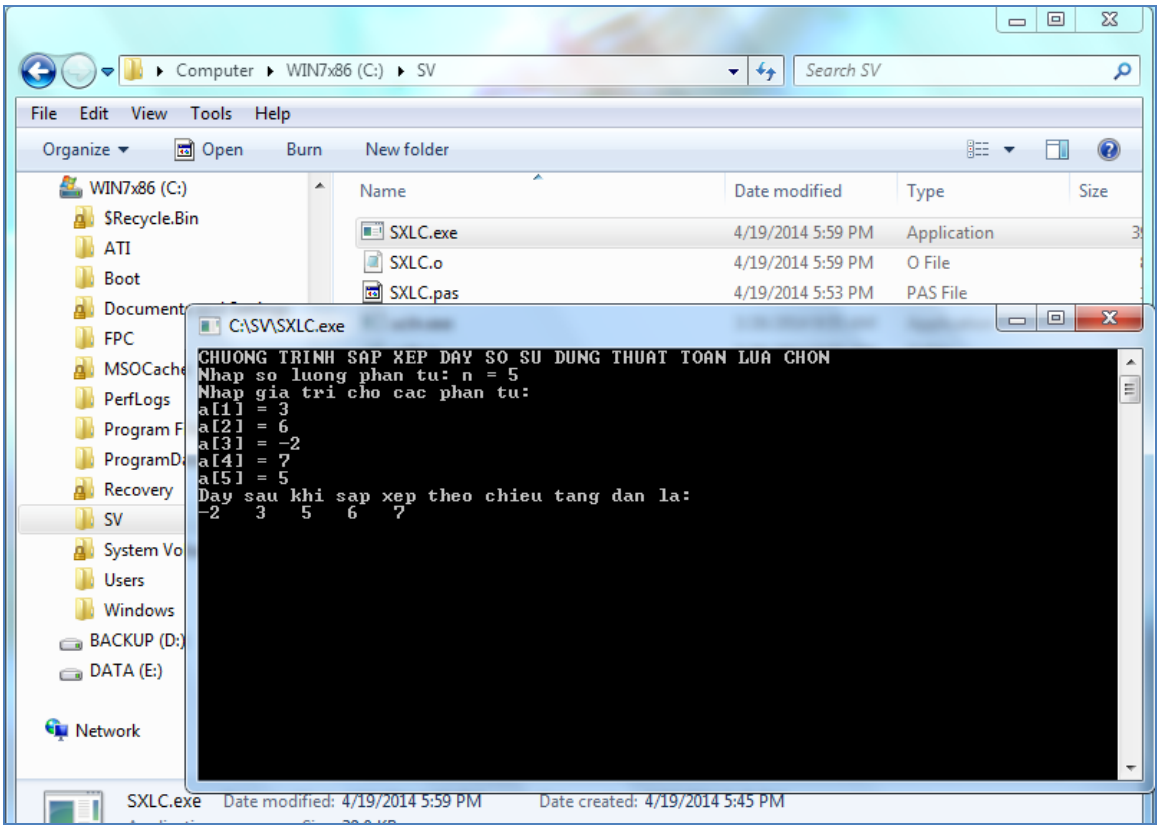
Hình 6.10. Soạn thảo chương trình sắp xếp dãy số với trình soạn thảo Notepad++

Bước 2: Biên dịch chương trình, sử dụng trình biên dịch Free Pascal Compiler. Khi chương trình nguồn không có lỗi cú pháp, hệ thống sẽ đưa ra thông báo quá trình biên dịch đã thành công, tệp đối tượng và tệp thực thi đã được tạo:



Hình 6.11. Biên dịch chương trình sắp xếp dãy số

Bước 3: Kích hoạt tệp thực thi để chạy thử chương trình:



Hình 6.11. Chạy thử chương trình sắp xếp dãy số

CÂU HỎI VÀ BÀI TẬP

1. Hãy trình bày phương pháp giải quyết vấn đề bằng máy tính.
2. Thuật toán là gì? Hãy trình bày những đặc trưng cơ bản của thuật toán.
3. Có mấy cách để diễn đạt một thuật toán, là những cách nào?
4. Hãy thiết kế thuật toán cho các bài toán sau:
   a. Cho n là một số nguyên dương, n>1, hãy tính giá trị n!
   b. Cho n là một số nguyên dương, n>1, hãy tính tổng S theo công thức:

      $$S = 1/2 + 1/3 + \ldots + 1/n$$

   c. Cho dãy n số nguyên a_1, a_2..., a_n (n>1). Hãy sắp xếp dãy số đã cho theo chiều giảm dần.
5. Nêu khái niệm độ phức tạp tính toán của thuật toán. Hãy xác định độ phức tạp tính toán cho các thuật toán đã xây dựng ở câu 4.
6. Ngôn ngữ lập trình là gì? Trong lịch sử phát triển, ngôn ngữ lập trình có thể chia làm mấy loại, là những loại nào?
7. Chương trình dịch dùng để làm gì? Có những loại chương trình dịch nào?
8. Khi lập trình để giải quyết một bài toán cụ thể, các lập trình viên cần thực hiện các công việc nào?

Chương 7

CÁC VẤN ĐỀ XÃ HỘI CỦA CÔNG NGHỆ THÔNG TIN

Công nghệ thông tin ngày nay đã xâm nhập vào mọi lĩnh vực trong đời sống thường ngày của con người từ cách thức chúng ta làm việc, tương tác, trao đổi với đồng nghiệp, bạn bè và người thân tới những hoạt động thuần túy mang tính cá nhân, chúng ta thực sự đang sống trong thời đại thông tin với những cư xử gắn liền với hạ tầng công nghệ. Song hành với những hình thức cư xử mới này, tất yếu sẽ xuất hiện những chuẩn mực đạo đức mới cũng như những loại hình phạm tội mới và những chế tài pháp luật mới. Do đó, việc nắm được những kiến thức cơ bản nhằm nhận biết và phân biệt cũng như khả năng tự bảo vệ trước các hình thức phạm tội mới cũng như các chế tài pháp luật nhằm hạn chế các hành vi phạm tội mới này là vô cùng cần thiết.

7.1. CÁC TÀI NGUYÊN CÓ THỂ BỊ XÂM PHẠM

7.1.1. Nội dung thông tin

Thông tin vốn đã quan trọng, thì ngày nay trong xã hội thông tin, lại càng trở nên quan trọng hơn bao giờ hết. Với sự tiện lợi trong lưu trữ, vận chuyển và chia sẻ, thông tin ngày càng được đưa vào nhiều hơn trong các hệ thống công nghệ thông tin. Từ những loại thông tin có tính chất công cộng tới những loại thông tin nghiệp vụ, thông tin bí mật chiến lược, tới những thông tin hết sức riêng tư, tất cả đều đã được đẩy vào các hệ thống thông tin. Cũng từ đây, những vấn đề tội phạm liên quan tới nội dung thông tin xuất hiện.

Nội dung thông tin bị tấn công thường là mục tiêu chiếm đoạt hoặc phá hủy thông tin. Chiếm đoạt thông tin là có được nội dung thông tin mà bản thân kẻ tấn công không có thẩm quyền để xem thông tin đó; Phá hủy thông tin là việc xóa bỏ hoặc thay đổi thông tin một cách trái phép. Các tấn công vào nội dung thông tin gây hậu quả vô cùng nghiêm trọng tới chính phủ, tổ chức và cá nhân. Gần đây Edward Snowden làm rò rỉ thông tin bí mật của cục tình báo Mỹ là một ví dụ điển hình ở mức chính phủ. Với tổ chức, công ty, việc để lộ hay phá hủy những thông tin chiến lược quan trọng, các bí quyết sản xuất, chế biến sẽ ảnh hưởng lớn tới sự tồn vong của tổ chức, công ty đó. Với cá nhân, việc lộ những thông tin riêng tư không chỉ gây khó chịu cho cá nhân mà thậm chí còn dẫn tới những hậu quả nghiêm trọng cho tính mạng của cá nhân đó. Như vậy, bảo vệ nội dung thông tin trở thành một vấn đề vô cùng quan trọng nhằm đảm bảo cuộc sống bình thường của mọi người trong xã hội thông tin.

7.1.2. Tài nguyên hạ tầng công nghệ thông tin

Xã hội thông tin ngày càng đẩy con người phụ thuộc vào hạ tầng công nghệ thông tin. Từ các giao dịch tài chính, nghiệp vụ tới các giao tiếp thông thường trong đời sống hàng ngày, tất thảy đều được thực hiện dựa trên hạ tầng công nghệ thông tin. Do vậy, việc hạ tầng này sụp đổ hoặc rơi vào trạng thái quá tải không thể đáp ứng, sẽ dẫn tới những hậu quả khôn lường.

Tấn công trên hạ tầng công nghệ thông tin thường tập trung vào hạ tầng tính toán và lưu trữ. Đối tượng tấn công sẽ tìm mọi cách để tiêu thụ hết tài nguyên tính toán và lưu trữ khiến hạ tầng công nghệ thông tin bị quá tải, thậm chí bị sụp đổ.

7.1.3. Định danh người dùng

Trong môi trường mạng, nhiều khi chúng ta sử dụng một định danh nhất định gắn với bản thân chúng ta trong đời sống thực. Định danh này làm cơ sở cho những hoạt động giao tiếp và giao dịch trong đời sống thực có thể được đưa vào các hệ thống thông tin. Nhờ có định danh này

mà thông tin được trao đổi có tính tin cậy. Do đó, việc bị đánh cắp định danh hay giả mạo định danh sẽ gây ra những hậu quả khôn lường. Trước tiên sẽ là đánh mất uy tín của người sử dụng định danh đó, sau đó là những hiểm họa với bản thân người sở hữu định danh và các cá nhân tổ chức thực hiện giao dịch liên quan tới định danh đó.

7.2. CÁC HÌNH THỨC TẤN CÔNG

Ở mục trên, chúng ta đã thấy được những loại tài nguyên thông tin có thể bị xâm phạm trong các hệ thống thông tin. Ở mục này, chúng ta sẽ xem xét những kẽ hở cũng như các cách thức mà kẻ xấu thực hiện tấn công vào các tài nguyên thông tin.

7.2.1. Tận dụng các lỗ hổng phần mềm

Mặc dù được xây dựng với mục tiêu tốt đẹp, các sản phẩm phần mềm vẫn có thể chứa đựng những lỗi hoặc những điểm yếu mà kẻ xấu có thể lợi dụng để thực hiện các hành vi xâm phạm tới các tài nguyên thông tin kể trên.

Ngày nay, hệ thống máy tính thường được cài đặt một lượng lớn các sản phẩm phần mềm để phục vụ các nhu cầu sử dụng khác nhau, do đó nguy cơ tiềm ẩn các lỗ hổng trong các hệ thống máy tính là rất lớn. Những lỗ hổng có thể đến từ bản thân thiết kế của sản phẩm, những lỗi lập trình trong quá trình phát triển, hay những lỗi trong quá trình cài đặt, cấu hình và vận hành sản phẩm. Các lỗ hổng cũng có thể đến từ hạ tầng đóng vai trò làm nền cho sản phẩm như hệ điều hành, hệ quản trị cơ sở dữ liệu hay những công cụ, thư viện được sử dụng trong quá trình phát triển sản phẩm phần mềm như ngôn ngữ lập trình, trình biên dịch.

Những lỗ hổng đến từ bản thân thiết kế của sản phẩm hay logic chức năng của sản phẩm được gọi là những lỗ hổng logic ứng dụng, chúng rất đa dạng và biến đổi tùy thuộc vào bản thân ứng dụng, do đó rất khó để phát hiện. Để loại bỏ những lỗ hổng logic, ứng dụng cần sự tham gia của chuyên gia an ninh, ứng dụng không thể sử dụng các công cụ tự động. Ngược lại, những lỗ hổng đến từ hạ tầng, công cụ và thư viện có thể được phát hiện dễ dàng hơn. Thông thường dựa trên những lỗ hổng đã được biết trên hạ tầng, công cụ và thư viện, một ứng dụng tự động có thể thực hiện một tìm kiếm một cách có hệ thống để phát hiện các lỗ hổng đó.

Mặc dù sự phát triển mạnh mẽ của kỹ nghệ phần mềm cũng như sự hỗ trợ của các công cụ kiểm thử hiện đại, sản phẩm phần mềm vẫn có thể tồn tại những lỗ hổng. Do đó, khai thác lỗ hổng phần mềm vẫn là một phương pháp hữu hiệu nhằm tấn công vào các hệ thống thông tin.

7.2.2. Sử dụng các phần mềm độc hại

Phần mềm độc hại là phần mềm được xây dựng với mục đích xấu, được sử dụng như công cụ để tấn công vào các hệ thống thông tin. Ở cách thức tấn công này, phần mềm độc hại phải được cài đặt lên hệ thống máy tính của người dùng và phải được kích hoạt để chạy. Thông thường, phần mềm độc hại được cài đặt bởi tin tặc thông qua những lỗ hổng phần mềm hoặc trực tiếp bởi người sử dụng. Phần mềm độc hại có thể được kích hoạt trực tiếp bởi người sử dụng hoặc thông qua những lệnh khởi động của hệ điều hành.

Phần mềm độc hại phát triển mạnh về số lượng và sự đa dạng, tính tới năm 2008 số lượng phần mềm độc hại đã vượt mốc 1 triệu, theo báo cáo của GData, chỉ trong nửa đầu năm 2010, đã phát hiện tới 1.017.208 phần mềm độc hại mới, con số này lớn hơn một nửa tổng số phần mềm độc hại năm 2009. Phần mềm độc hại thực sự trở thành mối nguy hại lớn với các hạ tầng công nghệ thông tin. Dưới đây chúng ta xem xét một số loại phần mềm độc hại chính.

a. Virus máy tính

Virus máy tính (thường được người sử dụng gọi tắt là virus hay vi-rút) là những chương trình hoặc đoạn mã lệnh được thiết kế để bám vào một tệp tin nào đó. Virus sẽ thi hành khi

những thao tác nhất định xảy ra trên tệp tin mà nó lây nhiễm được thực hiện, chẳng hạn như người sử dụng yêu cầu hệ điều hành thi hành tệp tin đó hay mở tệp tin đó bằng một trình ứng dụng nào đó. Virus thi hành sẽ thực hiện hai nhiệm vụ chính:

- Thực hiện chức năng mà virus được thiết kế để thực hiện. Những chức năng này có thể đơn giản là một trò đùa, cũng có thể là những hành động phá hoại với những hậu quả khôn lường. Nhiều virus cài đặt kỹ thuật đặt bẫy, để chức năng của virus chỉ thực sự hoạt động khi một số điều kiện cụ thể được thỏa mãn, chẳng hạn như virus Doodle Yankee đúng 17h là hát quốc ca.

- Thực hiện tìm kiếm các tệp tin trên hệ thống máy tính, tạo ra các nhân bản của nó và bám vào các tệp tin được lựa chọn. Cơ chế nhân bản có thể đơn giản là tạo ra một bản sao của chính bản thân virus, cũng có thể vô cùng phức tạp nhằm giúp cho mỗi lần nhân bản có được những khác biệt nhất định so với virus ban đầu.

Thuật ngữ Virus máy tính lần đầu tiên được đưa ra trong bài báo của Fred Cohen năm 1984 với tiêu đề Computer Viruses – Theory and experiments. Sau hơn 30 năm, virus máy tính cũng có sự phát triển mạnh mẽ về số lượng song hành cùng sự phát triển của phần mềm độc hại nói chung. Để phân loại virus máy tính cũng có nhiều cách khác nhau. Ở đây, ta phân loại virus thành 2 nhóm chính là virus biên dịch (compiled virus) và virus thông dịch (interpreted virus).

Virus biên dịch:

Virus biên dịch là loại virus có thể được thi hành trực tiếp bởi hệ điều hành. Để làm được điều đó, mã lệnh của virus biên dịch phải được biên dịch thành tệp tin có thể được thi hành bởi hệ điều hành. Với đặc trưng này, virus biên dịch chỉ có thể lây nhiễm trên một dòng hệ điều hành nhất định với một kiến trúc vi xử lý nhất định. Virus biên dịch lại có thể chia là ba nhóm chính là virus tệp tin (file virus), virus khởi động (boot virus) và virus đa năng (multipartite virus).

- Virus tệp tin là những loại virus biên dịch lây nhiễm tới các tệp tin thi hành trên hệ thống máy tính như ứng dụng soạn thảo văn bản, bảng tính hay các chương trình trò chơi, các chương trình chát trên mạng... Virus tệp tin lây lan đơn giản bằng cách gắn vào tệp tin thi hành. Khi tệp tin thi hành bị nhiễm virus tệp tin được kích hoạt để thực hiện, virus cũng sẽ thi hành, thực hiện chức năng của nó và lây lan sang các tệp tin thi hành khác. Hai ví dụ tiêu biểu cho virus tệp tin là Jerusalem và Cascade. Jerusalem là virus được phát hiện tại Jerusalem năm 1987, lây nhiễm các tệp tin thi hành trong môi trường DOS, nó có thể đơn giản là in ra các thông điệp hoặc xóa các tệp tin. Cascade là virus lây nhiễm rộng rãi suốt thập kỷ 1980 và những năm đầu của thập kỷ 1990. Cải tiến quan trọng của virus Cascade là việc sử dụng thuật toán mã hóa để lẩn tránh các phần mềm diệt virus.

- Virus khởi động là loại virus biên dịch lây nhiễm vào phân vùng khởi động của các thiết bị lưu trữ. Như ta đã biết, phân vùng khởi động lưu trữ những thông tin về thiết bị lữu trữ, nó sẽ được các chương trình khởi động đọc để khởi tạo hệ điều hành hoặc để hệ điều hành lấy thông tin về thiết bị. Điểm mạnh của virus khởi động là nó có thể được kích hoạt tự động bởi các chương trình khởi động mà không cần chờ người sử dụng kích hoạt. Michelangelo và Stoned là những ví dụ tiêu biểu cho virus khởi động. Virus Michelangelo được phát hiện ngày mùng 4 tháng 2 năm 1991 và nó đã khiến thế giới máy tính nín thở chờ đợi ngày mùng 6 tháng 3 năm 1992 (ngày sinh của nghệ sĩ Michelangelo) - ngày virus Michelangelo sẽ hủy diệt thế giới máy tính. Tuy nhiên trên thực tế, theo thống kê chỉ khoảng 20.000 trường hợp xuất hiện sự cố mất dữ liệu được ghi nhận vào ngày 6 tháng 3 năm 1992.

- Virus đa năng là loại virus biên dịch sử dụng nhiều phương thức lây nhiễm, thường bao gồm cả lây nhiễm theo tệp tin và lây nhiễm trên phân vùng khởi động. Virus đa năng tổ hợp

những thuộc tính của virus tệp tin và virus khởi động. Những ví dụ tiêu biểu cho loại virus này là Flip và Invader.

Ngoài ra, các virus biên dịch có thể cư trú trong bộ nhớ thi hành của hệ thống máy tính đã bị lây nhiễm và chiếm quyền điều khiển tệp tin của hệ điều hành. Do đó, khi hệ điều hành chạy một chương trình chưa bị lây nhiễm, hệ điều hành do bị virus chiếm quyền từ trước sẽ không thi hành ngay chương trình mà tiến hành lây nhiễm lên tệp tin thi hành đó trước. Với cách hoạt động như vậy, virus biên dịch còn được gọi là virus cư trú trong bộ nhớ (memory resident virus).

Virus thông dịch:

Trái ngược với virus biên dịch, virus thông dịch chứa đựng mã nguồn chương trình và chỉ được thi hành bởi một ứng dụng hay dịch vụ cụ thể nào đó. Do vậy, virus thông dịch trở nên phổ biến bởi đặc tính rất dễ để viết và sửa chữa. Một tin tặc với ít kỹ năng cũng có thể tìm kiếm trên mạng một virus thông dịch, đọc, sửa chữa và phát tán. Do đó, trên mạng có thể có hàng tá những biến thể khác nhau của cùng một virus thông dịch, hầu hết những khác biệt là rất nhỏ.

Các virus thông dịch có thể được chia làm hai loại chính là virus macro và virus script. Virus macro là loại virus thông dịch khá phổ biến. Virus macro bám vào các tệp tin tài liệu chẳng hạn như các tệp tin văn bản, các bảng tính và sử dụng các trình thông dịch ngôn ngữ macro của ứng dụng để thi hành và lây lan. Một số loại phần mềm thường hỗ trợ người dùng viết các macro đơn giản để tự động hóa những nhiệm vụ phức tạp có tính chất lặp, đây chính là cơ sở cho các virus macro hoạt động. Một ví dụ tiêu biểu là ứng dụng Microsoft Office, một phần mềm được sử dụng rộng rãi và cho phép người sử dụng tạo các macro bằng ngôn ngữ VB.Script. Hơn nữa, đặc tính thường xuyên chia sẻ các tài liệu ứng dụng chính là cơ sở cho phép các virus macro lây lan nhanh chóng. Ngoài ra, trong trường hợp ứng dụng hỗ trợ template (template là một khuôn mẫu được trình ứng dụng sử dụng để mở hoặc tạo mới tệp tin), virus macro thường lây nhiễm các tệp tin template. Một khi tệp tin template đã bị lây nhiễm, mọi tài liệu được tạo hoặc mở với template sẽ bị lây nhiễm. Một số virus macro tiêu biểu có thể kể tới là Cocept, Marker và Melissa.

Virus script về cơ bản không có gì khác biệt nhiều với virus macro. Sự khác biệt chính yếu là virus macro được biết bởi ngôn ngữ được hiểu bởi một ứng dụng cụ thể, do đó tính lây lan chỉ hạn chế trong phạm vi ứng dụng đó, ngược lại virus script lại được viết bởi những ngôn ngữ được hiểu bởi một dịch vụ nào đó, chạy bởi hệ điều hành. Chẳng hạn Windows Scripting Host là một dịch vụ của một vài phiên bản hệ điều hành Microsoft Windows có thể thi hành các kịch bản được viết bằng VBScript, do đó một virus script viết bằng VBScript có thể lây nhiễm trên mọi hệ thống máy tính cài đặt hệ điều hành Windows có hỗ trợ Windows Scripting Host. First và Love Stages là những ví dụ tiêu biểu cho virus script.

b. Sâu máy tính

Khác với virus là những chương trình, những đoạn mã lệnh sống bám trên một tệp tin nào đó và chỉ hoạt động khi tệp tin đó hoạt động, sâu máy tính (worm, thường được gọi tắt là sâu) là một chương trình hoàn chỉnh độc lập có khả năng nhân bản và di chuyển từ hệ thống máy tính này sang hệ thống máy tính khác. Do là một chương trình độc lập, sâu có thể tự nhân bản mà không cần chờ đợi sự kích hoạt của vật chủ như cơ chế nhân bản và lây lan của virus, chính ưu điểm này đã cho phép sâu lây lan với tốc độ nhanh hơn và được các tin tặc ưa chuộng hơn. Các sâu máy tính tận dụng mạng Internet để lây lan trên phạm vi lớn. Thông thường các sâu khai thác những lỗ hổng đã được biết, những điểm yếu trong cấu hình, các hệ thống thư điện tử và các hệ thống chia sẻ tệp tin để lây lan.

Hầu hết các sâu máy tính được định hướng để tiêu thụ tài nguyên của máy tính và tài nguyên mạng. Tuy nhiên, một số sâu được sử dụng để cài đặt backdoor (chi tiết được đưa ra trong mục c) cho phép thực hiện một tấn công từ chối dịch vụ từ xa tới một máy chủ nào đó, hoặc xử lý những hoạt động nguy hiểm khác trên hệ thống máy tính. Các sâu máy tính được chia làm hai loại chính là sâu dịch vụ mạng (network service worm) và sâu thư điện tử (mass mailing worm), tương ứng với hai hình thức lan truyền chính của sâu.

- Sâu dịch vụ mạng là những sâu máy tính lan truyền bằng cách khai thác những lỗ hổng trong một dịch vụ mạng gắn kết với hệ điều hành hoặc một ứng dụng nào đó. Sau khi sâu lây nhiễm vào hệ thống, nó thường sử dụng hệ thống đó để tìm kiếm những hệ thống khác đang chạy dịch vụ tương tự và tìm cách lây nhiễm vào các hệ thống đó. Hình thức lan truyền này hoàn toàn không cần bất kỳ sự tác động nào của người sử dụng, nên sâu dịch vụ mạng thường lan truyền nhanh hơn các loại phần mềm độc hại khác. Sasser và Witty là hai ví dụ cho sâu dịch vụ mạng. Sasser xuất hiện trong tháng 4 năm 2004, khai thác lỗ hổng tràn bộ đệm trong dịch vụ LSASS (Local Security Authority Subsystem Service) trên các hệ điều hành Windows (XP và 2000). Cũng trong năm 2004, sâu Witty lại khai thác lỗ hổng trên một loạt sản phẩm bảo mật mạng cụ thể là Internet Security Systems (hiện tại lấy tên là IBM Internet Security Systems). Witty có tốc độ lan truyền một cách khủng khiếp, chỉ với nửa giờ đã lây lan sang 12 ngàn máy tính và sinh ra lưu thông mạng lên tới 90 Gbits/s.

- Sâu thư điện tử là những sâu máy tính thực hiện lan truyền dựa trên cơ chế phát tán thư điện tử. Về cơ bản, sâu thư điện tử có cơ chế lan truyền giống với những virus phát tán qua thư điện tử. Tuy nhiên, với virus thì tệp tin đính kèm trong thư điện tử đã bị nhiễm virus nhưng với sâu thì tệp tin đính kèm là một chương trình sâu hoàn chỉnh. Một khi sâu thư điện tử đã lây nhiễm một hệ thống, nó sẽ tự động tìm kiếm trên hệ thống những địa chỉ thư điện tử và sau đó gửi một bản copy của nó tới những địa chỉ đó. Để gửi thư, nó có thể sử dụng hệ thống thư điện tử trên máy nạn nhân hoặc sử dụng một chức năng gửi thư đơn giản được chứa sẵn trong bản thân nó. Các sâu thư điện tử thường gửi một bản copy của nó tới nhiều người nhận một lần. Bên cạnh việc làm tràn ngập các hệ thống máy chủ thư điện tử và mạng bằng một lượng lớn các thư điện tử được gửi qua lại, các sâu thư điện tử cũng có thể là nguyên nhân gây nên những vấn đề hiệu năng cho hệ thống bị lây nhiễm. Beagle là một sâu thư điện tử xuất hiện năm 2004 có thể lây nhiễm trên tất cả các phiên bản của hệ điều hành Windows. Chủng đầu tiên Beagle.A không lây nhiễm rộng rãi. Tuy nhiên, một biến thể có nó là Beagle.B lại lây lan rộng rãi và rất nguy hiểm. Beagle chứa trong nó một cài đặt giao thức SMTP riêng để gửi thư tới các địa chỉ mà nó thu thập được trên máy tính đã bị lây nhiễm. Beagle cũng mở một backdoor trên cổng 6777 (Beagle.A) và 8866 (Beagle.B). Ngoài ra, Mydoom và Nestky cũng là những sâu thư điện tử tiêu biểu.

c. Trojan

Trojan được lấy tên theo tên con ngựa gỗ trong truyền thuyết Trojan Horse (Con ngựa thành Troa) trong thần thoại Hy Lạp. Trojan sử dụng một chiến lược khác biệt hoàn toàn với virus và sâu, đó là nó hoàn toàn không có khả năng nhân bản, Trojan thường tỏ ra vô hại, thậm chí là có lợi cho người dùng nhưng ẩn trong nó là những mục đích xấu. Trojan là một phần mềm hoàn chỉnh có thể được cài đặt theo các lỗ hổng an ninh vào máy tính do sự sơ suất của người dùng khi truy cập mạng máy tính. Trojan cũng có thể núp danh một phần mềm tiện ích và được người dùng cài đặt một cách bình thường. Việc sử dụng các phần mềm không bản quyền, được download từ những nguồn không rõ xuất xứ là nơi cư ngụ của rất nhiều Trojan.

Khi được cài đặt vào hệ thống máy tính, các chức năng xấu được tiến hành một cách âm thầm bên dưới sự hoạt động bình thường của các tiện ích thông thường mà nó cung cấp. Nhiều Trojan được định hướng để thay thế những tệp tin thi hành tồn tại trên hệ thống bằng những

phiên bản độc hại (hoặc có lỗ hổng hoặc bị nhiễm virus), hoặc tự động cài đặt những ứng dụng khác tới hệ thống. Tuy nhiên, phần lớn Trojan đóng vai trò như một gián điệp trong hệ thống máy tính. Tùy theo mục đích khác nhau mà các Trojan có tên khác nhau, dưới đây là một số loại Trojan thường gặp:

- Spyware (phần mềm gián điệp) đóng vai trò là gián điệp, nó thu thập những thông tin cần thiết trên hệ thống bị lây nhiễm và gửi thông tin đó tới một hệ thống nào đó.

- Adware (phần mềm quảng cáo) đóng vai trò quảng cáo, nó thường hoạt động bằng cách bật những quảng cáo trên hệ thống bị lây nhiễm.

- Key logger có nhiệm vụ ghi lại các phím đã được gõ trên bàn phím và gửi tới hệ thống phân tích nào đó bên ngoài.

- Backdoor (cửa hậu) có nhiệm vụ mở ra một cổng sau để tin tặc có thể khai thác hệ thống máy tính bị lây nhiễm.

- Rootkit được sử dụng để thu thập các tệp tin được cài đặt lên hệ thống và thay thế chúng, việc thay thế các tệp tin của rootkit đôi khi gây ra những hậu quả rất lớn. Tuy nhiên, nhiều trường hợp những thay đổi này nhằm che dấu một cuộc tấn công hay sự hoạt động của một phần mềm độc hại nào đó hay xóa bỏ những bằng chứng về sự hiện diện của chính bản thân rootkit. Những rootkit tiêu biểu có thể kể tới như LRK5, Knark, Adore và Hacker Defender.

Trojan thường khó để phát hiện, bởi vì chúng được thiết kế để che dấu sự tồn tại trên hệ thống và xử lý những chức năng có vẻ hợp lý, nên người sử dụng và quản trị hệ thống thường không phát hiện ra.

7.2.3. Tấn công từ chối dịch vụ

Ngày nay Internet được sử dụng trong hầu hết mọi khía cạnh của đời sống, nó trở thành một tài nguyên quan trọng có sức ảnh hưởng lớn. Do vậy việc phá hỏng nguồn tài nguyên Internet tạm thời dù chỉ là trong ít phút cũng có thể gây mất mát to lớn về tài chính, thậm chí gây xáo động đời sống con người. Cũng bởi thế nó nhanh chóng trở thành đối tượng đặc biệt quan tâm của giới tin tặc cũng như giới tội phạm. Sự thật là đã có nhiều đợt tấn công vào nguồn tài nguyên này trên phạm vi quốc gia và thế giới. Ngày 7 tháng 2 năm 2000, một hacker tuổi thiếu niên với nickname "mafiaboy" đã làm tê liệt website của yahoo trong gần 3 tiếng đồng hồ. Hai ngày sau, sáu website thương mại nổi tiếng khác như Amazon, CNN, Ebay, E*Trade và ZDNet cũng trở thành nạn nhân của mafiaboy. Năm 2001, sâu máy tính Code Red tấn công website của Nhà Trắng làm ảnh hưởng tới an ninh quốc gia của Mỹ, hay năm 2003, một đợt tấn công đã hạ gục Houston port system ở Texas đe dọa an ninh công cộng. Gần đây nhất, trong tháng 3 năm 2013, Spamhaus – tổ chức vốn chịu trách nhiệm duy trì danh sách đen các máy chủ chuyên gửi thư rác trên toàn cầu đang phải hứng chịu cuộc tấn công từ chối dịch vụ lớn chưa từng có với lưu lượng ở mức đỉnh lên tới 300 Gbits/s. Cuộc tấn công này đã làm trì trệ đường truyền mạng toàn thế giới trong đó khu vực châu Âu chịu ảnh hưởng nặng nề nhất.

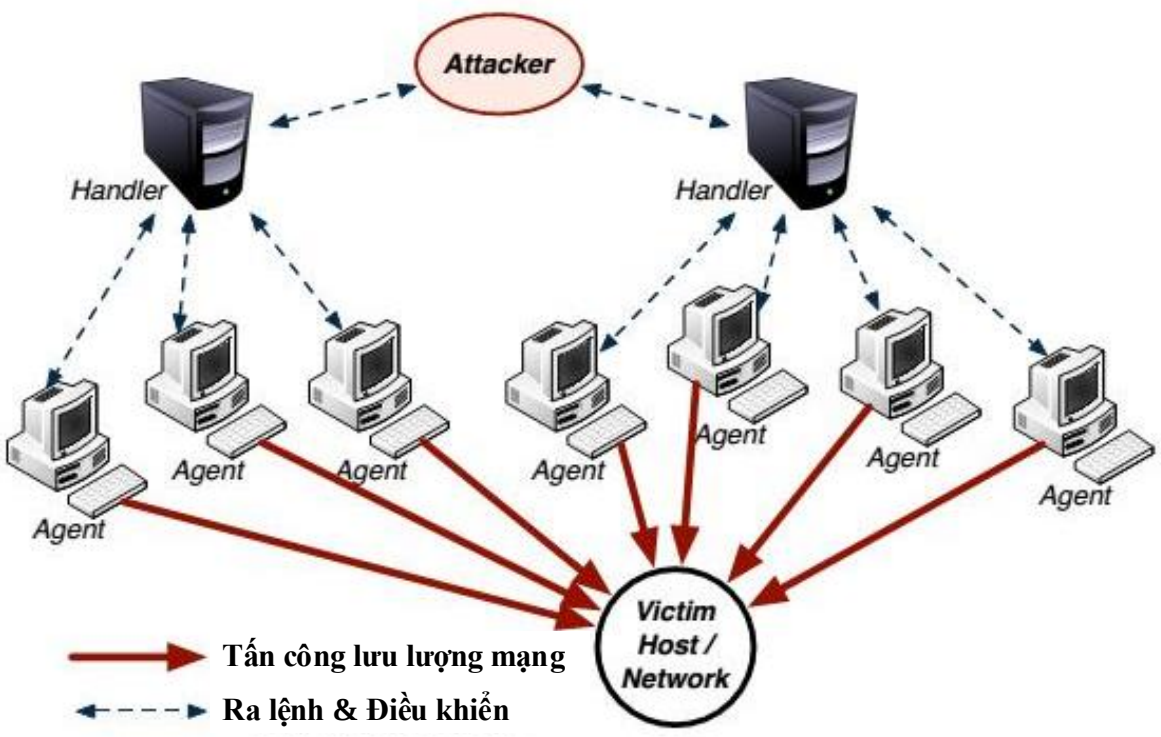
Những tấn công với mục tiêu là làm tê liệt các hệ thống máy tính hay các dịch vụ được gọi là tấn công từ chối dịch vụ - Denial of Service (DoS). Tấn công từ chối dịch vụ có thể được phân thành hai kiểu chính là tấn công dựa trên lỗ hổng phần mềm (vulnerability-based attack) và tấn công làm ngập lụt (flooding attack).

- Tấn công dựa trên lỗ hổng phần mềm, hay còn gọi là tấn công ngữ nghĩa (semantic attack), là hình thức tấn công khai thác một hoặc nhiều lỗ hổng trong chính sách an ninh hoặc trong kỹ thuật nhằm hiệu lực chính sách đó, hoặc những lỗi tiềm ẩn trong phần mềm. Lợi dụng những lỗ hổng này, tin tặc chỉ cần gửi tới hệ thống một vài yêu cầu đặc biệt, những yêu cầu này

sẽ tiêu thụ một lượng lớn tài nguyên của hệ thống và khiến hệ thống tê liệt. Ví dụ tiêu biểu của hình thức tấn công này là Ping-of-Death xuất hiện trong năm 1996, tin tặc đơn giản gửi tới hệ điều hành một yêu cầu theo giao thức ICMP (Internet Control Message Protocol) với kích thước lớn vượt mức cho phép.

- Tấn công làm ngập lụt, hay còn gọi là brute-force attack, là hình thức tấn công từ chối dịch vụ bằng cách tạo ra một một lượng lớn yêu cầu hợp lệ (thường là giống nhau) nhằm tiêu thụ một tài nguyên mục tiêu nào đó trên hệ thống khiến tài nguyên đó bị quá tải không thể đáp ứng những yêu cầu đến từ những người dùng hợp lệ khác. Một ví dụ tiêu biểu là tấn công UDP (User Datagram Protocol), tin tặc sẽ gửi một lượng lớn gói tin UDP tới các cổng ngẫu nhiên của một máy chủ nào đó, điều này dẫn tới tiêu thụ hết băng thông của máy chủ, do đó người sử dụng bình thường sẽ không thể truy cập được vào máy chủ đó.

Để thực hiện một cuộc tấn công từ chối dịch vụ, tin tặc có thể sử dụng một hoặc nhiều máy chủ để tấn công. Khi những yêu cầu nhằm mục đích tấn công của tin tặc đến từ nhiều máy tính khác nhau được phân tán trên mạng, thì được gọi là tấn công từ chối dịch vụ phân tán – distributed denial of service (DDoS). Ngược lại, khi những yêu cầu nhằm mục đích tấn công này đến từ cùng một máy chủ thì được gọi là tấn công từ chối dịch vụ đơn nguồn - single-source denial of service (SDoS). Trong hầu hết các tài liệu, thuật ngữ DoS được sử dụng thay thế cho SDoS.



Hình 7.1. Mô hình tấn công từ chối dịch vụ phân tán

Thông thường các tấn công DDoS sử dụng hai kiểu thành phần là agent (máy tác tử) và handler (máy điều khiển) như mô tả trong hình 7.1. Các Agent đơn giản là các máy tính có thể được điều khiển bởi handler để sinh ra các yêu cầu tấn công và gửi yêu cầu tấn công tới máy chủ nạn nhân. Các handler đơn giản là các máy tính có cài đặt chương trình nhằm điều khiển các agent. Handler có nhiệm vụ thông báo cho các agent biết khi nào thực hiện tấn công, tấn công

mục tiêu nào và tấn công như thế nào. Các Agent cũng còn được gọi là các bot hay zombie (âm binh) và một tập hợp các agent được điều khiển bởi cùng một tin tặc được gọi là botnet (mạng ma).

Ở Việt Nam, tấn công từ chối dịch vụ cũng nổi lên mạnh mẽ trong khoảng thời gian gần đây. Tháng 4 năm 2006, lần đầu tiên Cục C15 của Bộ Công an đã bắt giữ một tin tặc tấn công từ chối dịch vụ làm tê liệt nhiều website thương mại của Việt Nam. Năm 2010-2011, Vietnamnet đã phải liên tục hứng chịu các cuộc tấn công từ chối dịch vụ quy mô lớn, đợt đầu từ cuối năm 2010 tới đầu năm 2011 và đợt 2 trong khoảng tháng 8 và 9 của năm 2011. Đỉnh điểm là ngày 27 tháng 1 năm 2011, Vietnamnet phải xử lý 1,5 triệu kết nối vào cùng một thời điểm. Gần đây nhất, trong tháng 7 năm 2013, các báo điện tử của Việt Nam như Dân trí, Vietnamnet, Tuổi trẻ cũng bị tấn công từ chối dịch vụ khiến dịch vụ thi thoảng bị ngắt quãng.

7.2.4. Lừa đảo

Lừa đảo (Phishing) là hình thức trong đó kẻ tấn công (hay kẻ lừa đảo – phisher) tìm cách để chiếm đoạt thông tin bí mật hoặc những ủy nhiệm nhạy cảm của người sử dụng một cách khéo léo. Thuật ngữ phishing lần đầu tiên xuất hiện vào năm 1995, khi những kẻ lừa đảo trên mạng Internet sử dụng thư điện tử làm mồi nhử để chiếm đoạt username, password và các thông tin tài chính từ hàng triệu người sử dụng dịch vụ mail trên Internet. Một ví dụ tiêu biểu là vụ lừa đảo dựa trên website AOL (AOL là một website thương mại lớn của Mỹ), những kẻ lừa đảo đã tạo ra một trang web giả danh AOL và gửi một loạt thư hoặc tin nhắn lừa đảo nhằm dụ người nhận mở một liên kết tới trang AOL giả nhằm chiếm đoạt username và password.

Sự phát triển của thương mại điện tử, đã khiến phishing trở thành tâm điểm của giới tội phạm tài chính trên mạng. Theo báo cáo của RSA Anti-Fraud Command Center (AFCC) đầu năm 2013, tổng số cuộc tấn công lừa đảo của năm 2012 là 445.004, cao hơn năm 2011 (258.461 cuộc) tới 59% và tổng thiệt hại khoảng 1,5 tỉ đô la.

Song song với sự phát triển về số lượng, là sự đa dạng và tinh xảo của các kỹ thuật phishing. Ngày nay Phishing không chỉ thực hiện qua thư điện tử mà còn được thực hiện trên nhiều hình thức khác như VOIP, SMS, instant messaging, các trang mạng xã hội và những trò chơi trực tuyến nhiều người chơi. Ở đây ta xem xét một vài loại phishing phổ biến.

a. Lừa đảo dùng bản sao (Clone phishing)

Trong phương thức này, kẻ lừa đảo tạo ra một thư nhân bản từ một bức thư hợp lệ nào đó. Nội dung bức thư hợp lệ này chứa đựng nội dung và địa chỉ người nhận, đơn giản nó là một bức thư hợp lệ được gửi trước đó, sau đó kẻ lừa đảo tiến hành thay thế các liên kết trong mail, để trỏ tới một địa chỉ giả nào đó với mục đích lừa đảo. Cuối cùng bức thư nhân bản được gửi bằng địa chỉ giả, do đó người nhận sẽ thấy thư được gửi đến từ người gửi hợp lệ nào đó. Ngoài ra nội dung thư có thể đề cập tới thông tin như việc gửi lại hoặc một phiên bản cập nhật của thư ban đầu, như một chiến thuật để lừa người nhận.

b. Lừa đảo hướng đối tượng (Spear Phishing)

Lừa đảo hướng đối tượng khác với hình thức lừa đảo thông thường là nó nhắm tới một nhóm cụ thể. Thay vì tiến hành gửi thư cho các mục tiêu ngẫu nhiên, Lừa đảo hướng đối tượng sẽ lựa chọn các nhóm người với một số điểm chung nào đó chẳng hạn những người thuộc cùng một tổ chức. Kẻ lừa đảo cũng lợi dụng những thông tin chung đó để tạo lòng tin với người nhận. Cuối năm 2010 và đầu năm 2011, nạn nhân của lừa đảo hướng đối tượng là Văn phòng thủ tướng chính phủ Úc, Chính phủ Canda, HBGary Federal và Thư viện quốc gia Oak Ridge.

c. Lừa đảo dùng điện thoại (Phone Phishing)

Đối với hình thức tấn công này, kẻ lừa đảo gửi tin nhắn thông báo với nội dung ngân hàng yêu cầu người sử dụng gọi điện tới một số máy cụ thể để giải quyết một vài vấn đề gì đó về tài khoản của họ. Kẻ lừa đảo sẽ tận dụng kỹ thuật VOIP (Voice over IP) để nắm bắt cuộc gọi và tiến hành lừa đảo.

7.3. SỞ HỮU TRÍ TUỆ

Những sản phẩm của ngành công nghệ thông tin là những sản phẩm đặc biệt, chúng là những sản phẩm trí tuệ và rất dễ để sao chép, nhân bản. Do đó, để tạo ra một môi trường công nghiệp công nghệ thông tin một cách lành mạnh cũng như thúc đẩy những sáng tạo cá nhân trong lĩnh vực này rất cần những điều luật nhằm đảm bảo quyền lợi cho các tác giả của những sản phẩm công nghệ thông tin. Những điều luật này được đặt trong hệ thống các điều luật về sở hữu trí tuệ. Mục này sẽ giới thiệu những khái niệm cơ bản về sở hữu trí tuệ và vấn đề vi phạm sở hữu trí tuệ trong ngành công nghệ thông tin.

7.3.1. Tài sản trí tuệ

Tài sản trí tuệ là các thành quả sáng tạo của cá nhân hoặc tổ chức. Theo tổ chức Sở hữu Trí tuệ Thế giới (WIPO), các tài sản trí tuệ được chia thành 2 loại chính:

- Tác phẩm: tác phẩm mang tính văn chương (thơ, tiểu thuyết, kịch, truyện, sách tham khảo, báo), tác phẩm nghệ thuật (tranh, ảnh, phim, ca khúc, điêu khắc, vở múa, quảng cáo), bản vẽ kiến trúc, phần mềm, cơ sở dữ liệu, chương trình ti vi, radio...

- Tài sản trí tuệ trong công nghiệp: sáng chế, thiết kế kiểu dáng công nghiệp, thương hiệu, bí mật kinh doanh, mạch tích hợp, chỉ dẫn địa lí...

Mỗi tài sản trí tuệ đều được thể hiện thông qua một phương tiện vật lí cụ thể nào đó. Chẳng hạn một tiểu thuyết có thể được thể hiện thông qua một bản thảo viết tay, một tệp tin văn bản trên máy tính, hay giọng đọc của phát thanh viên trên đài... Giá trị cốt lõi của tài sản trí tuệ không nằm ở phương tiện vật lí thể hiện mà ở ý tưởng sáng tạo chứa đựng trong nó. Do đó, cần phải có sự phân biệt rõ ràng giữa một tài sản trí tuệ với những biểu hiện vật lí cụ thể của tài sản trí tuệ đó.

7.3.2. Quyền sở hữu trí tuệ

Mỗi tài sản đều gắn với chủ sở hữu nhất định. Quyền sở hữu khẳng định những quyền lợi của chủ sở hữu với tài sản, bao gồm quyền sử dụng, quyền sửa đổi, quyền chuyển nhượng... Quyền sở hữu đối với tài sản trí tuệ được gọi là quyền sở hữu trí tuệ.

Như đã chỉ ra trong mục 7.3.1, tài sản trí tuệ được chia là hai loại là tác phẩm và tài sản trí tuệ công nghiệp. Quyền sở hữu đối với tác phẩm còn được gọi với thuật ngữ khác là quyền tác giả hay bản quyền. Quyền sở hữu đối với tài sản trí tuệ công nghiệp được gọi là quyền sở hữu công nghiệp.

7.3.3. Luật sở hữu trí tuệ

a. Vì sao phải bảo hộ quyền sở hữu trí tuệ

Tài sản trí tuệ được tạo ra thông những qua hoạt động của cá nhân hoặc tổ chức, do đó trước hết nó phải dưới sự sở hữu của cá nhân hoặc tổ chức đó. Ngoài ra, các tài sản trí tuệ rất dễ bị xâm phạm vì chúng dễ bị sao chép và đánh cắp. Do đó, bảo hộ quyền sở hữu trí tuệ là rất cần thiết.

Bảo hộ quyền sở hữu trí tuệ sẽ giúp đảm bảo quyền lợi cho các cá nhân hoặc tổ chức trong việc tạo ra các sản phẩm trí tuệ. Điều này giúp tạo môi trường cạnh tranh bình đẳng cho các tổ chức cũng như lợi ích và động lực cho các cá nhân. Từ đó thúc đẩy hoạt động sáng tạo, tạo đà cho việc phát triển những sản phẩm trí tuệ mới phục vụ nhu cầu ngày càng cao của con người. Ví dụ, trong ngành công nghệ thông tin, một sản phẩm phần mềm có thể là kết tinh của rất nhiều thành viên trong đội phát triển phần mềm, dưới sự hỗ trợ về tài chính của một công ty hay tổ chức nào đó. Nếu như sản phẩm phần mềm đó không được bảo hộ, mà bị sao chép một cách trái phép, sẽ khiến tổ chức hoặc công ty gặp khó khăn về tài chính, hậu quả là sẽ không có những đầu tư để nghiên cứu phát triển sản phẩm mới cũng như cải tiến sản phẩm cũ.

b. Luật Sở hữu trí tuệ

Luật Sở hữu trí tuệ là văn bản pháp lý được đề ra nhằm bảo hộ cho quyền sở hữu trí tuệ. Ở Việt Nam, Luật Sở hữu trí tuệ được ban hành vào năm 2005 và chính thức có hiệu lực từ ngày 1/7/2006. Bất kỳ cá nhân hoặc tổ chức nào xâm phạm quyền sở hữu trí tuệ của một cá nhân hoặc tổ chức khác, nếu có khiếu kiện sẽ được xử lý theo luật định.

Tuy nhiên, một văn bản luật được ban hành chỉ có giá trị trong phạm vị một quốc gia, nên chỉ đảm bảo được quyền lợi cho chủ sở hữu trong phạm vị quốc gia đó. Nếu sự vi phạm xảy ra ngoài phạm vi của quốc gia đó thì luật sở hữu trí tuệ không can thiệp được. Vì lý do đó, những hiệu ước và công ước quốc tế được thành lập. Việt Nam hiện tại đã ra nhập Công ước Berne và Hiệp định TRIPS (Hiệp định về các khía cạnh thương mại của sở hữu trí tuệ - Trade Related Aspects of Intellectual Property Rights).

Việc bảo hộ quyền sở hữu rõ ràng đem lại ích lợi to lớn cho các chủ sở hữu, tuy nhiên nếu quản lý quá chặt và không hợp lý có thể làm cho tài sản trí tuệ không tiếp cận được với cộng đồng, không phát huy được sức mạnh vốn có của tài sản trí tuệ. Để khắc phục nhược điểm này khái niệm sở hữu công và ngoại lệ (fair use) được đưa ra.

Thường thì quyền sở hữu trí tuệ quy định quyền lợi của cá nhân hoặc tổ chức đối với tài sản trí tuệ, tài sản trí tuệ lúc này được hiểu như một tài sản riêng của cá nhân hoặc tổ chức cụ thể. Tuy nhiên, có một số tài sản trí tuệ lại thuộc quyền sở hữu của tất cả mọi người mà không thuộc về riêng cá nhân hay tổ chức nào, được gọi là sở hữu công (public domain). Một tài sản trí tuệ được liệt vào mục sở hữu công nếu:

- Tài sản trí tuệ đó được tạo ra bởi cộng đồng chẳng hạn như Tiếng Anh, dân ca quan họ Bắc Ninh.
- Tài sản trí tuệ là chân lí, sự thật tuy được khám phá bởi cá nhân nhưng không thể đặt quyền sở hữu cá nhân lên được vì sẽ cản trở sự phát triển của nhân loại. Ví dụ: định luật Newton, thuyết tiến hóa Darwin.
- Tài sản trí tuệ thuộc sở hữu cá nhân nhưng đã hết hạn, chẳng hạn như kịch Shakespeare, nhạc Beethoven, đèn điện Edison. Việc giới hạn thời gian hiệu lực của quyền sở hữu cá nhân có ý nghĩa vô cùng quan trọng giúp tài sản trí tuệ dễ dàng được tiếp cận với cộng đồng, tạo điều kiện cho cộng đồng khai thác và gia tăng giá trị cho tài sản trí tuệ đó.

Trong một số trường hợp, việc xin phép hoặc mua bản quyền thay vì đem lại lợi ích nó có thể gây phiền phức cho cả người khai thác lẫn chủ sở hữu của tài sản trí tuệ. Ví dụ như để trích dẫn một câu nói hay một đoạn văn bản trong một tác phẩm nào đó cho bài giảng của mình mà giảng viên lại phải đi xin phép hay mua lại của chủ sở hữu thì sẽ trở nên quá phức tạp. Với những trường hợp như vậy, ngoại lệ fair use (sử dụng hợp lý) có thể được áp dụng. Ngoại lệ fair use cho phép mọi người khai thác tài sản trí tuệ của người khác mà không cần xin phép với điều kiện:

- Có mục đích đẹp: Người dùng sử dụng/trích dẫn tác phẩm vào mục đích giáo dục, nghiên cứu, nhân văn hoặc đưa tin thời sự, bình luận nhưng không được lạm dụng bằng cách sử dụng quá nhiều hoặc dùng để thu lời tài chính.
- Biết ơn tác giả: Khi sử dụng, trích dẫn phải nêu lại tên người giữ bản quyền/tác giả và tên tác phẩm.

7.4. CÁC QUY ĐỊNH, ĐIỀU LUẬT VỀ AN TOÀN THÔNG TIN VÀ SỞ HỮU TRÍ TUỆ

Khi xuất hiện những hình thức cư xử, những hành vi có thể gây hại tới đời sống bình thường của cộng đồng, những chế tài pháp luật sẽ được đưa ra nhằm hạn chế và loại bỏ những hành vi, cư xử này. Tương tự như vậy, với các hình thức phạm tội mới liên quan tới tin học và hạ tầng công nghệ thông tin, một số chế tài pháp luật cụ thể phải được ban hành.

7.4.1. Các điều trong Bộ luật hình sự

Dưới đây là một số điều luật chống tội phạm tin học được đưa ra trong Bộ luật hình sự của Việt Nam năm 1999, được sửa đổi năm 2009.

Điều 224. Tội phát tán vi rút, chương trình tin học có tính năng gây hại cho hoạt động của mạng máy tính, mạng viễn thông, mạng Internet, thiết bị số

1. Người nào cố ý phát tán vi rút, chương trình tin học có tính năng gây hại cho mạng máy tính, mạng viễn thông, mạng Internet, thiết bị số gây hậu quả nghiêm trọng, thì bị phạt tiền từ hai mươi triệu đồng đến hai trăm triệu đồng hoặc phạt tù từ một năm đến năm năm.

2. Phạm tội thuộc một trong các trường hợp sau đây, thì bị phạt tù từ ba năm đến bảy năm:

a) Có tổ chức;

b) Gây hậu quả rất nghiêm trọng;

c) Tái phạm nguy hiểm.

3. Phạm tội thuộc một trong các trường hợp sau đây, thì bị phạt tù từ năm năm đến mười hai năm:

a) Đối với hệ thống dữ liệu thuộc bí mật nhà nước; hệ thống thông tin phục vụ an ninh, quốc phòng;

b) Đối với cơ sở hạ tầng thông tin quốc gia; hệ thống thông tin điều hành lưới điện quốc gia; hệ thống thông tin tài chính, ngân hàng; hệ thống thông tin điều khiển giao thông;

c) Gây hậu quả đặc biệt nghiêm trọng.

4. Người phạm tội còn có thể bị phạt tiền từ năm triệu đồng đến năm mươi triệu đồng, cấm đảm nhiệm chức vụ, cấm hành nghề hoặc làm công việc nhất định từ một năm đến năm năm.

Điều 225. Tội cản trở hoặc gây rối loạn hoạt động của mạng máy tính, mạng viễn thông, mạng Internet, thiết bị số

1. Người nào thực hiện một trong các hành vi sau đây gây hậu quả nghiêm trọng nếu không thuộc trường hợp quy định tại Điều 224 và Điều 226a của Bộ luật này, thì bị phạt tiền từ hai mươi triệu đồng đến hai trăm triệu đồng hoặc phạt tù từ một năm đến năm năm:

a) Tự ý xoá, làm tổn hại hoặc thay đổi phần mềm, dữ liệu thiết bị số;

b) Ngăn chặn trái phép việc truyền tải dữ liệu của mạng máy tính, mạng viễn thông, mạng Internet, thiết bị số;

c) Hành vi khác cản trở hoặc gây rối loạn hoạt động của mạng máy tính, mạng viễn thông, mạng Internet, thiết bị số.

2. Phạm tội thuộc một trong các trường hợp sau đây, thì bị phạt tù từ ba năm đến bảy năm:

a) Có tổ chức;

b) Lợi dụng quyền quản trị mạng máy tính, mạng viễn thông, mạng Internet;

c) Gây hậu quả rất nghiêm trọng.

3. Phạm tội thuộc một trong các trường hợp sau đây, thì bị phạt tù từ năm năm đến mười hai năm:

a) Đối với hệ thống dữ liệu thuộc bí mật nhà nước; hệ thống thông tin phục vụ an ninh, quốc phòng;

b) Đối với cơ sở hạ tầng thông tin quốc gia; hệ thống thông tin điều hành lưới điện quốc gia; hệ thống thông tin tài chính, ngân hàng; hệ thống thông tin điều khiển giao thông;

c) Gây hậu quả đặc biệt nghiêm trọng.

4. Người phạm tội còn có thể bị phạt tiền từ năm triệu đồng đến năm mươi triệu đồng, cấm đảm nhiệm chức vụ, cấm hành nghề hoặc làm công việc nhất định từ một năm đến năm năm.”

Điều 226. Tội đưa hoặc sử dụng trái phép thông tin trên mạng máy tính, mạng viễn thông, mạng Internet

1. Người nào thực hiện một trong các hành vi sau đây xâm phạm lợi ích của cơ quan, tổ chức, cá nhân, xâm phạm trật tự, an toàn xã hội gây hậu quả nghiêm trọng, thì bị phạt tiền từ mười triệu đồng đến một trăm triệu đồng, cải tạo không giam giữ đến ba năm hoặc bị phạt tù từ sáu tháng đến ba năm:

a) Đưa lên mạng máy tính, mạng viễn thông, mạng Internet những thông tin trái với quy định của pháp luật, nếu không thuộc trường hợp quy định tại Điều 88 và Điều 253 của Bộ luật này;

b) Mua bán, trao đổi, tặng cho, sửa chữa, thay đổi hoặc công khai hoá những thông tin riêng hợp pháp của cơ quan, tổ chức, cá nhân khác trên mạng máy tính, mạng viễn thông, mạng Internet mà không được phép của chủ sở hữu thông tin đó;

c) Hành vi khác sử dụng trái phép thông tin trên mạng máy tính, mạng viễn thông, mạng Internet.

2. Phạm tội thuộc một trong các trường hợp sau đây, thì bị phạt tù từ hai năm đến bảy năm:

a) Có tổ chức;

b) Lợi dụng quyền quản trị mạng máy tính, mạng viễn thông, mạng Internet;

c) Thu lợi bất chính từ một trăm triệu đồng trở lên;

d) Gây hậu quả rất nghiêm trọng hoặc đặc biệt nghiêm trọng.

3. Người phạm tội còn có thể bị phạt tiền từ hai mươi triệu đồng đến hai trăm triệu đồng, cấm đảm nhiệm chức vụ, cấm hành nghề hoặc làm công việc nhất định từ một năm đến năm năm.

Điều 226a. Tội truy cập bất hợp pháp vào mạng máy tính, mạng viễn thông, mạng Internet hoặc thiết bị số của người khác

1. Người nào cố ý vượt qua cảnh báo, mã truy cập, tường lửa, sử dụng quyền quản trị của người khác hoặc bằng phương thức khác truy cập bất hợp pháp vào mạng máy tính, mạng viễn thông, mạng Internet hoặc thiết bị số của người khác chiếm quyền điều khiển; can thiệp vào chức năng hoạt động của thiết bị số; lấy cắp, thay đổi, hủy hoại, làm giả dữ liệu hoặc sử dụng trái phép các dịch vụ, thì bị phạt tiền từ hai mươi triệu đồng đến hai trăm triệu đồng hoặc phạt tù từ một năm đến năm năm.

2. Phạm tội thuộc một trong các trường hợp sau đây, thì bị phạt tù từ ba năm đến bảy năm:

a) Có tổ chức;

b) Lợi dụng chức vụ, quyền hạn;

c) Thu lợi bất chính lớn;

d) Gây hậu quả nghiêm trọng;

đ) Tái phạm nguy hiểm.

3. Phạm tội thuộc một trong các trường hợp sau đây, thì bị phạt tù từ năm năm đến mười hai năm:

a) Đối với hệ thống dữ liệu thuộc bí mật nhà nước; hệ thống thông tin phục vụ an ninh, quốc phòng;

b) Đối với cơ sở hạ tầng thông tin quốc gia; hệ thống thông tin điều hành lưới điện quốc gia; hệ thống thông tin tài chính, ngân hàng; hệ thống thông tin điều khiển giao thông;

c) Thu lợi bất chính rất lớn hoặc đặc biệt lớn;

d) Gây hậu quả rất nghiêm trọng hoặc đặc biệt nghiêm trọng.

4. Người phạm tội còn có thể bị phạt tiền từ năm triệu đồng đến năm mươi triệu đồng, cấm đảm nhiệm chức vụ, cấm hành nghề hoặc làm công việc nhất định từ một năm đến năm năm.

Điều 226b. Tội sử dụng mạng máy tính, mạng viễn thông, mạng Internet hoặc thiết bị số thực hiện hành vi chiếm đoạt tài sản

1. Người nào sử dụng mạng máy tính, mạng viễn thông, mạng Internet hoặc thiết bị số thực hiện một trong những hành vi sau đây, thì bị phạt tiền từ mười triệu đồng đến một trăm triệu đồng hoặc phạt tù từ một năm đến năm năm:

a) Sử dụng thông tin về tài khoản, thẻ ngân hàng của cơ quan, tổ chức, cá nhân để chiếm đoạt hoặc làm giả thẻ ngân hàng nhằm chiếm đoạt tài sản của chủ thẻ hoặc thanh toán hàng hoá, dịch vụ;

b) Truy cập bất hợp pháp vào tài khoản của cơ quan, tổ chức, cá nhân nhằm chiếm đoạt tài sản;

c) Lừa đảo trong thương mại điện tử, kinh doanh tiền tệ, huy động vốn tín dụng, mua bán và thanh toán cổ phiếu qua mạng nhằm chiếm đoạt tài sản của cơ quan, tổ chức, cá nhân;

d) Hành vi khác nhằm chiếm đoạt tài sản của cơ quan, tổ chức, cá nhân.

2. Phạm tội thuộc một trong các trường hợp sau đây, thì bị phạt tù từ ba năm đến bảy năm:

a) Có tổ chức;

b) Phạm tội nhiều lần;

c) Có tính chất chuyên nghiệp;

d) Chiếm đoạt tài sản có giá trị từ năm mươi triệu đồng đến dưới hai trăm triệu đồng;

đ) Gây hậu quả nghiêm trọng;

e) Tái phạm nguy hiểm.

3. Phạm tội thuộc một trong các trường hợp sau đây, thì bị phạt tù từ bảy năm đến mười lăm năm:

a) Chiếm đoạt tài sản có giá trị từ hai trăm triệu đồng đến dưới năm trăm triệu đồng;

b) Gây hậu quả rất nghiêm trọng.

4. Phạm tội thuộc một trong các trường hợp sau đây, thì bị phạt tù từ mười hai năm đến hai mươi năm hoặc tù chung thân:

a) Chiếm đoạt tài sản có giá trị từ năm trăm triệu đồng trở lên;

b) Gây hậu quả đặc biệt nghiêm trọng.

5. Người phạm tội còn có thể bị phạt tiền từ năm triệu đồng đến một trăm triệu đồng, tịch thu một phần hoặc toàn bộ tài sản, cấm đảm nhiệm chức vụ, cấm hành nghề hoặc làm công việc nhất định từ một năm đến năm năm.

7.4.2. Điều trong Nghị định Chính phủ

Về chế tài xử phạt với các hành vi phạm tội liên quan tới tin học và hạ tầng công nghệ thông tin, ngày 23/8/2001 Chính phủ đã ban hành nghị định 55/2001/NĐ-CP với nội dung như sau:

Điều 41. Các hành vi vi phạm, hình thức và mức xử phạt vi phạm hành chính về Internet được quy định như sau:

1. Phạt cảnh cáo hoặc phạt tiền từ 50.000 đồng đến 200.000 đồng đối với hành vi không khai báo làm thủ tục cấp lại khi giấy phép cung cấp dịch vụ Internet bị mất, hoặc bị hư hỏng.

2. Phạt tiền từ 200.000 đồng đến 1.000.000 đồng đối với một trong các hành vi vi phạm sau đây:

a) Sử dụng mật khẩu, khoá mật mã, thông tin riêng của người khác để truy nhập, sử dụng dịch vụ Internet trái phép.

b) Sử dụng các công cụ phần mềm để truy nhập, sử dụng dịch vụ Internet trái phép.

3. Phạt tiền từ 1.000.000 đồng đến 5.000.000 đồng đối với một trong các hành vi vi phạm sau đây:

a) Vi phạm các quy định của Nhà nước về tiêu chuẩn, chất lượng trong việc sử dụng dịch vụ Internet.

b) Vi phạm các quy định của Nhà nước về giá, cước trong việc sử dụng dịch vụ Internet.

c) Vi phạm các quy định của Nhà nước về quản lý tài nguyên Internet trong việc sử dụng dịch vụ Internet.

d) Vi phạm các quy định của Nhà nước về quản lý truy nhập, kết nối Internet trong việc sử dụng dịch vụ Internet.

đ) Vi phạm các quy định của Nhà nước về mã hoá và giải mã thông tin trên Internet trong việc sử dụng dịch vụ Internet.

e) Vi phạm các quy định của Nhà nước về an toàn, an ninh thông tin trên Internet trong việc sử dụng dịch vụ Internet.

4. Phạt tiền từ 5.000.000 đồng đến 10.000.000 đồng đối với một trong các hành vi vi phạm sau đây:

a) Ngừng hoặc tạm ngừng cung cấp dịch vụ Internet mà không thông báo cho người sử dụng dịch vụ Internet biết trước, trừ trường hợp bất khả kháng.

b) Sửa chữa, tẩy xóa làm thay đổi nội dung giấy phép cung cấp dịch vụ Internet.

c) Sử dụng quá hạn giấy phép cung cấp dịch vụ Internet.

5. Phạt tiền từ 10.000.000 đồng đến 20.000.000 đồng đối với một trong các hành vi vi phạm sau đây:

a) Vi phạm các quy định của Nhà nước về tiêu chuẩn, chất lượng dịch vụ Internet trong việc cung cấp dịch vụ Internet.

b) Vi phạm các quy định của Nhà nước về giá, cước dịch vụ Internet trong việc cung cấp dịch vụ Internet.

c) Vi phạm các quy định của Nhà nước về quản lý tài nguyên Internet trong việc cung cấp dịch vụ Internet.

d) Vi phạm các quy định của Nhà nước về quản lý truy nhập, kết nối Internet trong việc cung cấp dịch vụ Internet.

đ) Vi phạm các quy định của Nhà nước về mã hoá và giải mã thông tin trên Internet trong việc cung cấp dịch vụ Internet.

e) Vi phạm các quy định của Nhà nước về an toàn, an ninh thông tin trên Internet trong việc cung cấp dịch vụ Internet.

g) Sử dụng Internet để nhằm mục đích đe dọa, quấy rối, xúc phạm đến danh dự, nhân phẩm người khác mà chưa đến mức truy cứu trách nhiệm hình sự.

h) Đưa vào Internet hoặc lợi dụng Internet để truyền bá các thông tin, hình ảnh đồi trụy, hoặc những thông tin khác trái với quy định của pháp luật về nội dung thông tin trên Internet, mà chưa đến mức truy cứu trách nhiệm hình sự.

i) Đánh cắp mật khẩu, khoá mật mã, thông tin riêng của tổ chức, cá nhân và phổ biến cho người khác sử dụng.

k) Vi phạm các quy định về vận hành, khai thác và sử dụng máy tính gây rối loạn hoạt động, phong toả hoặc làm biến dạng, làm hủy hoại các dữ liệu trên Internet mà chưa đến mức truy cứu trách nhiệm hình sự.

6. Phạt tiền từ 20.000.000 đồng đến 50.000.000 đồng đối với một trong các hành vi vi phạm sau đây:

a) Thiết lập hệ thống thiết bị và cung cấp dịch vụ Internet không đúng với các quy định ghi trong giấy phép.

b) Tạo ra và cố ý lan truyền, phát tán các chương trình vi rút trên Internet mà chưa đến mức truy cứu trách nhiệm hình sự.

7. Phạt tiền từ 50.000.000 đồng đến 70.000.000 đồng đối với hành vi thiết lập hệ thống thiết bị và cung cấp dịch vụ Internet khi không có giấy phép.

8. Ngoài các hình thức xử phạt chính, tùy theo tính chất, mức độ vi phạm mà tổ chức, cá nhân còn có thể bị áp dụng một hay nhiều hình thức xử phạt bổ sung hoặc biện pháp khắc phục hậu quả sau đây:

a) Tạm đình chỉ hoặc đình chỉ việc cung cấp và sử dụng dịch vụ Internet đối với các hành vi vi phạm tại điểm a khoản 2, điểm b khoản 2, các điểm tại khoản 3, các điểm tại khoản 5 và điểm b khoản 6 Điều 41.

b) Tước quyền sử dụng giấy phép có thời hạn hoặc không thời hạn đối với hành vi vi phạm quy định tại điểm b khoản 4 và điểm a khoản 6 Điều 41.

c) Tịch thu tang vật, phương tiện được sử dụng để vi phạm hành chính đối với hành vi vi phạm quy định tại điểm b khoản 4, điểm a khoản 6 và khoản 7 Điều 41.

d) Buộc khôi phục lại tình trạng ban đầu đã bị thay đổi do vi phạm hành chính gây ra đối với hành vi vi phạm quy định tại điểm k khoản 5, điểm b khoản 6 Điều 41.

7.4.3. Các điều trong Luật Công nghệ thông tin

Năm 2006, Quốc hội Việt Nam ban hành Luật 67/2006/QH11: Luật Công nghệ thông tin bao gồm 79 điều. Dưới đây là một số điều có liên quan đến nghĩa vụ của công dân đối với các hoạt động công nghệ thông tin.

Điều 12. Các hành vi bị nghiêm cấm

1. Cản trở hoạt động hợp pháp hoặc hỗ trợ hoạt động bất hợp pháp về ứng dụng và phát triển công nghệ thông tin; cản trở bất hợp pháp hoạt động của hệ thống máy chủ tên miền quốc gia; phá hoại cơ sở hạ tầng thông tin, phá hoại thông tin trên môi trường mạng.

2. Cung cấp, trao đổi, truyền đưa, lưu trữ, sử dụng thông tin số nhằm mục đích sau đây:

a) Chống Nhà nước Cộng hoà xã hội chủ nghĩa Việt Nam, phá hoại khối đoàn kết toàn dân;

b) Kích động bạo lực, tuyên truyền chiến tranh xâm lược, gây hận thù giữa các dân tộc và nhân dân các nước, kích động dâm ô, đồi trụy, tội ác, tệ nạn xã hội, mê tín dị đoan, phá hoại thuần phong mỹ tục của dân tộc;

c) Tiết lộ bí mật nhà nước, bí mật quân sự, an ninh, kinh tế, đối ngoại và những bí mật khác đã được pháp luật quy định;

d) Xuyên tạc, vu khống, xúc phạm uy tín của tổ chức, danh dự, nhân phẩm, uy tín của công dân;

đ) Quảng cáo, tuyên truyền hàng hoá, dịch vụ thuộc danh mục cấm đã được pháp luật quy định.

3. Xâm phạm quyền sở hữu trí tuệ trong hoạt động công nghệ thông tin; sản xuất, lưu hành sản phẩm công nghệ thông tin trái pháp luật; giả mạo trang thông tin điện tử của tổ chức, cá nhân khác; tạo đường dẫn trái phép đối với tên miền của tổ chức, cá nhân sử dụng hợp pháp tên miền đó.

Điều 69. Bảo vệ quyền sở hữu trí tuệ trong lĩnh vực công nghệ thông tin

Việc bảo vệ quyền sở hữu trí tuệ trong lĩnh vực công nghệ thông tin phải thực hiện theo quy định của pháp luật về sở hữu trí tuệ và các quy định sau đây:

1. Tổ chức, cá nhân truyền đưa thông tin trên môi trường mạng có quyền tạo ra bản sao tạm thời một tác phẩm được bảo hộ do yêu cầu kỹ thuật của hoạt động truyền đưa thông tin và bản sao tạm thời được lưu trữ trong khoảng thời gian đủ để thực hiện việc truyền đưa thông tin;

2. Người sử dụng hợp pháp phần mềm được bảo hộ có quyền sao chép phần mềm đó để lưu trữ dự phòng và thay thế phần mềm bị phá hỏng mà không phải xin phép, không phải trả tiền bản quyền.

Điều 70. Chống thư rác

1. Tổ chức, cá nhân không được che giấu tên của mình hoặc giả mạo tên của tổ chức, cá nhân khác khi gửi thông tin trên môi trường mạng.

2. Tổ chức, cá nhân gửi thông tin quảng cáo trên môi trường mạng phải bảo đảm cho người tiêu dùng khả năng từ chối nhận thông tin quảng cáo.

3. Tổ chức, cá nhân không được tiếp tục gửi thông tin quảng cáo trên môi trường mạng đến người tiêu dùng nếu người tiêu dùng đó thông báo không đồng ý nhận thông tin quảng cáo.

Điều 71. Chống vi-rút máy tính và phần mềm gây hại

Tổ chức, cá nhân không được tạo ra, cài đặt, phát tán vi-rút máy tính, phần mềm gây hại vào thiết bị số của người khác để thực hiện một trong những hành vi sau đây:

1. Thay đổi các tham số cài đặt của thiết bị số;

2. Thu thập thông tin của người khác;

3. Xóa bỏ, làm mất tác dụng của các phần mềm bảo đảm an toàn, an ninh thông tin được cài đặt trên thiết bị số;

4. Ngăn chặn khả năng của người sử dụng xóa bỏ hoặc hạn chế sử dụng những phần mềm không cần thiết;

5. Chiếm đoạt quyền điều khiển thiết bị số;

6. Thay đổi, xóa bỏ thông tin lưu trữ trên thiết bị số;

7. Các hành vi khác xâm hại quyền, lợi ích hợp pháp của người sử dụng.

Điều 72. Bảo đảm an toàn, bí mật thông tin

1. Thông tin riêng hợp pháp của tổ chức, cá nhân trao đổi, truyền đưa, lưu trữ trên môi trường mạng được bảo đảm bí mật theo quy định của pháp luật.

2. Tổ chức, cá nhân không được thực hiện một trong những hành vi sau đây:

a) Xâm nhập, sửa đổi, xóa bỏ nội dung thông tin của tổ chức, cá nhân khác trên môi trường mạng;

b) Cản trở hoạt động cung cấp dịch vụ của hệ thống thông tin;

c) Ngăn chặn việc truy nhập đến thông tin của tổ chức, cá nhân khác trên môi trường mạng, trừ trường hợp pháp luật cho phép;

d) Bẻ khóa, trộm cắp, sử dụng mật khẩu, khóa mật mã và thông tin của tổ chức, cá nhân khác trên môi trường mạng;

đ) Hành vi khác làm mất an toàn, bí mật thông tin của tổ chức, cá nhân khác được trao đổi, truyền đưa, lưu trữ trên môi trường mạng.

Điều 73. Trách nhiệm bảo vệ trẻ em

1. Nhà nước, xã hội và nhà trường có trách nhiệm sau đây:

a) Bảo vệ trẻ em không bị tác động tiêu cực của thông tin trên môi trường mạng;

b) Tiến hành các biện pháp phòng, chống các ứng dụng công nghệ thông tin có nội dung kích động bạo lực và khiêu dâm.

2. Gia đình có trách nhiệm ngăn chặn trẻ em truy nhập thông tin không có lợi cho trẻ em.

3. Cơ quan nhà nước có thẩm quyền tiến hành những biện pháp sau đây để ngăn ngừa trẻ em truy nhập thông tin không có lợi trên môi trường mạng:

a) Tổ chức xây dựng và phổ biến sử dụng phần mềm lọc nội dung;

b) Tổ chức xây dựng và phổ biến công cụ ngăn chặn trẻ em truy nhập thông tin không có lợi cho trẻ em;

c) Hướng dẫn thiết lập và quản lý trang thông tin điện tử dành cho trẻ em nhằm mục đích thúc đẩy việc thiết lập các trang thông tin điện tử có nội dung thông tin phù hợp với trẻ em, không gây hại cho trẻ em; tăng cường khả năng quản lý nội dung thông tin trên môi trường mạng phù hợp với trẻ em, không gây hại cho trẻ em.

4. Nhà cung cấp dịch vụ có biện pháp ngăn ngừa trẻ em truy nhập trên môi trường mạng thông tin không có lợi đối với trẻ em.

5. Sản phẩm, dịch vụ công nghệ thông tin mang nội dung không có lợi cho trẻ em phải có dấu hiệu cảnh báo.

CÂU HỎI VÀ BÀI TẬP

1. Lỗ hổng phần mềm là gì?
2. Nêu các hình thức tấn công vào các tài nguyên thông tin hiện nay?
3. Virus máy tính là gì? Phân biệt giữa 2 nhóm virus: virus biên dịch và virus thông dịch? Kể tên một số loại virus máy tính.
4. Sâu máy tính là gì? Trình bày sự khác nhau giữa 2 sâu máy tính: sâu dịch vụ mạng và sâu thư điện tử? Kể tên một số loại sâu máy tính.
5. Trojan máy tính là gì? Kể tên và nêu cơ chế hoạt động của một số loại trojan máy tính.
6. Trình bày sự khác nhau về cơ chế hoạt động giữa 3 loại phần mềm độc hại: virus máy tính, sâu máy tính và trojan?
7. Tấn công từ chối dịch vụ - Denial of Service (DoS) là gì? Phân biệt 2 kiểu tấn công từ chối dịch vụ: dựa trên lỗ hổng phần mềm và làm ngập lụt?
8. Tấn công từ chối dịch vụ phân tán – distributed denial of service (DDoS) là gì? So sánh DDoS với DoS?
9. Phân biệt 3 loại hình thức lừa đảo: dùng bản sao, hướng đối tượng, dùng điện thoại?
10. Tài sản trí tuệ là gì? Quyền sở hữu trí tuệ là gì? Vì sao phải bảo hộ quyền sở hữu trí tuệ?

TÀI LIỆU THAM KHẢO

I. Sách

1. Carl Reynolds and Paul Tymannn (2008). *Schaum's Outline of Principles of Computer Science*. McGraw-Hill Companies, Inc.
2. Donald Ervin Knuth (1997). *The Art of Computer Programming, third edition, Volume 1/ Fundamental Algorithms*. Addison-Wesley Professional.
3. Đào Kiến Quốc, Bùi Thế Duy (2006). *Giáo trình Tin học cơ sở*. NXB Đại học Quốc gia Hà Nội.
5. Đào Kiến Quốc, Trương Ninh Thuận (2011), *Các khái niệm cơ bản của Tin học*, NXB Đại học Quốc gia Hà Nội.
6. Đỗ Thanh Liên Ngân, Hồ Văn Tú (2005).*Giáo trình môn học Tin học căn bản*, Đại học Cần Thơ.
7. Đỗ Thị Mơ và đồng nghiệp (2007). *Tin học đại cương*. NXB Nông nghiệp.
8. Đỗ Xuân Lôi (2004). *Cấu trúc dữ liệu và giải thuật*. NXB ĐH Quốc Gia Hà Nội.
9. Ian Sommerville (2010), *Software Engineering*. Addison-Wesley; 9th edition.
10. Hoàng Thị Hà (2011), *Bài giảng cơ sở dữ liệu 1*. NXB Khoa học Tự nhiên và Công nghệ
11. Hồ Thuần, Hồ Cẩm Hà (2004). *Các hệ cơ sở dữ liệu: Lí thuyết &thực hành; tập một*. NXB Giáo dục.
12. June Jamrich Parsons, Dan Oja (2013), *Computer Concepts*, 15th edition. Cengage Learning.
13. J. Glenn Brookshear (2012). *Computer science – An overview*, 11thedition. Pearson Education, Inc., publishing as Addison-Wesley.
14. June Jamrich Parsons, Dan Oja (2013). *New Perspectives on Computer Concepts*, 15th edition. Cengage Learning.
15. Nguyễn Đình Hóa (2004). *Giáo trình Cấu trúc dữ liệu và giải thuật*. NXB ĐH Quốc gia Hà Nội.
16. Nguyễn Thị Khiêm Hòa (2013), *Giáo trình Tin học đại cương*, Khoa CNTT Đại học Ngân Hàng TP.HCM.
17. Nguyễn Văn Linh (2003). *Giáo trình Ngôn ngữ lập trình*, Khoa CNTT Đại học Cần Thơ.
18. Phan Thị Thanh Hà (2008), *Bài giảng tóm tắt Internet và dịch vụ*, Đại học Đà Lạt.
19. V. ANTON SPRAUL (2012), *Computer Science Made Simple*, Broadway.
20. Viện Công nghệ thông tin và Truyền thông (2010), *Giáo trình Tin học đại cương*, Đại học Bách khoa Hà Nội.
21. Võ Thanh Tú, Hoàng Hữu Hạnh, *Giáo trình mạng máy tính* (2003), Đại học Khoa học Huế - Đại học Huế.

II. Internet

1. Nguyễn Hoài Tưởng, *Cách sử dụng hiệu quả các dịch vụ lưu trữ đám mây*, http://vietsciexdir.net/ovsed-blog/blog/2012/04/, trích dẫn 4/8/2013
2. Nguyễn Hứa Phùng (2006). *Ngôn ngữ lập trình/Chương1: Giới thiệu*, http://www.cse.hcmut.edu.vn/~nnlt/LectureNotes/Chuong_1.pdf, trích dẫn ngày 10/04/2014.
3. Nguyễn Việt Anh. *Ngôn ngữ lập trình và chương trình dịch*, http://uet.vnu.edu.vn/~anhnv/courses/thcs/9_Ngon_ngu_lap_trinh.pdf, trích dẫn ngày 10/04/2014.
4. NIIT (2007). *Bài giảng Ngôn ngữ lập trình – Các khái niệm, Giải thuật và Lưu đồ*, http://clip.vn/watch/Bai-1-Ngon-ngu-lap-trinh-Cac-khai-niem-Giai-thuat-va-Luu-do,i6J/, trích dẫn ngày 10/04/2014.