

ĐOÀN PHAN TÂN

TIN HỌC

trong hoạt động

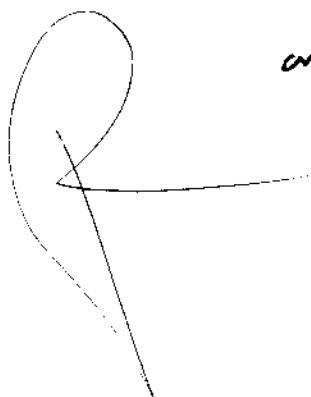
THÔNG TIN - THƯ VIỆN

*Giáo trình dành cho sinh viên ngành
thông tin-thư viện và quản trị thông tin*



NHÀ XUẤT BẢN ĐẠI HỌC QUỐC GIA HÀ NỘI

PGS.TS. ĐOÀN PHAN TÂN



chữ 2666

TIN HỌC TRONG HOẠT ĐỘNG THÔNG TIN - THƯ VIỆN

- *Tin học tư liệu*
- *CDS/ISIS 3.08 for DOS*
- *CDS/ISIS 1.31 for Windows*

NHÀ XUẤT BẢN ĐẠI HỌC QUỐC GIA HÀ NỘI - 2001

LỜI NÓI ĐẦU

Trong hơn nửa thế kỷ qua, các thư viện trên thế giới đã biến đổi sâu sắc và toàn diện dựa trên những thành tựu của công nghệ thông tin hiện đại: máy tính điện tử, liên lạc viễn thông, các kỹ thuật ghi và lưu giữ thông tin đa phương tiện v.v... Tin học hoá công tác thư viện đang là xu thế tất yếu và phát triển rất nhanh, đồng thời đem lại hiệu quả vô cùng to lớn.

Việc ứng dụng của tin học trong hoạt động thông tin - thư viện mới diễn ra trong vòng mấy chục năm gần đây nhưng đã làm thay đổi diện mạo hoạt động của ngành thông tin - thư viện: tăng nhanh tốc độ ở tất cả các công đoạn xử lý thông tin, cải tiến toàn bộ quy trình công nghệ hiện hành. Vì vậy việc nắm vững những kiến thức cơ bản về tin học hoá thư viện và có kỹ năng thực hành khai thác các hệ thống thông tin - thư viện tự động hoá, các mạng thông tin máy tính là yêu cầu không thể thiếu được đối với các cán bộ thông tin - thư viện ở nước ta hiện nay, bất kể ở trình độ đào tạo nào.

Giáo trình "Tin học trong hoạt động thông tin - thư viện" của tác giả do Trường Đại học Văn hoá Hà Nội phối hợp với Nhà xuất bản Văn hoá - Thông tin xuất bản năm 1997 đã phần nào đáp ứng yêu cầu cấp bách đó.

Về nội dung, "Tin học trong hoạt động thông tin - thư viện" đề cập đến những vấn đề cơ bản của Tin học tư liệu: các yếu tố

cấu thành hệ thống thông tin - thư viện tự động hoá, cấu trúc dữ liệu, lý thuyết về cơ sở dữ liệu (CSDL), các mô hình dữ liệu và vấn đề thiết kế CSDL, cấu trúc của hệ quản trị CSDL với tệp đảo, phân tích và thiết kế một hệ tin học ứng dụng trong công tác thư viện ở giai đoạn đầu. Ngoài ra cuốn sách còn hướng dẫn kỹ năng sử dụng phần mềm quản trị CSDL từ liệu CDS/ISIS để xây dựng và khai thác các CSDL quản lý các nguồn thông tin đã được tư liệu hoá. Đây là phần mềm tư liệu đang được sử dụng rộng rãi ở các thư viện và trung tâm thông tin ở nước ta hiện nay.

Ở lần xuất bản này, ngoài các nội dung trên, tác giả trình bày kỹ hơn về các phần mềm tư liệu và phần mềm tích hợp quản trị thư viện, vấn đề biên mục tự động và cấu trúc của khổ mẫu biên mục đọc máy MARC, UNIMARC. Ngoài mô hình cơ sở dữ liệu quan hệ, tác giả bổ sung thêm phần phương pháp thiết kế khái niệm CSDL bằng mô hình thực thể quan hệ. Đây là mô hình do Peter Chen đề xuất năm 1976 và đã nổi lên như một mô hình chuẩn để thiết kế khái niệm CSDL.

Để phù hợp với xu hướng phát triển hiện nay, trong lần xuất bản này, thay cho chương trình CDS/ISIS chạy trên môi trường DOS, tác giả giới thiệu chương trình CDS/ISIS chạy trên môi trường Windows (CDS/ISIS for Windows version 1.31, do UNESCO phát hành năm 1999). So với CDS/ISIS chạy trên DOS, CDS/ISIS for Windows có nhiều ưu điểm vượt trội: các công đoạn tạo cấu trúc CSDL được máy trợ giúp gần như hoàn toàn nên dễ thực hiện, các thao tác tìm tin, in ấn cũng đơn giản và hiệu quả hơn và với những tính năng được phát triển (các hàm mới, các lệnh liên kết siêu văn bản LINK,...) đã cho phép mở rộng rất nhiều khả năng của CDS/ISIS.

Đối với bạn đọc còn quan tâm đến CDS/ISIS chạy trên DOS vẫn có thể tham khảo cách sử dụng chương trình này trong

phần phụ chương. Ở đây tác giả trình bày khá chi tiết các bước xây dựng, sửa đổi cấu trúc, cập nhật và khai thác CSDL bằng chương trình CDS/ISIS version 3.08 do UNESCO phát hành năm 1997. Phụ chương còn giới thiệu một số bài tập về mô hình cơ sở dữ liệu quan hệ.

Cuốn sách này được viết trên cơ sở cuốn sách cùng tên của tác giả xuất bản năm 1997 và tập hợp các bài giảng về “Thiết kế và quản trị CSDL văn bản” mà tác giả đã trình bày cho sinh viên bậc sau đại học ngành thông tin - thư viện ở trường Đại học Văn hoá Hà Nội trong những năm gần đây.

Trong quá trình biên soạn cuốn sách chúng tôi nhận được sự đóng góp chân tình của các bạn đồng nghiệp ở bộ môn Thông tin học, khoa Thông tin - Thư viện, các chuyên gia đang làm việc tại các trung tâm thông tin và thư viện. Cho phép tôi được bày tỏ lời cảm ơn chân thành.

Mặc dù đã có nhiều cố gắng, song cuốn sách không thể tránh khỏi những thiếu sót. Rất mong sự góp ý của đồng nghiệp và bạn đọc để cuốn sách được tốt hơn.

Hà Nội, tháng 6 năm 2001

TÁC GIẢ

MỤC LỤC

LỜI NÓI ĐẦU	3
MỤC LỤC	7
Chương 1 TỔNG QUAN VỀ HỆ THỐNG THÔNG TIN - THƯ VIỆN TỰ ĐỘNG HOÁ	13
1. Hoạt động thông tin - thư viện dưới tác động của công nghệ thông tin hiện đại.	13
1.1. Bùng nổ thông tin và bùng nổ công nghệ	13
1.2. Xu hướng tin học hóa hoạt động thông tin - thư viện	16
2. Các yếu tố cấu thành hệ thống thông tin tự động hoá	19
2.1. Nhân sự trong hệ thống thông tin tự động hoá	19
2.2. Máy tính điện tử và các thiết bị ngoại vi	20
2.3. Các vật mang tin điện tử	24
2.4. Các phần mềm hệ thống và phần mềm chuyên dụng	26
2.5. Các thiết bị viễn thông	28
3. Khảo sát một trường hợp: Viện thông tin Khoa học Kỹ thuật Quốc gia Pháp	32

Chương 2	CẤU TRÚC DỮ LIỆU	40
1.	Thông tin và dữ liệu	40
1.1.	Thông tin	40
1.2.	Dữ liệu	48
2.	Cấu trúc dữ liệu	50
2.1.	Mảng	50
2.2.	Danh sách	51
2.3.	Cây	51
2.4.	Sắp	52
2.5.	Hàng	52
2.6.	Biểu ghi	52
2.7.	Tập	54
2.8.	Cơ sở dữ liệu	59
Chương 3	CƠ SỞ DỮ LIỆU VÀ HỆ QUẢN TRỊ CƠ SỞ DỮ LIỆU	60
1.	Mở đầu	60
2.	Từ quản lý tệp đến hệ thống quản trị dữ liệu	61
3.	Cơ sở dữ liệu	63
3.1.	Cơ sở dữ liệu và hệ quản trị cơ sở dữ liệu	63
3.2.	Phân loại CSDL	65
4.	Ngôn ngữ dữ liệu	68
5.	Các tiêu chuẩn để lựa chọn một hệ quản trị CSDL	70

Chương 4	PHẦN MỀM TƯ LIỆU VÀ KHỔ MẪU BIÊN MỤC ĐỌC MÁY MARC	74
1.	Phần mềm tư liệu và phần mềm tích hợp quản trị thư viện	74
2.	Biên mục tự động	79
2.1.	Chuẩn hoá công tác biên mục	79
2.2.	Biên mục tự động	83
3.	Khổ mẫu biên mục đọc máy MARC	86
Chương 5	MÔ HÌNH DỮ LIỆU VÀ VẤN ĐỀ THIẾT KẾ CSDL	96
1.	Các mô hình dữ liệu	96
1.1.	Một số khái niệm cơ bản	96
1.2.	Các loại mô hình dữ liệu	101
2.	Khái niệm chung về thiết kế CSDL	106
2.1.	Thiết kế CSDL trong vòng đời của hệ thống thông tin	106
2.2.	Trừu tượng hoá trong thiết kế khái niệm CSDL	108
2.3.	Các pha trong thiết kế CSDL	112
Chương 6	MÔ HÌNH THỰC THỂ QUAN HỆ VÀ MÔ HÌNH QUAN HỆ	115
1.	Mô hình thực thể quan hệ	115
1.1.	Các khái niệm cơ bản	116
1.2.	Phương pháp thiết kế khái niệm CSDL bằng mô hình thực thể quan hệ	117
2.	Mô hình quan hệ	125

2.1. Các khai niệm cơ bản	125
2.2. Đại số quan hệ	132
2.3. Tối ưu hoá câu hỏi	140
Chương 7 HỆ QUẢN TRỊ CƠ SỞ DỮ LIỆU CDS/ISIS FOR WINDOWS	146
1. Tổng quan về hệ thống lưu trữ và tìm kiếm thông tin CDS/ISIS for Windows	146
1.1. Chức năng của Win ISIS	147
1.2. Nguyên tắc hoạt động của WinISIS	148
1.3. Hệ thống các tệp của CSDL trong WinISIS	151
2. Cài đặt WinISIS trên máy vi tính	156
2.1. Tổ chức thư mục và yêu cầu hệ thống	156
2.2. Nội dung của tệp syspar.par	158
2.3. Thử tục cài đặt	165
2.4. Cài đặt trên mạng cục bộ (LAN)	167
2.5. Khởi động WinISIS	168
3. Ngôn ngữ tạo format	171
4. Bảng chọn trường	184
Chương 8 XÂY DỰNG VÀ KHAI THÁC CSDL BẰNG CHƯƠNG TRÌNH WINISIS	189
1. Xây dựng cấu trúc một CSDL mới bằng chương trình WinISIS	189
2. Nhập và xem dữ liệu	198
2.1. Nhập, hiệu đính dữ liệu	198

2.2. Xem dữ liệu	204
3. Tìm tin trong WinISIS	207
3.1. Nguyên tắc tìm tin theo tệp đảo	207
3.2. Ngôn ngữ tìm tin WinISIS	207
3.3. Phương tiện tìm tin trong WinISIS	211
3.4. Tìm tự do	214
3.5. Lưu kết quả tìm	215
4. Phương tiện in trong Win ISIS	216
4.1. Nguyên tắc chung	216
4.2. Hộp thoại in (Trang tổng quát)	218
4.3. Hộp thoại in (Trang trình bày)	220
4.4. Hộp thoại in (Trang đặt lề)	221
4.5. Hộp thoại in (Trang chế bản)	222
4.6. Hộp thoại in (Trang sắp xếp)	224
5. Trao đổi dữ liệu	228
5.1. Nguyên tắc trao đổi dữ liệu	228
5.2. Xuất dữ liệu	228
5.3. Nhập dữ liệu	230
6. Các phương tiện làm việc với tệp chủ	232
6.1. Nhập dữ liệu tự động theo MFN	232
6.2. Xoá dữ liệu tự động theo MFN	233
6.3. Thay thế dữ liệu tự động theo MFN	234

Chương 9	THẾT KẾ MỘT HỆ TIN HỌC ỨNG DỤNG TRONG CÔNG TÁC THƯ VIỆN	236
1. Mở đầu		236
2. Phân tích chức năng hệ thống		237
3. Thết kế một hệ tin học ứng dụng trong công tác thư viện		241
PHỤ CHƯƠNG		248
A	XÂY DỰNG VÀ KHAI THÁC CSDL BẰNG CHƯƠNG TRÌNH CDS/ISIS CHẠY TRÊN DOS	248
1. Xây dựng cấu trúc CSDL		248
2. Sửa đổi cấu trúc CSDL		262
3. Cập nhật dữ liệu		264
4. Tìm và xem dữ liệu		270
5. Sắp xếp và in		276
6. Trao đổi dữ liệu		283
B	BÀI TẬP VỀ MÔ HÌNH DỮ LIỆU	287
TÀI LIỆU THAM KHẢO		295

Chương 1

TỔNG QUAN VỀ HỆ THỐNG THÔNG TIN - THƯ VIỆN TỰ ĐỘNG HOÁ

1. HOẠT ĐỘNG THÔNG TIN - THƯ VIỆN DƯỚI TÁC ĐỘNG CỦA CÔNG NGHỆ THÔNG TIN HIỆN ĐẠI

1.1. Bùng nổ thông tin và bùng nổ công nghệ

Cuộc cách mạng khoa học kỹ thuật phát triển như vũ bão ngày nay thực chất là cuộc cách mạng công nghệ. Nét nổi bật của nó là khoa học đã trở thành lực lượng sản xuất trực tiếp. Sự phát triển của khoa học đã thu hút một bộ phận lớn nhân lực vào lĩnh vực này. Do đó đội ngũ những người làm khoa học gia tăng nhanh chóng. Lực lượng những người làm khoa học tăng lên theo cấp số cộng thì tài liệu khoa học, những sản phẩm nghiên cứu của họ tăng lên theo cấp số nhân. Tất cả đã tạo nên một khối lượng thông tin khổng lồ và không ngừng phát triển, dẫn đến hiện tượng bùng nổ thông tin. Ngoài ra cộng đồng khoa học được bổ sung thêm nhiều loại người dùng tin: các nhà quản lý, các nhà nghiên cứu, các nhà công nghệ, các nhà giáo dục, các nhà sản xuất kinh doanh... Họ không chỉ là những người dùng tin, mà còn là những người sản sinh ra những thông tin mới.

Một nền công nghiệp sản xuất tri thức khoa học, mà cơ sở của nó là truyền tri thức thông tin, tiếp tục tăng nhanh trong

một thế giới xây dựng trên cơ sở của tiến bộ khoa học và công nghệ.

Sự gia tăng nhanh chóng khối lượng tri thức khoa học không thể không ảnh hưởng tới hoạt động của các cơ quan thông tin thư viện. Trước hết nó tác động đến thành phần cơ cấu của kho tài liệu. Ngoài các sách báo và các ấn phẩm định kỳ xuất bản theo chu trình thương mại truyền thống, xuất hiện một loạt các tài liệu thuộc đủ các loại, không xuất bản được phân phối ở mức độ hẹp như: các báo cáo, luận văn, tổng kết hội nghị, giáo trình, tài liệu nghiên cứu, tài liệu chuẩn bị xuất bản... Chúng tạo thành nguồn tài liệu "xám", còn gọi là tài liệu không công bố. Những tài liệu này thường chứa các thông tin có giá trị cao. Mặc dù khó có thể biết được số lượng của những tài liệu này là bao nhiêu, nhưng người ta biết chắc rằng số lượng của chúng hiện nay tăng lên rất đáng kể.

Một hệ quả nữa của hiện tượng bùng nổ thông tin là sự rút ngắn đáng kể thời gian hữu ích của một tài liệu. Do đó người ta phải thường xuyên bổ sung vốn tài liệu và không ngừng phải xử lý chúng, hoặc bằng thủ công hoặc bằng phương tiện tự động hoá.

Ngoài ra, bên cạnh những tài liệu văn bản in trên giấy, còn có thêm những tài liệu không ở dạng sách như: đĩa, ảnh, băng từ, đĩa từ, đĩa quang... Sự xuất hiện của những tài liệu loại này trong sự chuyển giao thông tin tương ứng với một yếu tố quan trọng trong xã hội ngày nay là: sự xuất hiện các phương tiện nghe nhìn. Với một tương lai rất hứa hẹn, chúng đặt ra cho người làm công tác thông tin tư liệu những vấn đề về xử lý và phổ biến thông tin, dựa trên những kỹ thuật đặc biệt và dựa trên những kênh thông tin rất đa dạng.

Sự bùng nổ thông tin gắn liền với sự bùng nổ công nghệ đặc biệt trên ba lĩnh vực liên hệ chặt chẽ với công tác thông tin - thư viện là: tin học, viễn thông và vi xử lý, hạt nhân của công nghệ thông tin hiện đại.

Việc sử dụng các công cụ do con người chế tạo ra để thu thập, sản sinh, ghi chép, sắp xếp lại, truyền và khai thác thông tin đã có từ rất lâu. Nhưng chỉ từ khi máy tính điện tử ra đời (1946) và đặc biệt với cuộc cách mạng vi xử lý vào giữa những năm 70 tạo cơ sở cho sự ra đời hàng triệu, rồi hàng chục, hàng trăm triệu máy vi tính với năng lực ngày càng cao, giá ngày càng rẻ, thâm nhập khắp mọi nơi trên thế giới, mới thực sự mở ra một chân trời mới cho mỗi cá nhân và xã hội trong việc nắm bắt được những thông tin về những sự kiện và ý tưởng mới và mới tạo khả năng hiện thực cho việc hiện đại hoá hoạt động của các thư viện.

Cuối những năm 80 sang đầu những năm 90, sự phát triển bùng nổ các mạng viễn thông truyền dữ liệu quốc gia và quốc tế trên cơ sở kỹ thuật cáp sợi quang, vệ tinh và vi ba số đã tạo khả năng nối mạng không những giữa các trung tâm tính toán, mà còn nối được đến máy vi tính của từng cá nhân. Xuất hiện viễn cảnh của các siêu "xa lộ thông tin" liên kết hàng trăm triệu người trong từng quốc gia cũng như trong phạm vi khu vực và toàn cầu, mà tiêu biểu là liên mạng thông tin toàn cầu Internet.

Cùng với việc phát triển mạnh mẽ các kỹ thuật vi xử lý, việc sử dụng ngày càng nhiều các vật mang tin điện tử: băng từ, đĩa từ, đĩa quang đã loại bỏ được những khó khăn về sự quá tải của các kho chứa, đồng thời tạo điều kiện thuận lợi cho việc chuyển giao và phân phối thông tin với những kênh rất đa dạng.

1.2. Xu hướng tin học hoá hoạt động thông tin - thư viện

Vai trò của tin học trong các đơn vị thông tin - thư viện không ngừng gia tăng và phát triển với tốc độ rất nhanh. Trực tiếp hay gián tiếp các hoạt động và dịch vụ thông tin ngày nay đều dựa trên sự hỗ trợ của máy tính điện tử. Đồng thời ta càng thấy rõ vai trò của các mạng lưới tích nhập thông tin tự động hoá đang phát triển trong những năm gần đây.

Nhiệm vụ của các đơn vị thông tin là: thu thập tài liệu, xử lý thông tin, sản xuất ra các sản phẩm thông tin, tổ chức các dịch vụ tìm và phổ biến thông tin. Đặc điểm của các hoạt động này là các đơn vị thông tin thường phải quản lý một khối lượng tài liệu rất lớn và chúng được khai thác lặp đi lặp lại nhiều lần. Cũng như các đơn vị hành chính sự nghiệp khác, các đơn vị này cũng cần giải quyết các công việc mang tính chất quản lý, hành chính, văn phòng. Các công việc này rất thích hợp với khả năng ứng dụng của máy tính điện tử.

Mở đầu việc ứng dụng tin học trong công tác thông tin - thư viện thường tập trung vào việc lưu trữ, tìm kiếm thông tin và tạo ra các sản phẩm thông tin thư mục. Sau đó mở rộng dần ra các hoạt động kỹ thuật khác, các công việc quản lý lưu thông tài liệu và các dịch vụ phổ biến thông tin. Ngày nay ta thường gặp các hệ thống thông tin tự động hoá hoàn toàn hoặc từng phần các công việc như: bổ sung, biên mục, đánh chỉ số, lập các bộ phiếu, tạo ra các sản phẩm thông tin, các hoạt động tìm và khai thác dữ liệu, các hoạt động kiểm tra và quản lý hành chính thông thường.

Việc **ứng dụng máy tính điện tử** trong xử lý thông tin tư liệu mới diễn ra trong vòng gần 40 năm lại đây, nhưng đã đem lại hiệu quả thật là to lớn: tập trung thông tin trong những bộ nhớ lớn, những CSDL và ngân hàng dữ liệu (NHDL); tăng

nhANH tốc độ ở tất cả các công đoạn xử lý thông tin. Sự phát triển của những bộ nhớ lớn truy cập trực tiếp tạo cho khả năng tra cứu ngay, tại thời điểm bất kỳ những thông tin mà người dùng tin yêu cầu. Sự tiến bộ về chất trong quan hệ giữa người và máy, cùng với giá thành ngày càng hạ, giúp cho việc sử dụng máy tính trong công tác thông tin - thư viện ngày càng trở nên phổ cập.

Sự kết hợp giữa máy tính và viễn thông dẫn đến sự hình thành và phát triển các hệ thống và mạng lưới thông tin tự động hoá, cho phép các thư viện liên kết với nhau trên mạng máy tính để chia sẻ nguồn lực thông tin. Ở thư viện nhiều nước tiên tiến, người ta xây dựng các thư mục công cộng truy nhập trực tuyến, gọi tắt là OPAC (Online Public Access Catalog). Đó là các CSDL thư mục được khai thác trên mạng, giúp người sử dụng có thể truy nhập các thông tin thư mục một cách trực tiếp mà không cần sự hỗ trợ trung gian của nhân viên thư viện.

Ngày nay với sự phát triển của công nghệ thông tin, ngoài những tài liệu thông thường đã xuất hiện **nguồn tài liệu điện tử** như các sách, tạp chí điện tử ghi trên các đĩa máy tính và đĩa quang CD-ROM. Nhờ có các tính ưu việt như dung lượng nhớ lớn, độ bền vật lý cao, thao tác vận hành đơn giản, có khả năng lưu trữ văn bản, âm thanh, hình ảnh,... các đĩa quang ngày càng được sử dụng rộng rãi để lưu trữ và phổ biến thông tin, đặc biệt là thông tin đa phương tiện (multimedia). Trong những năm qua, một lượng thông tin khổng lồ, bao gồm các CSDL, các từ điển bách khoa, các sách tham khảo, các cẩm nang kỹ thuật, các chương trình phần mềm,... đã được phát hành dưới dạng CD-ROM. Điều đó sẽ ảnh hưởng tới các quy trình công nghệ và xử lý thông tin truyền thống, đồng thời cũng mở rộng khả năng và nâng cao chất lượng dịch vụ của các cơ quan thông tin thư viện.

Bước phát triển mới đây của thư viện là sự xuất hiện thư viện điện tử (electronic library) và thư viện kỹ thuật số (digital library). Đó có thể coi là xu hướng quan trọng nhất của tự động hoá thư viện trong tương lai.

Tin học hoá hoạt động thông tin - thư viện là xu thế phát triển tất yếu của các cơ quan thông tin - thư viện hiện nay và đang diễn ra với tốc độ ngày càng nhanh chóng. Theo số liệu của Tạp chí Thư viện (Library Journal) tính đến năm 1981, toàn thế giới chỉ có 301 thư viện tự động hoá, đến năm 1992 thế giới đã có 8789 thư viện tự động hoá, nghĩa là tăng 29 lần sau 10 năm.

Tuy nhiên, các hệ thống thông tin - thư viện tự động hoá phải được thiết kế, xây dựng, cung cấp thông tin và khai thác bởi con người. Có nghĩa là các hệ thống này không thể thay thế được con người mà trái lại nó còn đòi hỏi ở con người một chất lượng cao hơn, chuyên sâu hơn trong cả lĩnh vực thông tin học và tin học, để có thể khai thác được khả năng to lớn mà các hệ thống này đem lại.

Một điều cần nhấn mạnh là công nghệ thông tin phát triển rất nhanh và đang xâm nhập vào mọi lĩnh vực hoạt động của con người. Ngày nay cùng với việc giảm giá thành và điều kiện sử dụng thuận lợi, máy tính điện tử không còn là vật xa xỉ, mà đã đi vào từng gia đình. Sử dụng máy tính điện tử trở thành nhu cầu của mọi người như nói tiếng mẹ đẻ. Đó là cơ sở thuận lợi cho việc tin học hoá hoạt động thông tin thư viện.

Việc ứng dụng của tin học trong hoạt động thông tin - thư viện đã tạo điều kiện cho sự hình thành và phát triển một phân ngành của tin học, đó là Tin học tư liệu.

2. CÁC YẾU TỐ CẤU THÀNH HỆ THỐNG THÔNG TIN TỰ ĐỘNG HOÁ

Một hệ thống tin học bao gồm các yếu tố sau:

- Các cán bộ chuyên môn.
- Các thiết bị xử lý thông tin tự động: máy tính điện tử, các thiết bị ngoại vi, các vật mang tin điện tử.
- Các phần mềm hệ thống và phần mềm chuyên dụng.
- Các phương tiện viễn thông.

Các yếu tố trên đây có thể thay đổi về số lượng, chất lượng, hiệu quả sao cho thích hợp với yêu cầu ứng dụng riêng của từng đơn vị.

Một đơn vị thông tin có thể sử dụng phương tiện kỹ thuật của một đơn vị khác để tin học hoá hoạt động của mình mà không cần phải trang bị một hệ tin học đầy đủ với chi phí cao bằng cách hợp tác với các trung tâm tin học hoặc tham gia vào một mạng lưới thông tin.

2.1. Nhân sự trong hệ thống thông tin tự động hoá

Một hệ thống thông tin tự động hoá không thể thiếu yếu tố con người. Đó là các cán bộ chuyên môn có trách nhiệm thiết kế, xây dựng và khai thác hệ thống thông tin. Để xây dựng một hệ thống thông tin tự động hoá cần phải có các loại cán bộ sau:

- Các kỹ sư tin học, có trách nhiệm giám sát việc vận hành và bảo trì hệ thống, tức là giám sát sự hoạt động của nó.

- Các nhà phân tích hệ thống, có trách nhiệm nghiên cứu các ứng dụng, tức là nghiên cứu các công việc phải tin học hoá và chuẩn bị cho nó thích ứng với máy tính điện tử.

- Các nhà lập trình, có trách nhiệm biên soạn các chương trình cho máy tính nhằm giải quyết các công việc do các nhà phân tích hệ thống đặt ra.

- Các thao tác viên. Đó là những người vận hành và khai thác sự hoạt động của hệ thống. Một nhiệm vụ quan trọng của nhóm này là nhập dữ liệu và sử dụng hệ thống khi đã hoàn thiện.

- Các cán bộ chuyên môn trong lĩnh vực thông tin - thư viện.

Trong điều kiện hiện nay khó có thể có được một đội ngũ cán bộ chuyên môn đầy đủ về mặt chức danh như thế. Tùy theo quy mô của hệ thống, một số chức năng có thể do một người đảm nhiệm. Ví dụ: phân tích hệ thống và lập trình. Mặt khác các cơ quan thông tin - thư viện cần hợp tác với các trung tâm tin học để nhận được sự hỗ trợ về mặt kỹ thuật, thiết kế và cài đặt hệ thống. Đối với các trung tâm lớn cần có các kỹ sư tin học được đào tạo chính quy để có thể bảo trì sự vận hành của hệ thống và xây dựng các phần mềm chuyên dụng. Các cán bộ này cần có sự hợp tác về nghiệp vụ của các cán bộ thư viện và thông tin.

2.2. Máy tính điện tử và các thiết bị ngoại vi

Máy tính điện tử (MTĐT) là một thiết bị điện tử xử lý rất nhanh các thông tin đưa vào và hoạt động dưới sự điều khiển của chương trình lưu trữ trong bộ nhớ. Chương trình bao gồm những lệnh được sắp xếp một cách hợp lý và có thể giao cho máy thực hiện. Máy tính thu nhận, lưu trữ các dữ liệu và thực hiện các phép toán số học hay logic trên các dữ liệu đó mà không cần có sự can thiệp trực tiếp của con người.

Máy tính điện tử có ba khả năng: quản lý, xử lý và trao đổi dữ liệu.

Máy tính điện tử bao gồm hai bộ phận cơ bản sau:

- Bộ xử lý trung tâm, dùng để xử lý thông tin.
- Các thiết bị ngoại vi, bảo đảm việc vào, ra dữ liệu và lưu trữ thông tin.

Bộ xử lý trung tâm (CPU) có ba khối:

- Khối điều khiển (CU): cấp trình tự các thao tác nhỏ nhất cần làm đối với mỗi lệnh của MTĐT bằng các tín hiệu điều khiển tương ứng.

- Khối tính toán (ALU): thực hiện các phép toán số học và logic.

- Bộ nhớ trong (RAM, ROM): chứa các chương trình và dữ liệu.

ROM (Read Only Memory) là bộ nhớ chứa các chương trình điều khiển căn bản, các phần mềm hệ thống và ứng dụng do hãng sản xuất cung ứng. Các chương trình trong ROM không thể thay đổi, không bị mất đi khi tắt máy. ROM chỉ cho phép đọc các thông tin chứa trong đó.

RAM (Random Access Memory) là bộ nhớ chứa các chương trình và dữ liệu khi làm việc. Khi tắt máy thông tin trong RAM sẽ bị mất đi, do đó các thông tin cần lưu trữ phải được ghi trên đĩa cứng hoặc đĩa mềm. Dung lượng của RAM quyết định tốc độ và tính linh hoạt của máy tính.

Các thiết bị ngoại vi có:

- Bộ nhớ ngoài: băng từ, đĩa từ dùng để lưu trữ thông tin.

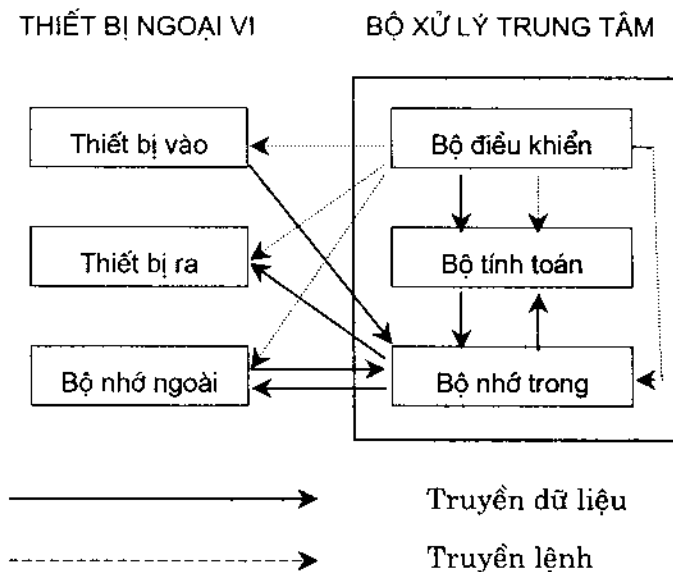
- Bộ nhớ trong chỉ lưu trữ thông tin đang được bộ vi xử lý làm việc. Để lưu trữ một lượng thông tin lớn hơn (các chương trình và dữ liệu) người ta phải sử dụng bộ nhớ ngoài. Trong máy

vì tính đó là các đĩa từ, bao gồm đĩa cứng và đĩa mềm, kèm theo là các thiết bị dùng để đọc và ghi gọi là ổ đĩa (Disk Drive).

- Thiết bị vào: bàn phím, con chuột, máy quét hình,... dùng để đọc thông tin cho MTĐT.

- Thiết bị ra: màn hình, máy in, máy vẽ,... dùng để đưa thông tin ra.

Ta có thể biểu diễn mối quan hệ giữa chúng bằng sơ đồ sau:



Hình 1. Quan hệ giữa bộ xử lý trung tâm và các thiết bị ngoại vi

Các thông tin đưa vào máy tính được biểu diễn dưới dạng số nhị phân, nghĩa là chỉ gồm hai số 0 và 1, ứng với hai trạng thái của vật mang tin từ tính có bị cảm ứng hay không và do đó có nhận thông tin hay không. Khi đó mỗi ký tự được biểu diễn dưới dạng một chuỗi các số nhị phân. Chuỗi đó gọi là mã. Có nhiều hệ thống mã dùng cho các máy tính điện tử khác nhau. Có hai hệ thống mã quan trọng là:

- EBCDIC (Extended Binary Coded Decimal Interchange Code).

- ASCII (American Standart Code for Information Interchange).

Ví dụ trong hệ mã ASCII :

số 1 có mã là : 00110001

chữ A có mã là: 01000001

chữ a có mã là : 01100001

Mỗi yếu tố thông tin 0 hoặc 1 gọi là một bit, 8 bit là một byte. Mỗi byte biểu diễn một ký tự.

Các đơn vị bội của byte (viết tắt là B) là: Kilobyte (KB) bằng 1024 B, Megabyte (MB) bằng 1024 KB, Gigabyte (GB) bằng 1024 MB.

Những đặc tính của một thiết bị tin học được thể hiện ở những yếu tố sau:

- Dung lượng của bộ nhớ, thể hiện bằng số ký tự hay số byte mà bộ nhớ có thể lưu trữ.
- Thời gian tiếp cận thông tin, ứng với thời gian thực hiện một lệnh hay thời gian truy nhập thông tin trong bộ nhớ, thể hiện bằng đơn vị một phần triệu giây.
- Tốc độ đọc, ghi và truyền thông tin. Chẳng hạn tốc độ đọc có thể từ 20.000 đến 200.000 ký tự /giây, tốc độ in có thể từ 100 đến 1500 dòng/phút.

2.3. Các vật mang tin điện tử

Công nghệ thông tin sử dụng các vật mang tin đặc biệt để ghi các thông tin dưới dạng có thể đọc được bằng máy. Đó có thể là băng đục lỗ, băng từ, đĩa từ hoặc đĩa quang.

Băng đục lỗ là băng giấy rộng 25,4 mm, trên đó có thể đục bảy hàng lỗ và một hàng dùng để kiểm tra. Mỗi cột lỗ có thể nhận một ký tự. Băng đục lỗ hoạt động theo nguyên lý của quang điện và được sử dụng đối với máy tính ở thế hệ đầu.

Băng từ là băng nhựa trên đó có phủ một chất có khả năng nhiễm từ. Mặt băng rộng 12,7 mm được chia thành 7 hay 9 rãnh từ, rãnh cuối cùng dùng để kiểm tra. Các ký tự phân cách nhau bằng một khoảng cách xác định. Một băng dài 730 m có thể ghi được 46 triệu ký tự.

Đĩa từ là một đĩa bằng kim loại hoặc chất dẻo, trên mặt đĩa có phủ một chất có khả năng nhiễm từ. Việc lưu trữ thông tin trên đĩa từ được thực hiện dựa trên cơ sở công nghệ điện từ.

Có hai loại đĩa từ là đĩa cứng và đĩa mềm. Thông tin được ghi trên đĩa theo từng vòng tròn đồng tâm, gọi là rãnh (track), mỗi rãnh lại chia thành cung (sector). Cung là đơn vị nhỏ nhất để ghi thông tin, mỗi cung có thể ghi được 512 byte. Những đặc trưng của các loại đĩa này là tốc độ quay, tốc độ truy nhập và chuyển giao thông tin, dung lượng nhớ. Chúng thay đổi tùy theo loại đĩa.

Hiện nay các máy tính cá nhân thường sử dụng loại đĩa mềm có đường kính 3,5 inch, dung lượng nhớ 1,4 MB, tốc độ quay 600 vòng/phút, tốc độ truyền tin khoảng 25 KB/giây... Đĩa mềm có thể nạp vào và lấy ra từ ổ đĩa mềm, còn đĩa cứng thì gắn liền với ổ đĩa. Tốc độ quay của đĩa cứng lên tới 3600

vòng/phút và dung lượng nhớ có thể từ 40 MB đến hàng chục GB, tốc độ truyền tin khoảng 5 MB/giây.

Đĩa quang là vật mang tin quang học. Đĩa quang xuất hiện vào đầu những năm 80. Đĩa quang được chế tạo dựa trên cơ sở công nghệ laser. Ở đây các dữ liệu số được ghi lại bằng cách đốt cháy một dãy các lỗ cực nhỏ nhờ một chùm tia laser chiếu vào một phim kim loại mỏng láng trên một đĩa nhựa có đường kính 12 cm. Bằng cách đó các thông tin trên đĩa từ được mã hoá và ghi lại trên một đĩa chủ, rồi từ đó được nhân bản lên. Có nhiều loại đĩa quang. Đĩa quang được sử dụng nhiều trong hoạt động thông tin khoa học hiện nay là CD-ROM. Đó là thiết bị nhớ có khả năng ghi một lần nhưng có thể đọc nhiều lần (ROM - Read Only Memory). Nó thường được dùng để lưu trữ các thông tin với khối lượng lớn nhưng không có nhu cầu phải thay đổi hoặc cập nhật bởi người sử dụng.

Với các lợi thế nổi bật như kích thước nhỏ, dung lượng lớn, độ bền cao, đã được tiêu chuẩn hoá, có phần mềm khai thác nằm ngay trên đĩa, các đĩa quang CD-ROM ngày nay được sử dụng rộng rãi để lưu trữ, phân phối thông tin và triển khai các ứng dụng multimedia.

CD-ROM có dung lượng lớn, khoảng 600 MB thông tin, tương đương với khoảng 300.000 trang in (600 quyển sách, mỗi quyển 500 trang), do đó ứng dụng chính của CD-ROM là dùng để lưu trữ thông tin. Bất cứ dạng dữ liệu nào, chữ viết, âm thanh, hình ảnh tĩnh và động, sau khi số hoá đều có thể được lưu trữ trên CD-ROM dưới những dạng thức nhất định. Vì vậy sách báo, tạp chí, âm nhạc, phim ảnh, chương trình máy tính v.v... đều có thể được ghi vào đĩa CD-ROM. Mỗi đĩa CD-ROM chứa được khoảng 700 triệu ký tự hoặc hàng ngàn hình ảnh đồ hoạ, hoặc 18 giờ âm thanh, hoặc 74 phút phim với hình ảnh động.

Điều đặc biệt có ý nghĩa là mỗi CD-ROM đều có một phần mềm khai thác lưu trữ ngay trên đĩa (thường chiếm khoảng 1/2 MB), với giao diện sử dụng thích hợp giúp ta dễ dàng tìm kiếm và truy nhập tới các thông tin ghi trên đĩa.

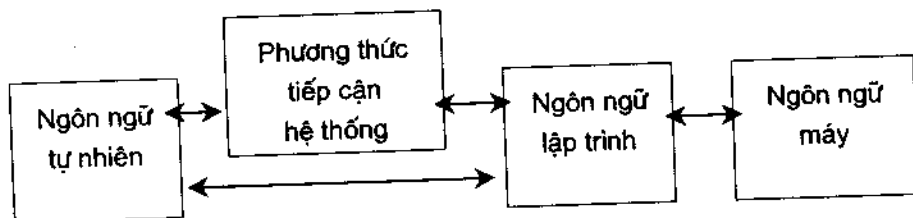
Trong những năm qua, lượng thông tin khổng lồ chứa trong các CSDL, các sách tham khảo, cẩm nang kỹ thuật, các phần mềm máy tính, v.v... đã được phát hành dưới dạng CD-ROM.

Công nghệ CD-ROM cung cấp một phương tiện tốt để lưu trữ các bộ bách khoa toàn thư, các bộ từ điển lớn. Một bách khoa toàn thư hàng chục tập chỉ cần lưu trữ trên một CD-ROM. Các thông tin ở đây có thể phân phối dưới dạng có thể đọc lướt qua, được dò tìm theo các đề mục, các thông tin tham khảo chéo với hình ảnh, âm thanh hay kết nối với các thông tin khác chi tiết hơn.

2.4. Các phần mềm hệ thống và phần mềm chuyên dụng

Phần mềm là tập hợp có cấu trúc các lệnh cho phép thực hiện các công việc tính toán và quản lý. Các lệnh này có thể biểu diễn bằng ngôn ngữ máy, dựa trên các số nhị phân hoặc bằng ngôn ngữ lập trình, nó sẽ được máy tính dịch ra ngôn ngữ máy.

Sự giao lưu giữa người sử dụng và máy tính trải qua các khâu sau:



Hình 2. Từ ngôn ngữ tự nhiên đến ngôn ngữ máy

Có hai loại phần mềm là phần mềm hệ thống và phần mềm chuyên dụng.

Phần mềm hệ thống là tập hợp các tiến trình điều phối, quản lý, cấp phát, khai thác các tài nguyên của máy (bộ xử lý trung tâm, bộ nhớ, các thiết bị vào ra, các chương trình tiện dụng, các bộ chương trình dịch...) và đáp ứng các ứng dụng của người sử dụng. Có thể nói phần mềm hệ thống là tập hợp các chương trình cần thiết để điều khiển toàn bộ hoạt động của máy tính như: khởi động máy tính, quản lý bộ nhớ, điều khiển việc vào ra dữ liệu, điều khiển việc thực hiện các chương trình v.v...

Hạt nhân của phần mềm hệ thống là hệ điều hành và các bộ chương trình dịch. Hiện có nhiều hệ điều hành khác nhau như DOS, UNIX, Windows,...

Phần mềm chuyên dụng là phần mềm được biên soạn dành cho người sử dụng nhằm giải quyết một nhiệm vụ xác định. Phần mềm chuyên dụng có rất nhiều. Một số được cung cấp bởi các nhà sản xuất, một số có thể do các tổ chức chuyên môn xây dựng, một số khác có thể do người sử dụng soạn thảo tùy theo yêu cầu của họ.

Ta có thể kể một số phần mềm chuyên dụng quen thuộc:

Soạn thảo văn bản có WORD, WordPerfect.

Bảng tính điện tử có LOTUS, QUATTRO, EXCEL.

Thư tín điện tử có MS.MAI, OUTLOOK EXPRESS.

Thiết kế có AUTOKED.

Quản lý các cơ sở dữ liệu có FOXPRO, ACCESS,...

Soạn nhạc có ENCORE.

v.v...

Về nguyên tắc, mỗi phần mềm dùng cho một ứng dụng xác định. Cũng có phần mềm có nhiệm vụ giải quyết nhiều khâu công việc của một hệ thống phức hợp.

Mỗi phần mềm là một tổ hợp các chương trình, mỗi chương trình giải quyết một nhiệm vụ cơ bản nào đó. Việc biểu diễn dữ liệu vào, ra cũng như cấu trúc của các tệp dữ liệu phải được xây dựng chặt chẽ bởi các format. Một số phần mềm có thể chuyển từ format này sang format khác. Phân tích hệ thống là kỹ thuật cơ bản để biên soạn các phần mềm. Phân tích hệ thống bao gồm hai khâu là phân tích chức năng và phân tích tổ chức. Người ta dùng sơ đồ khối để mô tả các kết quả phân tích.

Chương trình được thể hiện bằng một dãy có cấu trúc các lệnh. Mỗi chương trình đều mô tả các dữ liệu cần xử lý, cho chúng những địa chỉ xác định trong bộ nhớ và chỉ ra một dãy các lệnh tương ứng với một dãy công việc cần thực hiện.

Trong các hệ thống thông tin - thư viện tự động hoá, phần mềm quản lý và tìm kiếm tài liệu gọi là phần mềm tư liệu. Ví dụ CDS/ISIS là một phần mềm tư liệu đang được sử dụng rộng rãi ở nước ta hiện nay. Ngoài ra còn có những phần mềm tích hợp quản trị thư viện, cho phép quản lý toàn diện các chức năng hoạt động của một thư viện như: bổ sung, biên mục, quản lý việc lưu thông tài liệu, quản lý kho, tìm tin, v.v...

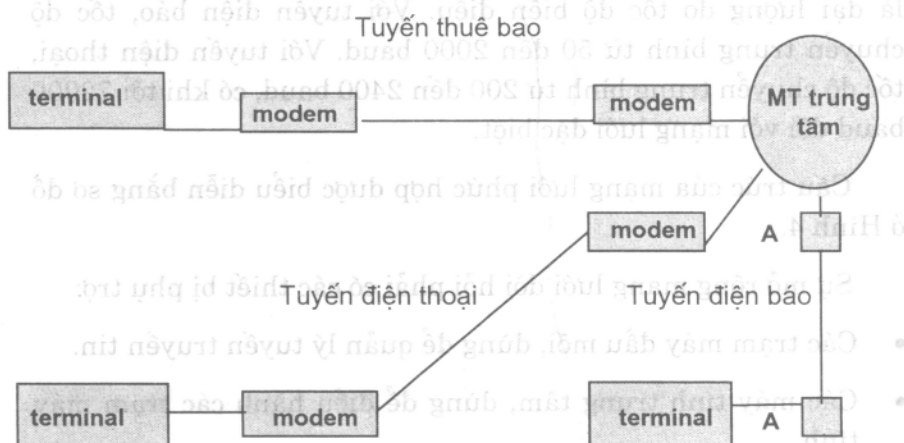
Ta sẽ trình bày sâu hơn các phần mềm này ở chương sau.

2.5. Các thiết bị viễn thông

Sự kết hợp giữa máy tính và các hệ thống truyền thông, đặc biệt là viễn thông đã tạo ra sự chuyển biến có tính cách mạng trong các hệ thống giao lưu thông tin. Đó là sự ra đời của các mạng thông tin máy tính, ở đó các máy tính điện tử ở đầu cuối được nối với đơn vị trung tâm bằng mạng lưới viễn thông, cho

phép con người có thể chia sẻ nguồn lực thông tin và khai thác thông tin từ xa.

Cấu trúc của mạng viễn thông đơn giản được biểu diễn bằng sơ đồ ở Hình 3.



Hình 3. Sơ đồ mạng viễn thông đơn giản

Thiết bị của hệ thống thông tin bao gồm:

- Một thiết bị cho phép chuyển tín hiệu nhị phân sản sinh bởi máy tính điện tử thành tín hiệu có thể truyền theo tuyến viễn thông và ngược lại. Đó có thể là các biến điệu điện báo (adapter telegraphic), nhưng phổ biến nhất là các modem (modulator - demodulator). Chúng có thể nối trực tiếp với tuyến điện thoại và ngược lại bằng phương tiện của một bộ ghép nối âm thanh (coupler acoustic). Modem là thiết bị đặt giữa máy tính điện tử hay đầu cuối và một tuyến viễn thông, nó dùng để biến tín hiệu số thành tín hiệu điện, cho phép truyền các tín hiệu đã số hóa.

- Các tuyến viễn thông có thể là mạng lưới điện thoại công cộng hay mạng lưới dành cho việc chuyển giao dữ liệu, hay tuyến thuê bao dành cho người dùng tin.

Năng lực của các tuyến viễn thông thể hiện bằng baud. Đó là đại lượng đo tốc độ biến điệu. Với tuyến điện báo, tốc độ chuyển trung bình từ 50 đến 2000 baud. Với tuyến điện thoại, tốc độ chuyển trung bình từ 200 đến 2400 baud, có khi tới 72000 baud đối với mạng lưới đặc biệt.

Cấu trúc của mạng lưới phức hợp được biểu diễn bằng sơ đồ ở Hình 4.

Sự mở rộng mạng lưới đòi hỏi phải có các thiết bị phụ trợ:

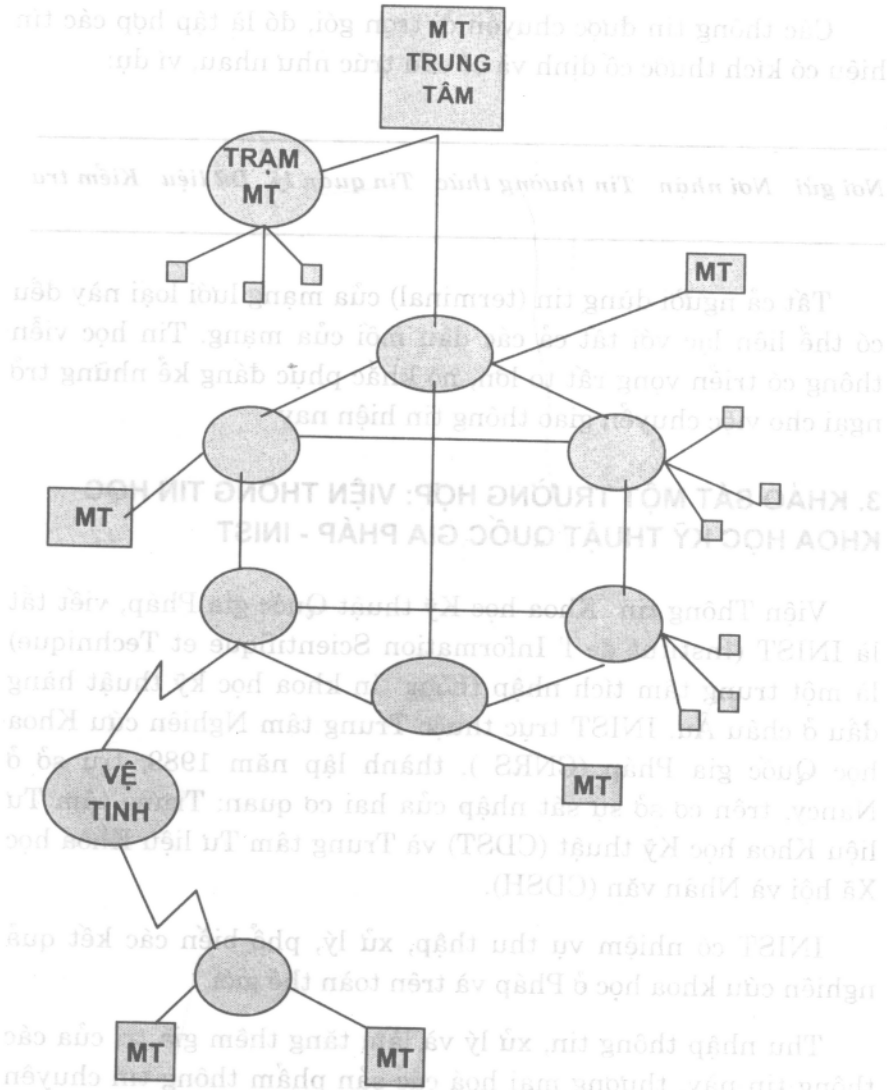
- Các trạm máy đầu mối, dùng để quản lý tuyến truyền tin.
- Các máy tính trung tâm, dùng để điều hành các trạm máy tính.

Sự phát triển của tin học viễn thông dẫn đến xuất hiện các terminal “thông minh”, nó thực sự là các máy tính điện tử có thể bảo đảm được một vài chức năng xử lý.

Mạng lưới phức hợp có hai mức độ:

- Mức độ truyền tin, bao gồm các trạm máy tính đầu mối và các tuyến truyền tin nối giữa chúng, có chức năng chuyển giao thông tin trong mạng lưới viễn thông.

- Mức độ xử lý và sử dụng, tạo thành bởi các trạm đầu mối và các máy tính điện tử ở đầu cuối. Ở đây các cơ quan dịch vụ thông tin (server) sẽ làm môi giới giúp người dùng tin tiếp cận các cơ sở dữ liệu.



Hình 4. Sơ đồ của mạng viễn thông phức hợp

Các thông tin được chuyển đi trọn gói, đó là tập hợp các tin hiệu có kích thước cố định và có cấu trúc như nhau, ví dụ:

Nơi gửi Nơi nhận Tin thường thức Tin quản lý Dữ liệu Kiểm tra

Tất cả người dùng tin (terminal) của mạng lưới loại này đều có thể liên lạc với tất cả các đầu mối của mạng. Tin học viễn thông có triển vọng rất to lớn, nó khắc phục đáng kể những trở ngại cho việc chuyển giao thông tin hiện nay.

3. KHẢO SÁT MỘT TRƯỜNG HỢP: VIỆN THÔNG TIN HỌC KHOA HỌC KỸ THUẬT QUỐC GIA PHÁP - INIST

Viện Thông tin Khoa học Kỹ thuật Quốc gia Pháp, viết tắt là INIST (Institut de l'Information Scientifique et Technique) là một trung tâm tích nhập thông tin khoa học kỹ thuật hàng đầu ở châu Âu. INIST trực thuộc Trung tâm Nghiên cứu Khoa học Quốc gia Pháp (CNRS), thành lập năm 1989, trụ sở ở Nancy, trên cơ sở sự sát nhập của hai cơ quan: Trung tâm Tư liệu Khoa học Kỹ thuật (CDST) và Trung tâm Tư liệu Khoa học Xã hội và Nhân văn (CDSH).

INIST có nhiệm vụ thu thập, xử lý, phổ biến các kết quả nghiên cứu khoa học ở Pháp và trên toàn thế giới.

Thu nhập thông tin, xử lý và làm tăng thêm giá trị của các thông tin này, thương mại hoá các sản phẩm thông tin chuyên dạng... Tất cả tạo thành một kho tàng tri thức cho phép các nhà chuyên môn tiếp cận các kết quả nghiên cứu trên tất cả các lĩnh vực. Đó là mục tiêu của INIST. Tham vọng này của INIST dựa trên một hệ thống công nghệ cao, một chính sách hợp tác tích

cực, nhưng trước hết được bảo đảm bởi các chuyên gia của INIST.

Nhân sự

INIST có 400 nhân viên có năng lực trong lĩnh vực thông tin: các cán bộ thư viện, cán bộ thông tin, các nhà khoa học, các kỹ sư, bác sĩ, các kỹ sư tin học, các nhà doanh nghiệp...

Phương tiện kỹ thuật và công nghệ thông tin

Cơ sở hạ tầng kỹ thuật tin học của INIST bao gồm nhiều thiết bị kết nối với nhau thành mạng lưới có khả năng trao đổi thông tin với khối lượng lớn.

- Một máy tính trung tâm IBM 9121 . 210 bảo đảm các chức năng quản lý: cung cấp tài liệu cấp một, cơ sở dữ liệu thư mục, chọn lọc các thông tin để sản xuất các sản phẩm thông tin có chất lượng cao (các bộ tra cứu thư mục, các ấn phẩm thông tin, CD - ROM...). Hơn một trăm đầu cuối được nối với hệ thống này cho phép 40 Minitel và đầu cuối có thể tiếp cận thông tin cùng một lúc.

- Một phần mềm quản lý thư viện tự động hoá GEAC 9000, dùng để quản lý việc bổ sung tài liệu, biên mục tự động, quản lý việc mượn và phân phối ấn phẩm thông tin.

- Một mạng lưới gồm 130 trạm máy tính (PC 386 và 486) nối với một máy chủ HP 9000/UNIX/TCP . IP cho phép đánh chỉ số các sách chuyên khảo và ấn phẩm định kỳ để xây dựng các cơ sở dữ liệu thư mục.

- Một hệ thống lưu trữ thông tin dưới dạng số hóa trên đĩa quang (DON – Disque Optique Numérique).

- Các phương tiện viễn thông và thiết bị văn phòng.

Vốn tài liệu của INIST

Theo số liệu năm 1994, INIST lưu trữ một khối lượng tài liệu xếp trên giá dài tới 34 km. Mỗi năm có tới từ 2 đến 3 triệu trang in dưới dạng số hoá của 1700 tạp chí quan trọng nhất được lưu trữ trên đĩa quang.

Kho tài liệu của INIST bao gồm:

- 28.000 tệp tạp chí, trong đó 19.000 tên tạp chí đang đặt mua hàng năm.
- 41.000 báo cáo khoa học của Pháp.
- 50.000 tổng kết của các hội nghị quốc tế.
- 69.000 luận án khoa học của Pháp.

Các tài liệu trên bao quát các lĩnh vực sau:

- Khoa học ứng dụng	17%
- Hóa học	6%
- Vật lý	5%
- Khoa học xã hội và nhân văn	23%
- Khoa học thông tin	2%
- Khoa học về trái đất, đại dương, vũ trụ	9%
- Tâm lý học	3%
- Sinh học	19%
- Y học, Dược học	16%

Sản xuất cơ sở dữ liệu

INIST sản xuất hai cơ sở dữ liệu thư mục đa ngành, đa ngôn ngữ, đó là cơ sở dữ liệu PASCAL và FRANCIS.

PASCAL là cơ sở dữ liệu đa ngành, đa ngôn ngữ, bao quát các lĩnh vực khoa học, công nghệ và y học.

- Bắt đầu xây dựng năm 1973.
- Số biểu ghi tra cứu đến năm 1994 là hơn 10 triệu.
- Mức tăng trưởng hàng năm: 600.000 biểu ghi.
- Chu kỳ cập nhật: 1 tháng.
- Ngôn ngữ sử dụng: Pháp, Anh, Tây Ban Nha.

Dữ liệu trong CSDL PASCAL là các bài báo của 8200 tạp chí khoa học của Pháp và thế giới, các báo cáo khoa học, các luận án, các tổng kết hội nghị, các công trình nghiên cứu.

FRANCIS là tập hợp 19 cơ sở dữ liệu thư mục đa ngành, bao quát các lĩnh vực khoa học nhân văn và kinh tế.

- Bắt đầu xây dựng năm 1972.
- Số biểu ghi tra cứu đến năm 1994 là 1,5 triệu.
- Mức tăng trưởng hàng năm: 80.000 biểu ghi.
- Chu kỳ cập nhật: 3 tháng.
- Ngôn ngữ sử dụng: Pháp, Anh.

85% dữ liệu trong CSDL FRANCIS là các bài báo của 7200 tạp chí khoa học của Pháp và thế giới.

Các dịch vụ thông tin của INIST

INIST thực hiện các dịch vụ thông tin phong phú và đa dạng:

1. Dịch vụ cung cấp tư liệu

Mỗi ngày INIST nhận và xử lý trung bình 3000 yêu cầu tin. Ngay sau khi nhận được, yêu cầu của người dùng tin được xử lý

tự động, tài liệu lưu trữ trong kho sẽ được chụp và gửi cho người dùng tin bằng đường bưu điện hoặc fax. Yêu cầu của người dùng tin sẽ được đáp ứng trong vòng 48 giờ với dịch vụ thông thường (Service courant) và trong vòng 24 giờ với dịch vụ nhanh (Service express). Mỗi năm INIST cung cấp gần 600.000 bản sao chụp tài liệu cho 6000 khách hàng ở Pháp và trên thế giới.

2. Dịch vụ biên dịch tài liệu

INIST nhận dịch tài liệu khoa học kỹ thuật cho mọi đối tượng hoặc tìm kiếm các bản dịch đã có sẵn để cung cấp cho họ. INIST hợp tác với Trung tâm Dịch thuật Quốc tế ITC (Internationl Translation Centre) sản xuất cơ sở dữ liệu WTI (World Traslation Index) cho phép xác định các bản dịch đã có trước khi tiến hành dịch theo yêu cầu của khách hàng. Từ năm 1979 WTI đã tiến hành dịch 300.000 tài liệu khoa học kỹ thuật ra tiếng các nước phương Tây.

3. Dịch vụ tìm tin on-line

Tìm tin on-line là tìm tin trên cơ sở dữ liệu thông qua một cơ quan dịch vụ thông tin (server) dựa trên một mạng lưới truyền dữ liệu. Bằng phương thức này người dùng tin có thể tiếp cận trực tiếp trên 11 triệu bản tra cứu thư mục trong hầu hết các lĩnh vực của tri thức trên hai cơ sở dữ liệu PASCAL và FRANCIS.

Có hai phương thức tìm tin on-line là:

a. Tìm tin on- line qua cơ quan dịch vụ thông tin

PASCAL do QUESTEL, DIALOG và ESAIRS đảm nhiệm.

FRANCIS do QUESTEL đảm nhiệm.

Để tìm tin on-line qua các cơ sở dữ liệu PASCAL và FRANCIS bằng phương thức này cần phải:

- Ký hợp đồng với các cơ quan dịch vụ trên để nhận được mật khẩu (Password).

- Sử dụng các thiết bị sau: một đầu cuối của máy tính, một máy vi tính có trang bị modem hay một Minitel.

- Biết sử dụng ngôn ngữ hỏi riêng của từng cơ quan dịch vụ.

Tìm tin on-line qua cơ quan dịch vụ thông tin là phương thức tìm tin cho phép nhận ngay được thông tin thích hợp một cách nhanh chóng và linh hoạt.

b. Tìm tin on-line bằng Minitel

Tìm tin on-line trên cơ sở dữ liệu PASCAL và FRANCIS bằng phương thức này chỉ cần sử dụng Minitel với ký số 36 29 36 01.

Để giúp bạn sử dụng các dịch vụ tìm tin trên INIST cung cấp một loạt các tài liệu trợ giúp sau:

- Sách hướng dẫn sử dụng Minitel hoặc CD-ROM.

- Sách hướng dẫn hỏi đáp trên cơ sở dữ liệu PASCAL hay FRANCIS.

- Tập từ vựng các từ khoá.

- Khung phân loại của PASCAL hay FRANCIS.

Ngoài ra còn có một bộ phận tư vấn và đào tạo để giúp cho người dùng tin có thể sử dụng thuận tiện các dịch vụ tìm tin của INIST.

4. Tìm tin trên CD- ROM

Từ năm 1987 cơ sở dữ liệu PASCAL được ghi trên CD-ROM. Mỗi đĩa chứa nội dung trong một năm của cơ sở dữ liệu này, bao gồm khoảng 500.000 biểu ghi. Với một PC có trang bị đầu đọc

đĩa quang, CD-ROM PASCAL cho phép bạn có thể tìm ngay được các tra cứu thư mục và in các kết quả ra đĩa hay ra giấy mà không gặp trở ngại về thời gian.

Đĩa quang đầu tiên của FRANCIS ghi lại nội dung của cơ sở dữ liệu này trong 7 năm (1984 - 1990), chứa khoảng 500.000 biểu ghi. Hàng năm, một đĩa cập nhật toàn bộ nội dung của FRANCIS được xây dựng trong năm đó. CD-ROM PASCAL và FRANCIS sử dụng:

- Hai hình thức tra cứu là: tra cứu có trợ giúp và tra cứu ở trình độ cao (tra cứu kiểu chuyên gia).
- Hai ngôn ngữ hỏi: Anh hoặc Pháp. .
- Nhiều dấu hiệu tìm kiếm như: lĩnh vực, từ vựng, tác giả, cơ quan, từ khoá, ngôn ngữ, ngày xuất bản, loại hình tài liệu, ký hiệu phân loại.

5. Xuất bản ấn phẩm thư mục

Hàng tháng cơ sở dữ liệu PASCAL xuất bản 63 ấn phẩm thư mục theo từng lĩnh vực dưới dạng giấy hay microfiche. Các ấn phẩm này được tập hợp theo 4 chủ đề lớn:

- Vật lý, hoá học, khoa học ứng dụng.
- Sinh học cơ bản và ứng dụng.
- Y học.
- Khoa học về trái đất, đại dương và vũ trụ.

Ba tháng một lần, cơ sở dữ liệu FRANCIS xuất bản 16 ấn phẩm thư mục theo từng lĩnh vực dưới dạng giấy hoặc microfiche. Các thư mục này được sắp xếp theo khung phân loại có kèm theo chỉ số hướng dẫn, cho phép tìm tin một cách nhanh chóng.

6. Phổ biến thông tin có chọn lọc

Dịch vụ phổ biến thông tin có chọn lọc phục vụ yêu cầu thông tin theo chủ đề với sự trợ giúp của các chuyên gia được đào tạo chuyên để tra cứu các cơ sở dữ liệu Pascal và Francis. Dịch vụ này được tiến hành dựa trên các thư mục hồi cố của cá nhân, các profil cá nhân, các profil chuẩn.

7. Cung cấp thông tin khoa học kỹ thuật của Nhật Bản

Nhờ mối quan hệ chặt chẽ với giới khoa học, kinh tế, kỹ nghệ Nhật Bản, bộ phận Nhật Bản của INIST bảo đảm cung cấp cho người dùng tin các thông tin khoa học cần thiết của Nhật Bản với cơ sở dữ liệu JAPINFO.

Dưới tác động của khoa học kỹ thuật hiện đại và công nghệ thông tin mới, thế giới chúng ta đang thay đổi từng ngày hướng tới xã hội của một nền văn minh mới: xã hội thông tin. Trong sự thay đổi đó có phần đóng góp tích cực của INIST.

Chương 2

CẤU TRÚC DỮ LIỆU

1. THÔNG TIN VÀ DỮ LIỆU

1.1. Thông tin

Thông tin (Information) là khái niệm cơ bản của khoa học cũng là khái niệm trung tâm của xã hội trong thời đại chúng ta. Mọi quan hệ, mọi hoạt động của con người đều dựa trên một hình thức giao lưu thông tin nào đó. Mọi tri thức đều bắt nguồn bằng một thông tin về những điều đã diễn ra, về những cái người ta đã biết, đã nói, đã làm và điều đó luôn xác định bản chất và chất lượng của những mối quan hệ của con người.

Vậy thông tin là gì?

Theo nghĩa thông thường: Thông tin là tất cả các sự việc, sự kiện, ý tưởng, phán đoán làm tăng thêm sự hiểu biết của con người. Thông tin hình thành trong quá trình giao tiếp: một người có thể nhận thông tin trực tiếp từ người khác thông qua các phương tiện thông tin đại chúng, từ các ngân hàng dữ liệu hoặc từ tất cả các hiện tượng quan sát được trong môi trường xung quanh.

Trên quan điểm triết học: Thông tin là tất cả những gì được phản ánh trong thế giới vật chất. Thông tin là nội dung của sự phản ánh mà con người nhận thức được phụ thuộc vào đặc điểm và tính chất của đối tượng được phản ánh. Như vậy mọi đối

tượng vật chất đều có thể coi là những nguồn tin (vì nó có tính chất tự phản ánh - theo Lê Nin). Nhưng đó là thông tin ở dạng tiềm năng. Để có được thông tin cần phải có đối tượng thu nhận tin.

Quá trình tác động qua lại giữa nguồn tin và đối tượng thu nhận tin gọi là quá trình thông tin. Do đó lịch sử phát triển của thông tin gắn liền với lịch sử phát triển của kỹ thuật truyền tin. Các mốc ghi dấu lịch sử phát triển kỹ thuật truyền tin là: tiếng nói, chữ viết, nghề in và kỹ thuật thông tin hiện đại.

Giao lưu thông tin

Thông tin tiềm tàng khắp nơi trong xã hội. Đó là các nguồn thông tin về lao động, đất đai, tài nguyên, môi trường; thông tin về các tổ chức và các hoạt động kinh tế, xã hội; thông tin về khoa học và công nghệ; thông tin về sản xuất, kinh doanh v.v... Nhưng thông tin chỉ có giá trị và ý nghĩa khi nó được truyền đi, phổ biến và được sử dụng. Có thể nói bản chất của thông tin nằm trong sự giao lưu của nó. Nói cách khác *thuộc tính cơ bản của thông tin là giao lưu*. Thông tin được giao lưu bằng ngôn từ, hình ảnh, tiếng động, thư tín, điện tín, điện thoại hoặc bất kỳ một phương tiện vật lý cụ thể nào khác.

Để phân biệt nội dung thông tin cần truyền đi và cách thức chuyển giao thông tin người ta tách nội dung thông tin ra khỏi hình thức biểu diễn nó. Các hình thức biểu diễn thông tin (các ký hiệu, dấu hiệu, hình ảnh ...) là hữu hạn nhưng nội dung của thông tin (khái niệm, ý tưởng, sự kiện, tên v.v...) thì vô hạn. Trong trường hợp thông tin có một hình thức biểu diễn, quá trình chuyển giao thông tin chính là quá trình truyền các ký hiệu biểu diễn nó. Những ý tưởng mới sẽ được truyền đi bằng một tổ hợp mới của một số hữu hạn các ký hiệu (chữ cái, chữ số...). Trong đời sống hàng ngày thông tin được biểu diễn bằng

ngôn ngữ. Khi đó thông tin được diễn tả bằng cú pháp và ngữ nghĩa của ngôn ngữ mà người ta sử dụng.

Lý thuyết thông tin xác nhận rằng càng nhiều tín hiệu sinh ra từ nguồn tin thì càng có nhiều thông tin được truyền đi. Khi đó thông tin được mô tả bằng sự thống kê và tổ hợp các dấu hiệu phát ra từ nguồn.

Sự giao lưu thông tin giữa con người trong thực tế bị chi phối bởi không gian và thời gian. Nó chỉ có thể bền vững nếu nó để lại một dấu vết, tức là nó phải được ghi trên một cái giá vật chất, gọi là vật mang tin. Đó có thể là sách vở, tranh ảnh, đĩa, băng từ,... Tóm lại đó là một tư liệu (document).

Chọn lọc từ khối lượng lớn thông tin mang trên những giá đó những tri thức mà người ta yêu cầu, cung cấp cho người dùng tin vào lúc mà họ cần, đồng thời giữ cho nó không bị lạc hậu bằng cách bổ sung, cập nhật hàng ngày. Đó là chức năng của hoạt động thông tin tư liệu hay còn gọi là thông tin khoa học.

Nhu cầu thông tin

Trong đời sống con người, nhu cầu thông tin là một nhu cầu rất cơ bản. Nhu cầu đó không ngừng tăng lên cùng với sự gia tăng các mối quan hệ trong xã hội. Mỗi người sử dụng thông tin lại tạo ra thông tin mới. Các thông tin đó lại được truyền cho người khác trong quá trình thảo luận, truyền đạt mệnh lệnh, trong thư từ và tài liệu hoặc qua các phương tiện truyền thông khác. Thông tin được tổ chức tuân theo một số quan hệ logic nhất định, trở thành một bộ phận của tri thức, đòi hỏi phải được khai thác và nghiên cứu một cách hệ thống.

Sự quan tâm đến hiện tượng thông tin gia tăng đột biến vào thế kỷ XX và ngày nay chúng trở thành đối tượng nghiên cứu của nhiều ngành chuyên môn, trong đó có triết học, vật lý, sinh

học, ngôn ngữ học, thông tin học và tin học, kỹ thuật điện tử và truyền thông, khoa học quản lý và nhiều ngành khoa học xã hội. Về phương diện thương mại, công nghiệp dịch vụ thông tin đã trở thành một nền công nghiệp mới mẻ mang tính toàn cầu. Ngày nay hầu như không một ngành công nghiệp, sản xuất và dịch vụ nào lại không quan tâm đến thông tin. Những quan điểm và hiện tượng khác nhau của lĩnh vực này đã dẫn đến những khái niệm và định nghĩa khác nhau về thông tin.

Giá trị của thông tin

Trong một nghiên cứu mới đây, người ta thấy có bốn yếu tố tác động đến chất lượng thông tin và đem lại giá trị cho nó. Đó là: tính chính xác, phạm vi bao quát của nội dung, tính cập nhật và tần số sử dụng. Trong đó quan trọng nhất là nội dung, thứ đến là tính chính xác.

Trên bình diện tổng quát, ta thấy rằng thông tin có giá trị là những thông tin có tính chất riêng biệt và thông tin có tính chất dự báo. Tính chất riêng biệt làm cho thông tin phù hợp với yêu cầu của người sử dụng. Còn tính chất dự báo cho phép người ta có thể lựa chọn một quyết định trong nhiều khả năng cho phép. Có thể nói thêm rằng giá trị nhận thức của thông tin dự báo liên quan mật thiết đến tính đúng đắn của việc lựa chọn quyết định.

Trên một ý nghĩa rộng hơn, giá trị của thông tin nằm trong quyền lực tổ chức của nó. Thông tin phản ánh cái xác định, trật tự trong các mối quan hệ của tổ chức. Thông tin có giá trị cao cho phép người ta có thể làm môi trường tốt lên và có thể ra những quyết định đối phó được với sự thay đổi của hoàn cảnh. Tính chất quyền lực này của thông tin còn nằm trong cách nó có thể tượng trưng cho những kiến trúc vật chất và tinh thần và được phản ánh trong các định nghĩa của từ điển về động từ

"thông tin": "Thông tin là sắp xếp, hình thành, tạo thành (trí tuệ và tính cách...) bằng cách truyền đạt kiến thức" (*Oxford English Dictionary*).

Sự vật luôn vận động, ở trạng thái bất định và chứa đựng tính ngẫu nhiên. Tăng lượng tin tức về một hiện tượng nào đó cũng là giảm độ chưa biết hoặc độ bất định của nó. Vì vậy trên quan điểm của *lý thuyết thông tin* thì thông tin là sự loại trừ tính bất định của hiện tượng ngẫu nhiên.

Tính trật tự đối lập với cái bất định và ngẫu nhiên là thuộc tính cơ bản của thông tin mà Lý thuyết thông tin của Claude E. Shannon phát hiện. Với ý nghĩa đó thông tin là lượng đo trật tự nhân tạo chống lại sự hỗn độn của tự nhiên. Chính điều đó giải thích ý nghĩa to lớn của thông tin trong mọi lĩnh vực hoạt động của con người.

Vai trò của thông tin

Hiện nay người ta thừa nhận rằng vật chất, năng lượng, thông tin và bản sắc văn hoá dân tộc là các nhân tố quyết định sự tồn tại và phát triển của mỗi quốc gia. Đặc biệt trong điều kiện cách mạng và công nghệ đang diễn ra với quy mô lớn như hiện nay, khoa học và công nghệ đang trở thành lực lượng sản xuất trực tiếp của xã hội thì thông tin khoa học và công nghệ thật sự trở thành nguồn lực quan trọng tạo nên những ưu thế kinh tế và chính trị của mỗi nước.

Nếu như trước đây các nền kinh tế đều dựa chủ yếu vào các nguồn tài nguyên thiên nhiên, lấy việc khai thác và chế biến các tài nguyên thiên nhiên là nguồn chủ yếu tạo ra các của cải và sự giàu có cho xã hội, thì từ những năm cuối của thế kỷ XX, thông tin đã được xem là một nguồn tài nguyên kinh tế, giống như các tài nguyên khác như vật chất, lao động, tiền vốn... Bởi vì việc sở hữu, sử dụng và khai thác thông tin có thể đem lại

hiệu quả kinh tế cho nhiều quá trình vật lý và nhận thức. Ngày nay các hoạt động liên quan đến xử lý thông tin trong công nghiệp chế tạo cũng như trong việc giải quyết các vấn đề xã hội và con người tăng lên đáng kể. Khác với các nguồn tài nguyên khác, tài nguyên thông tin có thể mở rộng phát triển không ngừng và hầu như chỉ bị hạn chế bởi thời gian và khả năng nhận thức của con người. Khả năng mở rộng này thể hiện ở các *thuộc tính* sau đây:

1. Thông tin lan truyền một cách tự nhiên;
2. Khi sử dụng thông tin không bao giờ bị cạn đi, mà trái lại càng trở nên phong phú do được tái tạo và bổ sung thêm các nguồn thông tin mới;
3. Thông tin có thể chia sẻ, nhưng không mất đi trong giao dịch.

Với khả năng thay thế các nguồn tài nguyên khác, khả năng truyền với tốc độ rất cao và khả năng đem lại ưu thế cho người nắm giữ nó, thông tin đã thực sự trở thành cơ sở cho nhiều hoạt động xã hội như nghiên cứu, giáo dục, xuất bản, tiếp thị và cả hoạt động chính trị nữa. Mối quan tâm của xã hội đối với việc sử dụng có hiệu quả các nguồn tài nguyên thông tin đã mở rộng từ các lĩnh vực truyền thống như thư viện, lưu trữ sang các tổ chức, các cơ quan, các ngành. Bây giờ ở đâu người ta cũng quan tâm đến việc quản lý và khai thác các nguồn tài nguyên thông tin.

Một khía cạnh nhận thức thứ hai về vai trò của thông tin trong những năm gần đây là ngày nay ở nhiều nước thông tin đã trở thành hàng hoá. Điều đó đã thúc đẩy hình thành một bộ phận mới trong nền kinh tế quốc dân, đó là khu vực dịch vụ thông tin. Khu vực này ngày càng tạo ra nhiều sản phẩm và dịch vụ thông tin đa dạng và đóng góp một tỷ trọng đáng kể

trong nền kinh tế quốc dân. Người ta thấy rằng khối lượng và chất lượng của các sản phẩm và dịch vụ thông tin đã trở thành tiêu chí đánh giá trình độ phát triển kinh tế của mỗi quốc gia.

Các yếu tố cơ bản trong xử lý thông tin

Con người nhận thông tin thông qua các giác quan: âm thanh qua thính giác; hình ảnh và văn bản qua thị giác; trạng thái, nhiệt độ, cảm xúc qua xúc giác; mùi vị qua khứu giác. Để giải thích và hiểu được các tín hiệu nhận được từ các giác quan, con người phải phát triển và học các hệ thống ngôn ngữ phức hợp, nó bao gồm một "bộ chữ cái" các tín hiệu và các tác nhân kích thích cùng với các quy tắc sử dụng chúng. Điều đó cho phép người ta nhận ra các đối tượng mà họ nhìn thấy, hiểu được các thông báo mà họ đọc hoặc nghe, cảm nhận được các tín hiệu nhận được qua xúc giác và khứu giác.

Các vật mang thông tin chuyển tải tín hiệu tới người nhận có thể là sóng điện từ, sóng ánh sáng, sóng âm và các tác nhân kích thích hoá và điện hoá. Cho đến trước khi máy tính điện tử ra đời, các tín hiệu truyền đi thông qua các vật mang tin trên là những tín hiệu được lưu trữ và xử lý dưới dạng tương đồng, dựa trên công nghệ in, chụp ảnh và điện thoại. Với công nghệ thông tin hiện đại, thông tin được biểu diễn dưới dạng các tín hiệu số nhị phân, dựa trên kỹ thuật số. Đó có thể coi là bước chuyển biến mang ý nghĩa lịch sử vào cuối thế kỷ XX trong cách thức mà con người sáng tạo, tiếp cận và sử dụng thông tin.

Thao tác thông tin

Trong quá trình nghiên cứu một đối tượng ta cần phải tìm kiếm, thu thập các thông tin về đối tượng ấy, chọn lọc rút ra những thông tin cần thiết và khi cần thì đem ra sử dụng.

Các thao tác thông tin cơ bản là: thu thập, xử lý, lưu trữ, tìm kiếm và phổ biến thông tin.

Để thao tác thông tin, thông tin phải được biểu diễn dưới một dạng vật chất cụ thể. Đó là các ký hiệu. Các ký hiệu đó cụ thể là:

- Các chữ số: 0,1,2,3,4,5,6,7,8.
- Các chữ cái: A, B, C,... a, b, c,...
- Các ký hiệu đồ họa: /, #, [,], ∇, ...
- Các ký hiệu toán học: α , β , γ , Σ , $\subset$, $\in$, $\geq$, $\neq$, $+$, $\dots$, x , $\div$, $\cap$, $\cup$, ...

Đo thông tin

Khi thông tin được biểu diễn thông qua các bộ ký hiệu (các chữ cái, các chữ số ...), ta nói thông tin được mã hoá. Thông tin được lưu trữ hoặc trao đổi thông qua mã của nó. Việc khôi phục thông tin từ thông tin đã mã hóa trở lại thông tin ban đầu gọi là giải mã.

Trong xử lý thông tin tự động, dạng mã quan trọng được dùng là dạng mã nhị phân. Thông tin được mã hóa trên bảng chữ cái chỉ gồm hai ký hiệu là chữ số 0 và chữ số 1.

Khả năng hai chữ số đó được sử dụng như nhau. Lượng thông tin vừa đủ để nhận biết một trong hai trạng thái có khả năng xuất hiện như nhau gọi là một bit. Đó là đơn vị dùng để đo thông tin. Từ bit là từ viết tắt của Binary Digit (chữ số nhị phân).

Trong tin học ta thường dùng một số đơn vị bội của bit sau đây:

TÊN GỌI	VIẾT TẮT	GIÁ TRỊ
Byte	B	8 bit
Kilobyte	KB	$1024 = 2^{10}B$
Megabyte	MB	$1024 = 2^{10}MB$
Gigabyte	GB	$1024 = 2^{10}KB$

1.2. Dữ liệu

Dữ liệu (data) là một khái niệm cơ bản của tin học. Sự biểu diễn thông tin bằng một tập hợp các ký hiệu có thể thao tác được trên máy tính điện tử gọi là dữ liệu.

Như vậy, dữ liệu nhấn mạnh vào các ký hiệu, còn thông tin nhấn mạnh vào sự việc và ý tưởng được biểu diễn bằng các ký hiệu ấy.

Trong hoạt động thông tin dữ liệu còn là các thông tin thô ban đầu chưa qua xử lý.

Khái niệm dữ liệu trong các hệ thống lưu trữ và xử lý thông tin có các đặc trưng cơ bản sau:

- Là một thông tin: phản ánh tự nhiên và xã hội.
- Có cấu trúc: được biểu diễn hình thức hoá và được cố định bằng một số ký hiệu.
- Có thể biến đổi: được xử lý, lưu trữ và lưu truyền trong các hệ thống và mạng lưới thông tin.

Có thể nói dữ liệu là “món ăn” để máy tính điện tử tiêu thụ và cũng là kết quả sinh ra từ quá trình xử lý đó. Nói cách khác thì đầu vào, đầu ra, cũng như các thông tin vận động trong máy tính điện tử đều là dữ liệu.

Ví dụ: Ta hãy xem thông tin nào được biểu diễn bởi các dữ liệu sau:

104 Vũ Hoài Nam 1973 165 19/20 Thái Bình

Để hiểu được thông tin biểu diễn bởi các dữ liệu trên ta phải biết được cấu trúc của dữ liệu. Cấu trúc này cho ta hình dung được thông tin được biểu diễn bởi dãy dữ liệu trên song chưa cho ý nghĩa cụ thể của từng dữ liệu.

Cấu trúc của dữ liệu trên là:

MÃ SỐ	HỌ VÀ TÊN	NĂM SINH	CHIỀU CAO	THỊ LỰC	QUÊ QUÁN
(999)		(9999)	(cm)	(K/20)	

Với cấu trúc trên ý nghĩa của dữ liệu trên là:

MÃ SỐ	HỌ VÀ TÊN	NĂM SINH	CHIỀU CAO	THỊ LỰC	QUÊ QUÁN
104	Vũ Hoài Nam	1973	165	19/20	Thái Bình

Kiểu dữ liệu

Khái niệm kiểu dữ liệu liên quan tới chính dữ liệu, nó biểu diễn một tập hợp các giá trị mà dữ liệu có thể nhận.

Trong ngôn ngữ lập trình PASCAL dữ liệu có thể nhận các kiểu chuẩn sau:

- Kiểu nguyên (Integer)
- Kiểu thực (Real)
- Kiểu ký tự (Char)
- Kiểu logic (Boolean)
- Kiểu chuỗi ký tự (String).

2. CẤU TRÚC DỮ LIỆU

Người ta phân biệt hai loại dữ liệu:

- *Dữ liệu sơ cấp* là dữ liệu có thể thao tác như một thực thể không thể chia nhỏ được.

Ví dụ: Một chữ số, một ký tự,...

- *Dữ liệu có cấu trúc* là một tập hợp các dữ liệu sơ cấp mà ở đó người ta có thể tổ chức, sắp xếp và truy nhập được.

Ví dụ: Các mảng, các biểu ghi, các danh sách, các tệp,...

Dưới đây ta điểm qua một số cấu trúc dữ liệu cơ bản.

2.1. Mảng

Mảng (array) là một danh sách các giá trị có cùng một kiểu và được đánh dấu bằng một chỉ số.

Có mảng một chiều và mảng hai chiều.

Mảng một chiều, còn gọi là véc tơ.

Ví dụ: mảng $A(i)$, với $i = 1, \dots, 8$

5	7,5	4	9	6,5	6	8,5	3
---	-----	---	---	-----	---	-----	---

$A(5)$ là phần tử thứ 5 của mảng A :

$$A(5) = 6,5$$

Mảng hai chiều, còn gọi là ma trận.

Ví dụ: Mảng $B(i,j)$, với $i = 1, \dots, 8, j = 1, \dots, 4$

14	5	18	58	47	30	13	5
10	8	20	45	83	11	58	31
4	15	3	27	19	12	0	13
89	74	9	1	25	76	82	16

$B(3,5)$ là phần tử thuộc dòng 3, cột 5:

$$B(3,5) = 19$$

2.2. Danh sách

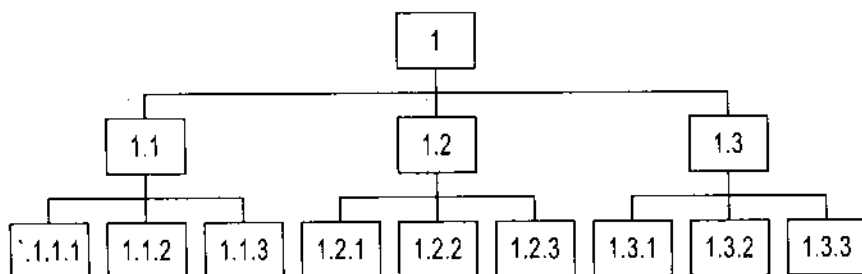
Danh sách (list) là một tập hợp sắp thứ tự của dữ liệu.

Ví dụ: Danh sách các công trình khoa học của một viện nghiên cứu.

2.3. Cây

Cây (tree) là một đồ thị có ít nhất 2 nút và không có chu trình.

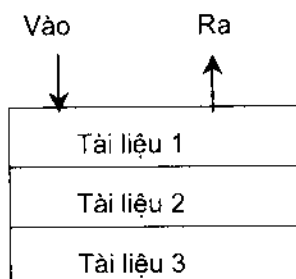
Ví dụ: Cấu trúc của các khung phân loại là cấu trúc cây:



Hình 5. Cấu trúc cây của khung phân loại

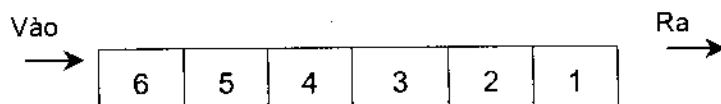
2.4. Sắp

Sắp (pile) là tập hợp các dữ liệu có cấu trúc tuyến tính, sao cho ở đó người ta chỉ có thể thêm vào hoặc bớt đi ở một đầu của sắp dữ liệu.



2.5. Hàng

Hàng (queue) là tập hợp các dữ liệu có cấu trúc tuyến tính, sao cho ở đó các dữ liệu được thêm vào ở một đầu và lấy ra ở đầu kia.



2.6. Biểu ghi

Biểu ghi (record) là một dữ liệu có cấu trúc được tạo thành bởi một số thành phần gọi là trường (Field).

Ví dụ: Biểu ghi về bản tra cứu thư mục của một cuốn sách có thể bao gồm các trường sau:

- Tác giả
- Tên sách
- Nơi xuất bản

- Nhà xuất bản
- Năm xuất bản
- Số trang
- Khổ, cỡ
- Ký hiệu phân loại
- Ký hiệu kho
- Tóm tắt
- Từ khoá.

Cùng với giá trị của các trường đó, biểu ghi sẽ cho ta thông tin thư mục về một cuốn sách cụ thể.

Trường

Trường (field) là thành phần của biểu ghi. Đó là một dãy các ký tự nối tiếp nhau để mô tả một tiêu thức hay một thuộc tính của đối tượng. Trường có thể coi là đơn vị dữ liệu được lưu trữ, nó cung cấp thông tin liên quan tới một khía cạnh hay thuộc tính của đối tượng mà biểu ghi mô tả.

Trên quan điểm sắp xếp người ta phân biệt có trường khoá chủ và trường khoá thứ.

Trường khoá chủ là trường quan trọng có ý nghĩa xác định đối tượng còn trường khoá thứ là trường phản ánh các thuộc tính đơn lẻ của đối tượng.

Ví dụ: Đối tượng là nhân viên thì:

Trường khoá chủ là: họ và tên hoặc số chứng minh.

Trường khoá thứ là: giới tính, năm sinh, học vấn,...

2.7. Tập

Tập (file) là một tập hợp dữ liệu gồm nhiều biểu ghi có chung các đặc tính cấu trúc và ngữ nghĩa, dùng để mô tả một lớp đối tượng.

Ví dụ: Tập hợp các biểu ghi thư mục tạo thành tập dữ liệu thư mục. Đó là thành phần cơ bản của một cơ sở dữ liệu thư mục.

Có nhiều cách tổ chức các tập dữ liệu, tuy nhiên chúng phải được tổ chức sao cho dễ truy nhập tới các phần tử dữ liệu.

Theo cách tổ chức, người ta phân ra các loại tập sau:

Tập kế tiếp

Tập kế tiếp (sequential file) là tập được trình bày thành một dãy các biểu ghi liên tiếp. Để đọc một biểu ghi, ta phải đọc lần lượt các biểu ghi trước nó.

Tập truy nhập trực tiếp

Tập truy nhập trực tiếp (direct access file) có thể coi là một bộ sưu tập các biểu ghi có đánh số thứ tự. Nó có các tính chất sau:

- Việc tra cứu một biểu ghi được tiến hành bằng cách chỉ rõ số thứ tự của nó.
- Người ta có thể thay thế, loại bỏ hoặc bổ sung dễ dàng một biểu ghi.

Tập đảo

Một cơ sở dữ liệu có cấu trúc tập đảo thường bao gồm:

- Các tập sơ cấp
- Mỗi tập sơ cấp lại có nhiều tập đảo kết hợp với nó.

Tập sơ cấp bao gồm các tình huống của thực thể. Một thực thể được tạo thành bởi các thuộc tính của nó. Tập đảo là tập mà biểu ghi của nó là các giá trị của một thuộc tính nào đó, nó xác định tình huống của thực thể được tạo thành ở tập sơ cấp. Các giá trị đó được xem là các điểm truy nhập thông tin. Đó có thể coi như là một tập hợp các bảng đảo của các thuộc tính trong các biểu ghi mà ta cần tìm.

Như vậy tổ chức của tập đảo xuất phát từ các giá trị riêng của các thuộc tính. Nó cho phép ta tiếp cận các biểu ghi của tập sơ cấp mà thuộc tính đang xét có giá trị riêng đó.

Người ta có thể tạo ra nhiều tập đảo. Tuy nhiên số tập đảo không thể vượt quá số trường có trong biểu ghi của tập sơ cấp.

Để tổ chức tập đảo thì ở tập sơ cấp mỗi biểu ghi phải được định vị bằng một “địa chỉ”. Nhờ các “địa chỉ” này mà xuất phát từ một giá trị của tập đảo ta có thể xác định được biểu ghi chứa các thông tin mà ta cần tìm. “Địa chỉ” này cũng giống như số trang trong mục lục hay trong các bảng chỉ mục (index) của một cuốn sách.

Ví dụ: Tập của một cơ sở dữ liệu quản lý bạn đọc ở thư viện được tạo thành bởi các thuộc tính sau:

ST:	Số thẻ
HT:	Họ và tên
NS:	Năm sinh
NN:	Nghề nghiệp
TĐ:	Trình độ văn hoá
CQ:	Cơ quan công tác

Giả sử dưới đây là một thể hiện của tệp đó:

Tệp sơ cấp BẠNĐOC

ST	HT	NS	NN	TĐ	CQ
001	Đỗ Cao Thắng	1965	Bác sĩ	Đại học	Bộ Năng lượng
002	Nguyễn Ngọc Hà	1960	Giáo viên	Thạc sĩ	ĐH Văn hoá
003	Lê Thu Trang	1964	Kỹ sư	Đại học	Bộ Nông nghiệp
004	Vũ Quế Anh	1962	Bác sĩ	Tiến sĩ	Viện Đông Y
005	Trần Văn Thành	1967	Kỹ sư	Đại học	Bộ Xây dựng
006	Hoàng Quốc Long	1958	Giáo viên	Tiến sĩ	ĐH Văn hoá
007	Nguyễn Duy Hiến	1961	Công nhân	Trung cấp	Dệt 8.3
008	Trần Hải Châu	1966	Giáo viên	Đại học	ĐH Sư phạm
009	Vũ Thị Uyên	1950	Kỹ sư	Thạc sĩ	Bộ Xây dựng
010	Chu Phương Diệp	1963	Công nhân	Trung cấp	Dệt 8.3

Từ tệp sơ cấp trên ta có thể tạo ra các tệp đảo sau:

- Một tệp đảo theo thuộc tính NN
- Một tệp đảo theo thuộc tính TĐ
- Một tệp đảo theo thuộc tính CQ

Sự tương ứng giữa tệp sơ cấp và tệp đảo thường được bảo đảm nhờ một thuộc tính khoá. Đó là thuộc tính xác định thực thể. Trong ví dụ đang xét thuộc tính khoá là ST (số thẻ).

Từ bảng trên ta thấy giá trị “Giáo viên” của thuộc tính NN ứng với các ST: 002, 006, 008. Do đó cấu trúc logic của biểu ghi ứng với giá trị đó trong tệp đảo là:

Giáo viên 002, 006, 008

Do đó cấu trúc logic của các tệp đảo được xác lập như sau:

Tệp đảo NN

NN	ST		
Giáo viên	002	006	008
Bác sĩ	001	004	
Kỹ sư	003	005	009
Công nhân	007	010	

Tệp đảo TĐ

TĐ	ST			
Tiến sĩ	004	006		
Thạc sĩ	002	009		
Đại học	001	003	005	008
Trung cấp	007	010		

Tệp đảo CQ

CQ	ST	
Bộ Năng lượng	001	
Bộ Nông nghiệp	003	
Bộ Xây dựng	005	009
Viện Đông Y	004	
ĐH Văn hoá	002	006
ĐH Sư phạm	008	
Dệt 8.3	007	010

Các cơ sở dữ liệu được xây dựng với cấu trúc tệp đảo thường sử dụng ngôn ngữ tìm dựa trên cơ sở của logic Boole. Một thuật ngữ tìm có liên quan đến biểu ghi, thực chất có thể xem như đại diện cho một lớp các biểu ghi có liên quan đến thuật ngữ đó. Do đó bằng cách sử dụng các phép toán logic liên kết các thuật ngữ với nhau, ta có thể xác định một cách chính xác lớp các biểu ghi cần tìm trả lời cho một câu hỏi.

Một câu hỏi có thể là: “Tìm bạn đọc là giáo viên có trình độ Thạc sĩ, công tác tại Trường ĐH Văn hoá Hà Nội”.

Bằng phương pháp biểu diễn của lý thuyết tập hợp ta có:

$$\text{Giáo viên} = \{ 002, 006, 008 \}$$

$$\text{Thạc sĩ} = \{ 002, 009 \}$$

$$\text{ĐH Văn hoá} = \{ 002, 006 \}$$

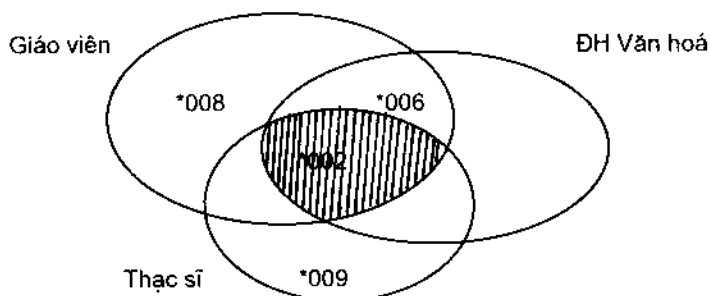
Câu trả lời của câu hỏi trên là:

$$\text{Giáo viên} \cap \text{Thạc sĩ} \cap \text{ĐH Văn hoá} = \{ 002 \}$$

Đó là biểu ghi:

002	Nguyễn Ngọc Hà	1960	Giáo viên	Thạc sĩ	ĐH Văn hoá
-----	----------------	------	-----------	---------	------------

Kết quả trên có thể biểu diễn bằng sơ đồ Venn:



2.8. Cơ sở dữ liệu

Cơ sở dữ liệu (Data base) là cấp tổ chức cao nhất của dữ liệu. Đó là tập hợp một số tệp quan hệ với nhau tới thực thể, thống nhất với nhau về mặt cấu trúc, được quản lý theo một cơ chế thống nhất, giúp cho việc xử lý và truy nhập dữ liệu được dễ dàng và nhanh chóng.

Ta sẽ trình bày chi tiết khái niệm này ở chương sau.

Chương 3

CƠ SỞ DỮ LIỆU VÀ HỆ QUẢN TRỊ CƠ SỞ DỮ LIỆU

1. MỞ ĐẦU

Ngày nay không ai còn nghi ngờ gì về vai trò của thông tin trong khoa học kỹ thuật và đời sống cũng như trong mọi vận động của xã hội dưới mọi quy mô, từ các đơn vị xí nghiệp cho đến quốc gia và cả quốc tế.

Việc nắm bắt thông tin nhanh chóng, đầy đủ, chính xác và kịp thời ngày càng đóng vai trò cốt tử trong quản lý và điều hành. Nói cách khác ngày nay nói quản lý thực chất là quản lý thông tin. Song mọi thông tin cần quản lý trên máy tính theo bất cứ quy trình cụ thể nào cũng đều được thể hiện bằng các dữ liệu ghi trên một vật mang tin nào đó. Bởi vậy khi nói đến quản lý thông tin tức là quản lý dữ liệu.

Cơ sở dữ liệu là một lĩnh vực của tin học nghiên cứu các cơ chế, nguyên lý và phương pháp tổ chức các nhóm dữ liệu nhằm phục vụ cho việc khai thác dữ liệu trong các hệ thống tin học ứng dụng cũng như trong các hệ thống lưu trữ và tìm kiếm thông tin ở các trung tâm thông tin - thư viện hoặc trong các hệ thống quản lý cơ quan xí nghiệp.

Những CSDL đầu tiên xuất hiện vào cuối những năm 60 và nhanh chóng trở thành một nguồn thông tin mới bên cạnh các kho lưu trữ thông tin truyền thống của các thư viện và cơ quan

lưu trữ. Sự ra đời và phát triển của các CSDL tạo nên một biến chuyển về chất trong hoạt động của các cơ quan thông tin - thư viện. Đánh giá về sự kiện này, tại hội thảo lần thứ 41 của Liên đoàn Tư liệu Quốc tế (FID) về chủ đề “Thách thức của công nghệ thông tin (CNTT) đối với hoạt động thông tin - thư viện”, tổ chức tại Hồng Kông năm 1982, J.Martin - một chuyên gia thông tin học người Mỹ đã nói: “Những nhà sử học sẽ nhìn nhận sự xuất hiện của các CSDL, cũng như khả năng phát triển của chúng là một bước biến đổi về chất của sự tiến hoá xã hội, và có lẽ, nó có tầm quan trọng lớn hơn so với sự tiến bộ của báo in”.

Ngày nay CSDL là bộ phận không thể thiếu được trong các hệ thống thông tin - thư viện tự động hoá.

2. TỪ QUẢN LÝ TẬP ĐẾN HỆ THỐNG QUẢN TRỊ DỮ LIỆU

Nhu cầu tích lũy và xử lý dữ liệu là nhu cầu cơ bản nảy sinh từ hoạt động thực tiễn của con người. Tự động hoá xử lý dữ liệu là một đòi hỏi tất yếu khi cần xử lý nhanh một khối lượng lớn dữ liệu.

Khởi đầu công tác tự động hoá xử lý dữ liệu, người ta lưu trữ thông tin trên các tập dữ liệu và dùng các chương trình để tìm kiếm, thao tác trên các tập dữ liệu đó. Đó là tiền thân của các hệ thống cơ sở dữ liệu.

Việc quản trị dữ liệu trên máy tính được thực hiện nhờ hai dạng chương trình sau:

- Các hệ thống quản lý tập.
- Các hệ thống quản trị cơ sở dữ liệu.

Hệ thống quản lý tập là chương trình lưu trữ, xử lý và in dữ liệu chứa trong các tập tách biệt. Hệ quản trị tập lưu trữ dữ liệu

trên bộ nhớ ngoài (băng từ, đĩa từ, đĩa quang) và khai thác dữ liệu bằng các chương trình. Chương trình là tập hợp các lệnh được viết bằng ngôn ngữ bậc cao như ngôn ngữ PASCAL, C... cho phép mô tả và xử lý dữ liệu. Các chương trình cần thoả mãn yêu cầu phục vụ nhiều tệp dữ liệu khác nhau. Chương trình đó bảo đảm tính độc lập so với dữ liệu, tức là có thể thay đổi dữ liệu của bộ nhớ ngoài mà không cần thay đổi chương trình. Một khi tệp được phân chia thành các biểu ghi thì tệp được xem là tập hợp các biểu ghi. Tuy nhiên không phải tệp nào cũng được sắp đặt như vậy. Tại một thời điểm chỉ có một tệp được truy nhập và xử lý.

Chức năng cơ bản của hệ quản lý tệp là xử lý các tệp dữ liệu, bao gồm:

- Tạo tệp mới.
- Mở, đóng các tệp.
- Xoá tệp.

Hệ thống quản trị tệp không tự trang bị mà nhờ hệ thống điều hành máy tính thực hiện các chức năng trên. Nó quan tâm nhiều đến các phương pháp tổ chức dữ liệu trên bộ nhớ ngoài. Có hai phương pháp chính tổ chức dữ liệu trên tệp là tổ chức tuần tự và tổ chức khai thác trực tiếp.

Sử dụng hệ thống quản lý tệp có nhiều hạn chế. Thứ nhất nó tốn thời gian khi phải khai thác các tệp chứa các thông tin khác. Thứ hai là nhiều khi phải lưu trữ lại một số dữ liệu, ví dụ như tên và mã số của các nhân viên. Thứ ba là độ tin cậy của dữ liệu không cao, ví dụ một thông tin phải lưu trữ ở nhiều nơi, do sơ suất không cập nhật được ở tất cả các nơi khác nhau, tạo nên sự không nhất quán về thông tin.

Khái niệm cơ sở dữ liệu và hệ quản trị cơ sở dữ liệu được hình thành để giải quyết những hạn chế liên quan tới hệ thống quản lý tệp.

Từ quản lý tệp đến quản lý dữ liệu bằng các hệ quản trị cơ sở dữ liệu là một bước tiến quan trọng của tin học, góp phần mở rộng các ứng dụng của tin học trong công tác quản lý và điều hành.

3. CƠ SỞ DỮ LIỆU

3.1. Cơ sở dữ liệu và hệ quản trị cơ sở dữ liệu

Cơ sở dữ liệu là tập hợp các dữ liệu về các đối tượng cần được quản lý, được lưu trữ đồng thời trên các vật mang tin của máy tính điện tử và được quản lý theo một cơ chế thống nhất, nhằm giúp cho việc truy nhập và xử lý dữ liệu được dễ dàng và nhanh chóng.

CSDL bao gồm một tệp hoặc một tập hợp các tệp dữ liệu. Thông tin trong các tệp này có thể chia nhỏ thành các biểu ghi, mỗi biểu ghi lại bao gồm một hoặc nhiều trường. Trường là đơn vị cơ sở của dữ liệu và mỗi trường thường chứa các thông tin liên quan một khía cạnh hay một thuộc tính của thực thể được mô tả bởi CSDL. Sử dụng các từ khoá hoặc các lệnh tìm, người dùng tin có thể nhanh chóng lựa chọn ra các biểu ghi thoả mãn yêu cầu tìm tin đặt ra.

CSDL được quản lý bởi một hệ quản trị CSDL. Đó là một hệ thống phần mềm, bao gồm các chương trình giúp người sử dụng quản lý và khai thác CSDL theo ba chức năng sau đây:

- Mô tả dữ liệu
- Cập nhật dữ liệu

- Tìm kiếm dữ liệu

Mô tả dữ liệu được xem như việc xác định tên, dạng và tính chất của dữ liệu. Việc mô tả dữ liệu tập trung vào các thuộc tính của đối tượng được mô hình hoá trong CSDL. Khác với hệ quản lý tệp, hệ quản trị CSDL cho phép mô tả dữ liệu mà không phụ thuộc vào việc xử lý dữ liệu trong các ứng dụng thực tế. Việc tách riêng chức năng mô tả dữ liệu ra khỏi các chức năng khác trong chương trình ứng dụng cho phép quản lý các dữ liệu một cách thống nhất.

Trong CSDL, dữ liệu có thể thay đổi theo yêu cầu của người sử dụng. Mỗi hệ quản trị CSDL đều cung cấp những phương tiện để cập nhật dữ liệu, tức là có thể thêm, bớt, sửa đổi dữ liệu.

Một khi CSDL đã được thiết lập, người sử dụng có thể tìm kiếm dữ liệu theo yêu cầu. Để diễn tả yêu cầu tìm kiếm thông tin, người ta dùng ngôn ngữ hỏi của hệ quản trị CSDL. Người sử dụng phải biết đặt điều kiện tìm, tức là xác định phạm vi mà các dữ liệu phải thoả mãn.

Hệ quản trị CSDL là một công cụ để quản lý các CSDL. Mục tiêu của hệ quản trị CSDL là:

- Thực hiện tích nhập dữ liệu sao cho người sử dụng có thể tiếp cận được dễ dàng.
- Bảo đảm chất lượng và tính vẹn toàn dữ liệu.
- Bảo đảm tính độc lập logic và vật lý của dữ liệu.
- Cho phép kiểm soát toàn bộ CSDL.

Tính độc lập dữ liệu là mục tiêu chủ yếu của các hệ quản trị CSDL. Có thể định nghĩa tính độc lập dữ liệu là “ tính bất biến

của các hệ ứng dụng đối với các thay đổi trong cấu trúc lưu trữ và chiến lược truy nhập dữ liệu” (Date).

Ngày nay trên thị trường có hàng trăm phần mềm quản lý CSDL. Trong các phần mềm quản lý CSDL văn bản ta thấy các hệ quản trị CSDL với cấu trúc tệp đảo có nhiều ưu điểm nổi bật và được sử dụng nhiều hơn cả vì nó cho khả năng truy nhập dữ liệu một cách mềm dẻo và linh hoạt. CDS/ISIS chính là hệ quản trị CSDL có cấu trúc tệp đảo, đang được sử dụng rộng rãi trong các cơ quan thông tin và thư viện ở nước ta.

Các hệ quản trị CSDL đều cung cấp một ngôn ngữ hình thức để trợ giúp người sử dụng và làm công cụ giao tiếp giữa người và máy. Đó là ngôn ngữ dữ liệu, nó cho phép người sử dụng xây dựng cấu trúc của CSDL, mô tả, cập nhật và tìm kiếm dữ liệu.

3.2. Phân loại CSDL

Có nhiều cách phân loại CSDL. Trong lĩnh vực thông tin - thư viện, căn cứ theo bản chất của thông tin được lưu giữ, người ta phân biệt ba loại CSDL chính sau:

CSDL thư mục

Dữ liệu trong CSDL thư mục là tin tức về bản thân tài liệu. Chúng chứa các thông tin cấp hai, tức là các dữ liệu thư mục và các dữ liệu bổ sung, chứ không phải là văn liệu gốc. CSDL thư mục được trình bày tương tự như trong ấn phẩm thư mục hay trong tạp chí tóm tắt. Nó bao gồm các dữ liệu thư mục (tác giả, nhan đề, các yếu tố xuất bản, các đặc trưng số lượng,...), các chỉ số phân loại, tóm tắt, chú giải, từ khoá... Đối tượng xử lý là các tài liệu chuyên khảo, bài trích báo, tạp chí, kỷ yếu hội nghị, hội thảo khoa học, báo cáo khoa học, luận án, sáng chế... Nói tóm lại là mọi đối tượng đã được tư liệu hoá.

CSDL thư mục bao gồm các tra cứu về tài liệu gốc, cùng với các chỉ dẫn giúp người dùng tin có thể nhận được tài liệu gốc ở một thư viện.

Các CSDL thư mục đưa lại nhiều lợi ích. Trước hết nó cho phép người sử dụng truy nhập trực tiếp và tức thì các thông tin thư mục trong kho dữ liệu và thông tin cần tìm hiển thị ngay trên màn hình. Mặt khác nó cho phép tạo ra nhiều sản phẩm trung gian, như là những sản phẩm đầu ra của CSDL, đó là các ấn phẩm thư mục, các tạp chí tóm tắt, các bộ phiếu mục lục hay các tài liệu vi dạng (microfilm, microfiche).

CSDL dữ kiện

Khác với CSDL thư mục, CSDL dữ kiện chứa các thông tin cấp một. Nội dung thông tin chứa trong CSDL là các thông tin dữ kiện. Đó là các số liệu hay các dữ kiện cụ thể về các đối tượng, các quy trình hoặc phương pháp. Thông tin dữ kiện được lưu trữ chủ yếu dưới dạng số, ngoài ra cũng được trình bày dưới dạng văn bản, hình ảnh, biểu đồ, đồ thị,....

Trong CSDL dữ kiện, thông tin đã được xử lý biên tập và có thể dùng trực tiếp, không cần tới một nguồn dữ liệu bổ sung nào. Lợi ích của nó là người dùng tin không cần mất thời gian tìm đến tài liệu gốc. Điều đó làm cho CSDL dữ kiện trở thành một công cụ hữu ích cho các nhà quản lý, các nhà khoa học và các kỹ sư.

CSDL toàn văn

CSDL toàn văn là sự mở rộng logic tới các hệ thống thông tin hiện đại. Chúng chứa các thông tin cấp một, đó là toàn văn bản của tài liệu cùng với các dữ liệu thư mục và các dữ liệu chủ đề khác. Ngày nay một số tạp chí và dịch vụ báo chí đã được

cung cấp dưới dạng CSDL toàn văn. Đó là trường hợp của các báo điện tử.

CSDL toàn văn có nhược điểm là nó không chứa các bảng biểu và các hình ảnh minh họa. Ngoài ra CSDL toàn văn đòi hỏi phải có bộ nhớ với dung lượng lớn, đầu tư ban đầu tốn kém. Tuy nhiên ngày nay, với sự phát triển của công nghệ thông tin, vấn đề này đã và đang được giải quyết.

Các CSDL toàn văn đã cải thiện đáng kể việc cung cấp các nguồn thông tin cấp một, mở ra triển vọng mới trong hoạt động của các hệ thống giao lưu thông tin.

Theo phạm vi bao quát của đề tài, các CSDL phân thành: CSDL đa ngành, CSDL đơn ngành, CSDL theo chủ đề. Ví dụ: CSDL mỏ và địa chất, CSDL về môi trường, CSDL về dân số, v.v...

Các CSDL cũng có thể phân theo loại hình tài liệu. Những CSDL này thường hướng vào một loại hình tài liệu cụ thể và thường có nội dung đa ngành. Ví dụ: CSDL sách, CSDL tạp chí, CSDL về các phát minh sáng chế, CSDL về các công trình khoa học, v.v...

Các CSDL được lưu trữ trong bộ nhớ của máy tính điện tử, tức là ghi trên các băng từ, đĩa từ. Từ đó CSDL cũng có thể được xuất bản dưới dạng một ấn phẩm, tức là được ghi ra trên giấy. Ngày nay, với sự phát triển của công nghệ thông tin hiện đại, CSDL còn được ghi và lưu trữ trên đĩa quang CD-ROM. Các CSDL toàn văn trên CD-ROM có thể lưu trữ các thông tin dưới dạng văn bản, âm thanh, hình ảnh tĩnh và động và truy nhập rất thuận lợi.

Với những ưu điểm nổi bật như kích thước nhỏ, dung lượng lớn, độ bền cao, có phần mềm khai thác thông tin ghi ngay trên

đĩa... các CD-ROM cơ sở dữ liệu trở thành một sản phẩm thông tin quan trọng, không thể thiếu được trong các dịch vụ phổ biến thông tin hiện nay.

4. NGÔN NGỮ DỮ LIỆU

Ngôn ngữ là phương tiện giao tiếp của con người. Có ngôn ngữ tự nhiên và ngôn ngữ hình thức.

Ngôn ngữ tự nhiên là ngôn ngữ của các bộ tộc người, thể hiện qua tiếng nói, chữ viết hoặc động tác, cử chỉ và tuân theo những quy tắc nhất định.

Ngôn ngữ hình thức là tập hợp các ký hiệu và quy định do con người đặt ra. Trong tin học ngôn ngữ hình thức là công cụ giao tiếp giữa người và máy hoặc giữa máy và máy.

Các hệ quản trị CSDL bao giờ cũng cung cấp một ngôn ngữ hình thức làm công cụ giao tiếp giữa người sử dụng và hệ thống. Đó là ngôn ngữ dữ liệu.

Ngôn ngữ dữ liệu thường bao gồm ba lớp ngôn ngữ con sau đây, ứng với ba chức năng chủ yếu của hệ quản trị CSDL là: mô tả dữ liệu, cập nhật dữ liệu và tìm kiếm dữ liệu.

a. Ngôn ngữ con mô tả dữ liệu (Data Definition Language - DDL).

Ngôn ngữ mô tả dữ liệu giúp cho người sử dụng khai báo tên đối tượng, cấu trúc của đối tượng và các quan hệ của đối tượng này với đối tượng khác.

b. Ngôn ngữ con cập nhật dữ liệu (Data Update Language - DUL).

Ngôn ngữ con cập nhật dữ liệu dùng để thực hiện các thao tác sau:

- Đổi tên, đổi kiểu của dữ liệu đã khai báo.
- Thêm bớt một số thành phần của cấu trúc đã có.
- Sửa, xóa, bổ sung dữ liệu trong CSDL.

c. *Ngôn ngữ con truy nhập dữ liệu hay ngôn ngữ hỏi (Query Language - QL).*

Ngôn ngữ con truy nhập dữ liệu giúp ta tìm kiếm thông tin trong CSDL. Thông thường ngôn ngữ hỏi có ba mức khai báo chủ yếu sau đây:

- *Đường dẫn*: Cho biết địa chỉ của các đối tượng cần xử lý. Có thể hiểu đường dẫn là địa chỉ theo tên. Biết tên thì biết nơi lưu trữ đối tượng.
- *Điều kiện*: Chứa các điều kiện dùng để lựa chọn dữ liệu trong CSDL.
- *Khuôn ra*: Quy định khuôn dạng cho dữ liệu cần đưa ra.

Các dạng thể hiện của ngôn ngữ dữ liệu

Ngôn ngữ là hình thức giao tiếp nên có thể có hai dạng thể hiện chính là chế độ hội thoại hoặc chế độ chương trình.

- *Chế độ hội thoại* là chế độ hỏi - trả lời trong thời gian thực. Một hình thức thể hiện của chế độ hội thoại được triển khai rộng rãi trong những năm gần đây là hội thoại theo thực đơn (menu). Ở đây hệ thống đưa ra các thực đơn bao gồm các “tùy chọn” (option) có sẵn. Người sử dụng tùy theo mục đích của mình sẽ lựa chọn một “tùy chọn” - một chức năng. Nếu chức năng lựa chọn là đơn giản thì hệ thống sẽ thực hiện chức năng đó và thông báo kết quả. Nếu không, tức là chức năng được lựa chọn

lại chứa nhiều khả năng, hệ thống lại phối ra một thực đơn thứ hai để người sử dụng tiếp tục lựa chọn những chức năng mình hơn.

- *Chế độ chương trình*: Người sử dụng viết một dãy lệnh dưới dạng một chương trình, rồi giao chương trình đó cho hệ thống thực hiện.

5. CÁC TIÊU CHUẨN ĐỂ LỰA CHỌN MỘT HỆ QUẢN TRỊ CSDL

Để lựa chọn một hệ quản trị CSDL cần căn cứ vào các đặc trưng chức năng của chúng. Các chức năng đó là:

a. *Tạo lập CSDL*

- Số CSDL: Bao nhiêu CSDL có thể mở cùng một lúc (Ví dụ: với Superdoc là 1, với Dbase là 10).
- Có khả năng thiết lập quan hệ giữa các CSDL khác nhau?
- Kích thước:
 - Giới hạn số các biểu ghi của một CSDL.
 - Độ dài của mỗi đề mục.
- Xác định cấu trúc của CSDL: Được xác định trước hay do người sử dụng thiết lập, giới hạn của số trường.
- Xác định các trường:
 - Độ dài của tên trường.
 - Độ dài của trường.
 - Kiểu trường: Chữ cái, chữ số, hỗn hợp, kiểu logic, kiểu ngày tháng, v.v...
 - Khả năng phân chia trường thành các trường con.

- Khả năng nhận biết trường lặp.
- Thiết lập các tệp đảo:
 - Có một hay nhiều tệp đảo.
 - Có khả năng tự tạo tệp đảo hay phụ thuộc người sử dụng.
 - Các thuật ngữ tìm có thể đưa vào một danh mục hay không?
 - Có thể mở danh mục các thuật ngữ tìm.
 - Trong tệp đảo chữ hoa và chữ thường như nhau hay phải phân biệt khi tìm tin.

b. Nhập dữ liệu

Khả năng nhập dữ liệu thể hiện ở các điểm sau:

- Việc nhập dữ liệu có thể thực hiện theo trình tự các trường hay theo một biểu mẫu nhập tin - Worksheet?
- Người sử dụng có thể thiết lập Worksheet ?
- Việc nhập dữ liệu thường được thực hiện bởi một chương trình soạn thảo văn bản nhúng trong phần mềm quản trị CSDL. Chất lượng của chương trình này làm cho việc nhập dữ liệu được thuận lợi và dễ dàng hơn.
- Các chỉ dẫn soạn thảo giúp cho việc nhập dữ liệu có thường xuyên hiện ra trên màn hình hay phải lựa chọn.
- Khả năng kiểm tra: Kiểm tra về cú pháp, kiểm tra ngữ nghĩa, kiểm tra trùng.

c. Tìm tin

Khả năng tìm thể hiện ở các phương tiện, phương thức tìm và tốc độ tìm.

- Tìm tin theo dòng: Đây không phải cách tìm chuẩn vì thường rất dài.
- Tìm tin trực tiếp trên tệp đảo: Liên kết các từ khoá bằng các toán tử tìm. Các toán tử tìm có thể là:
 - Toán tử Boole: AND, OR, NOT, các dấu ngoặc.
 - Toán tử chặt cụt.
 - Toán tử lân cận.
 - Toán tử so sánh với trường số: <, >, <=, >=
- Tìm tin với sự trợ giúp của danh mục các thuật ngữ tìm kết hợp các toán tử tìm.
- Thời gian tìm.

d. Hiện hình, in và xuất bản

- In: theo dòng hay theo lô.
- Khả năng sắp xếp theo chữ số, chữ cái, tăng hoặc giảm trên một hoặc nhiều trường.
- Khả năng tạo format hiện hình và format in.
- Khả năng đưa dữ liệu ra dưới dạng tệp văn bản? với format ISO 2709?
- Khả năng hiển thị danh mục các thuật ngữ tìm.

e. An toàn dữ liệu

- Mật khẩu : trên 3 mức độ:
 - Tạo lập cấu trúc CSDL.
 - Cập nhật.
 - Tìm tin.
- Khả năng sao dự phòng và phục hồi dữ liệu.

Chương 4

PHẦN MỀM TƯ LIỆU VÀ KHỔ MẪU BIÊN MỤC ĐỌC MÁY MARC

1. PHẦN MỀM TƯ LIỆU VÀ PHẦN MỀM TÍCH HỢP QUẢN TRỊ THƯ VIỆN

Thư viện là một ngành hoạt động có lịch sử lâu đời. Trên thế giới, thư viện là một trong những lĩnh vực được áp dụng sớm nhất những thành tựu của công nghệ thông tin do hoạt động thư viện là hoạt động mang tính đặc thù, rất phù hợp với khả năng lưu trữ thông tin với khối lượng lớn và khả năng tìm kiếm thông tin nhanh của máy tính. Việc ứng dụng tin học trong công tác thông tin thư viện thường trải qua hai giai đoạn kế tiếp nhau.

Mở đầu người ta thường tập trung vào việc lưu trữ, tìm kiếm tài liệu và tạo ra các sản phẩm thông tin thư mục. Phần mềm được sử dụng để thực hiện chức năng trên gọi là *phần mềm tư liệu*. Các phần mềm tư liệu bao gồm nhiều module. Module trung tâm dành cho việc tìm kiếm các biểu ghi. Các module khác cho phép thiết lập cấu trúc CSDL, cập nhật dữ liệu, hiển thị kết quả trên màn hình và in các kết quả ra giấy. Hiện nay có rất nhiều phần mềm tư liệu trên thị trường. Ta có thể kể ra ở đây vài phần mềm chạy trên môi trường DOS như: TEXTTO, SUPERDOC, INFORBANK. CSD/ISIS đang được sử dụng rộng rãi ở nước ta cũng là một phần mềm tư liệu do UNESCO cung cấp miễn phí cho các nước đang phát triển.

Các phần mềm tư liệu có các đặc tính cơ bản sau:

- Cấu trúc CSDL do người sử dụng thiết lập.
- Quản lý được nhiều CSDL.
- Tạo lập được bộ phiếu đảo cho phép truy nhập nhanh CSDL.
- Format nhập tin có thể thay đổi.
- Tìm tin trên các trường bằng toán tử Boole.
- Từ vựng các từ khoá có thể hiển thị và in ra.
- Có thể trình bày, sắp xếp và in các kết quả tìm.

Các phần mềm tư liệu thích ứng với việc triển khai các ứng dụng tin học ở giai đoạn đầu, với các ứng dụng mang tính cục bộ là xây dựng các CSDL để quản lý tài liệu của từng đơn vị.

Ở Việt Nam phần mềm tư liệu CDS/ISIS do UNESCO cung cấp đã được Trung tâm Thông tin tư liệu Khoa học và Công nghệ Quốc gia cài đặt tiếng Việt, được đưa vào sử dụng từ cuối những năm 80 ở một số thư viện lớn như Thư viện Quốc gia, Thư viện Khoa học Kỹ thuật Trung ương, Thư viện Khoa học tổng hợp Thành phố Hồ Chí Minh. Các cơ quan này đều sử dụng CDS/ISIS để xây dựng, quản lý và khai thác các CSDL tài liệu của mình. Đến cuối năm 1994, được sự giúp đỡ của Bộ Văn hoá - Thông tin, 61 tỉnh thành trong cả nước được trang bị máy tính, CDS/ISIS được triển khai ứng dụng ở thư viện các tỉnh. Cũng trong thời gian này thư viện ở hầu hết các trường đại học ở Hà Nội đều cũng triển khai ứng dụng CDS/ISIS để xây dựng các cơ sở dữ liệu để quản lý sách, giáo trình, tạp chí, luận văn...

Bên cạnh những ưu điểm nổi bật của một phần mềm quản trị các dữ liệu dạng văn bản có cấu trúc như: cấu trúc CSDL do người sử dụng tự xây dựng, có thể quản lý được các trường có độ

dài biến động nhiều, có khả năng nhận biết trường lặp, trường con, có thể tìm tin trên file đảo bằng ngôn ngữ tìm khá mềm dẻo và linh hoạt, có ngôn ngữ tạo format đủ mạnh để có thể in và hiện hình theo ý muốn v.v... CDS/ISIS vẫn có nhược điểm là không thực hiện các phép tính và thống kê, không có khả năng kiểm tra trùng. Ngoài ra CDS/ISIS chỉ là phần mềm quản lý tài liệu, không có khả năng thực hiện các chức năng quản lý khác của thư viện.

Để mở rộng ứng dụng tin học trong các chức năng quản lý khác của thư viện như: theo dõi việc bổ sung tài liệu, tổ chức biên mục tự động, cung cấp khả năng tra cứu tại chỗ hay từ xa cho bạn đọc, quản lý việc mượn tài liệu của bạn đọc, quản lý kho, quản lý lưu thông tài liệu, trao đổi thông tin thư mục với các hệ thống khác, v.v... người ta phải xây dựng các phần mềm có khả năng thực hiện các chức năng trên, gọi là *phần mềm tích hợp quản trị thư viện*.

Trong những phần mềm tích hợp quản trị thư viện khá nổi tiếng trên thế giới có thể kể đến GEAC, MULTILIS, CARD DATALOG, MEDIABOP,...

Ngày nay hầu hết các phần mềm tích hợp quản trị thư viện đều được viết để chạy trên máy vi tính hoặc một mạng máy tính là một hệ tích hợp bao gồm nhiều phân hệ, trong đó các phân hệ chủ yếu là:

- Phân hệ bổ sung
- Phân hệ biên mục
- Phân hệ tra cứu trực tuyến
- Phân hệ lưu thông
- Phân hệ quản lý ấn phẩm định kỳ

- Phân hệ mượn liên thư viện
- Phân hệ quản trị hệ thống

Phân hệ bổ sung có các chức năng: đặt tài liệu, nhận tài liệu, theo dõi hiện trạng thực hiện đơn đặt, kế toán ngân sách bổ sung, xây dựng hồ sơ về các cơ sở cung cấp tài liệu. Phân hệ này còn cung cấp các chức năng báo cáo, thống kê, cho phép quản lý các tài liệu bổ sung và quản lý tài chính có hiệu quả.

Phân hệ biên mục có các chức năng: nhập dữ liệu nhằm tạo các biểu ghi thư mục theo format nhập tin, tiếp nhận dữ liệu từ các CSDL khác trên đĩa từ, đĩa quang hoặc trên mạng, kiểm soát tính quy định thống nhất để bảo đảm tính nhất quán trong quá trình lập tiêu đề tác giả và định chỉ số. Các thông tin đưa vào biểu ghi thư mục bao gồm: các thông tin mô tả (nhân đề, tác giả, các yếu tố xuất bản, số trang, khổ,...), các thông tin chủ đề (ký hiệu phân loại, từ khoá, đề mục chủ đề,...) ngoài ra còn các mã số để xác định tài liệu, trong đó một bộ phận sẽ được lựa chọn để làm các điểm truy cập thông tin. Phân hệ này có khả năng hỗ trợ công tác biên mục theo các tiêu chuẩn và quy tắc mô tả thư mục khác nhau như ISBD, AACR2, TCVN 4734-89 và theo các khung phân loại khác nhau như khung phân loại thập phân của Dewey (DDC), khung phân loại thập phân bách khoa (UDC), khung đề mục chủ đề, v.v...

Phân hệ tra cứu trực tuyến có các chức năng: truy nhập các CSDL thư mục trên mạng thông qua hệ thống mục lục truy nhập công cộng trực tuyến (OPAC), truy nhập từ xa thông qua Internet. Phân hệ này cho phép tìm tin theo nhiều dấu hiệu tìm kiếm khác nhau bằng các công cụ tìm tin đáp ứng các chuẩn quốc tế về tìm tin như sử dụng toán tử logic Boole, toán tử lân cận, toán tử chặt cắt, v.v... cùng với khả năng viết các biểu thức

tìm tin phức hợp thoả mãn những yêu cầu tìm tin đa dạng, khác nhau của người sử dụng.

Phân hệ lưu thông có chức năng: quy định về chế độ phục vụ và cho mượn, theo dõi việc cho mượn và nhận tài liệu trả, quản lý hồ sơ bạn đọc, lập báo cáo thống kê về tình hình mượn đọc và sử dụng kho tài liệu. Đối với chức năng quản lý việc cho mượn tài liệu, phân hệ này cho phép thực hiện các thao tác phục vụ tài liệu cho bạn đọc một cách nhanh chóng và thuận lợi. Việc quản lý vốn tài liệu cũng được tích hợp vào trong phân hệ này.

Phân hệ quản lý ấn phẩm định kỳ có chức năng: đặt, nhận và đăng ký cập nhật từng số của ấn phẩm, biên mục các tên tài liệu mới, kiểm soát lưu thông nếu có chế độ cho mượn báo, tạp chí, lập báo cáo thống kê liên quan đến ấn phẩm định kỳ, kế toán ngân sách bổ sung.

Phân hệ mượn liên thư viện có chức năng: quy định chế độ mượn liên thư viện, kiểm soát việc xuất, nhập tài liệu, quản lý hồ sơ các thư viện mượn, theo dõi tình hình mượn và lập báo cáo thống kê.

Phân hệ quản trị hệ thống có chức năng: lập các báo cáo thống kê tổng hợp, phân tích thông tin thống kê để lập kế hoạch, ra quyết định và điều chỉnh. Phân hệ này cũng cho phép thực hiện các chức năng bảo mật, sao lưu và phục hồi dữ liệu, các biện pháp kỹ thuật bảo đảm hệ thống hoạt động thông suốt và liên tục.

Một yêu cầu quan trọng với các hệ quản trị thư viện tích hợp này là chúng phải được chuẩn hoá, tuân thủ mọi tiêu chuẩn của nghiệp vụ thư viện như ISO-2709, MARC hoặc UNIMARC.

Trong thời gian gần đây nhiều công ty phát triển phần mềm trong và ngoài nước đã quan tâm đến thị trường phần mềm

quản trị thư viện ở Việt Nam: Công ty InfoAccess - Singapore đã giới thiệu 2 phần mềm thư viện Elib và VTLS, công ty Công nghệ tin học Tinh Vân giới thiệu phần mềm thư viện Libol. Elib được phát triển dựa trên công nghệ web, có khả năng đáp ứng các chức năng và hoạt động của một thư viện điện tử. Libol là giải pháp phần mềm tự động hoá thư viện tổng thể với 7 phân hệ chức năng, có khả năng quản lý được các loại tài liệu đa dạng. Để thích ứng với thị trường Việt Nam các phần mềm này đều hỗ trợ tiếng Việt và có khả năng chuyển đổi CSDL từ CDS/ISIS.

Các cuộc giới thiệu này, có kèm theo trình diễn, đã tạo điều kiện cho các nhà quản lý thư viện làm quen với các phần mềm quản trị thư viện, định hướng cho sự lựa chọn và chuẩn bị cho quá trình tin học hoá thư viện ở mức độ cao hơn: tin học hoá toàn diện hoạt động thư viện.

2. BIÊN MỤC TỰ ĐỘNG

2.1. Chuẩn hoá công tác biên mục

Nhu cầu tiếp cận một cách thuận tiện và phổ dụng các thông tin thư mục và phát triển sự hợp tác giữa các đơn vị thông tin đã dẫn đến phải chuẩn hoá công tác biên mục.

Quy tắc mô tả thư mục quốc tế ISBD

Năm 1960, với sự cố gắng của tổ chức tiêu chuẩn quốc tế ISO/TC46 và sự hỗ trợ của Hiệp hội các thư viện quốc tế IFLA, quy tắc mô tả thư mục theo tiêu chuẩn quốc tế ISBD được biên soạn.

Đầu tiên là các quy tắc mô tả thư mục dành cho sách (Monographies) - ISBD (M), rồi đến các quy tắc dành cho các ấn

phẩm định kỳ (Serials) - ISBD (S), sau đó mở rộng ra cho các loại hình tài liệu chuyên dạng khác như:

ISBD (NBM): mô tả tài liệu không ở dạng sách (Non-book Materials).

ISBD (CM): mô tả tài liệu dạng đồ biểu (Cartographic Materials).

ISBD (PM): mô tả ấn phẩm âm nhạc (Printed Music).

ISBD (A): mô tả sách cổ (Antiquarian).

ISBD (CP): mô tả các bài trích (Component Parts).

ISBD (ER): mô tả các nguồn tài liệu điện tử (Electronic Resources).

ISBD là một tập hợp các quy tắc trình bày các dữ liệu thư mục theo một quy định chặt chẽ, cùng với các dấu hiệu dùng để xác định chúng. Nó phân chia các dữ liệu thư mục thành từng vùng và vùng con với những thuyết minh đủ chính xác về các vùng đó. Ngoài ra ISBD còn đưa vào một hệ thống các dấu phân cách để báo hiệu chỗ bắt đầu hoặc kết thúc một vùng, vùng con.

Các vùng mô tả của ISBD

Các dữ liệu thư mục trong ISBD được phân thành 8 vùng mô tả:

1. Vùng nhan đề và thông tin về trách nhiệm.
2. Vùng thông tin về lần xuất bản và trách nhiệm liên quan đến lần xuất bản.
3. Vùng thông tin đặc thù (dành cho ấn phẩm định kỳ và tài liệu chuyên dạng).
4. Vùng địa chỉ xuất bản (phát hành).

5. Vùng mô tả vật lý hay đặc trưng số lượng (số trang, minh hoạ, khổ cỡ).
6. Vùng tít thư (sách bộ).
7. Vùng phụ chú.
8. Vùng chỉ số ISBN, ISSN và điều kiện có được tài liệu.

Trong các vùng trên có những vùng được dùng thường xuyên, nhưng cũng có nhiều vùng ít được sử dụng. Mỗi vùng lại chứa một số yếu tố dữ liệu, còn gọi là vùng con. Ví dụ vùng thông tin về xuất bản bao gồm các yếu tố: nơi xuất bản, nhà xuất bản, năm xuất bản.

ISBD sử dụng các dấu ký hiệu để:

- Chỉ rõ sự bắt đầu của mỗi vùng.
- Phân cách các yếu tố trong một vùng.
- Xác định các yếu tố đặc thù bởi dấu ký hiệu đứng trước chúng.

Ví dụ trong vùng mô tả vật lý: minh hoạ bao giờ cũng đặt sau dấu hai chấm (:), khổ sách bao giờ cũng đặt sau dấu chấm phẩy (;) v.v...

Ví dụ: Dưới đây là một mô tả thư mục theo ISBD:

The vision and dream / Marguerite Hargrove ; photographs by Harry and Claude Frauca. - Cammeray, N.S.W. : Horwitz, 1980.- 304 p. — : ill. ; 18 cm.- (Bucks books)

ISBN 0-7255-0934-1 : \$15.95.

Trong mô tả trên:

Nhan đề:	The vision and dream
Tác giả thứ nhất:	/ Marguerite Hargrove
Tác giả thứ hai:	; photographs by Harry and Claude Frauca
Các yếu tố xuất bản:	.- Cammeray, N.S.W. : Horwitz, 1980
Mô tả vật lý:	.- 304 p. - : ill ; 18 cm.
Tùng thư:	- (Bucks books)
Chỉ số ISBN, giá:	ISBN 0-7255-0934-1 : \$15.95.

Chức năng của ISBD là đặc trưng hoá các yếu tố mô tả dùng để xác định tài liệu, gán một thứ tự cho các yếu tố ấy và quy định các dấu ký hiệu.

Mục đích của ISBD là tạo thuận lợi cho việc trao đổi quốc tế các ấn phẩm thông tin thư mục, cho phép các mô tả này không bị hàng rào ngôn ngữ hạn chế.

Quy tắc mô tả thư mục ISBD có những hạn chế nhất định, bởi vì nó chỉ cho những quy tắc chung. Vì vậy người ta phải lập ra các ủy ban ISBD để nghiên cứu và tổ chức việc áp dụng nó trong các thư viện hay các trung tâm thông tin.

Trong các nước nói tiếng Pháp có tổ chức AFNOR (Hiệp hội các nước nói tiếng Pháp về tiêu chuẩn hoá) đã biên soạn và xuất bản các tập sách hướng dẫn việc ghi chú thư mục, gọi là tiêu chuẩn AFNOR. Quy tắc AFNOR được đưa ra áp dụng từ năm 1976.

Trong các nước nói tiếng Anh, người ta biên soạn ra quy tắc mô tả thư mục Anh - Mỹ AACR (Anglo-American Cataloguing Rules). AACR (còn gọi là AACR1) xuất hiện năm 1967, dựa trên cơ sở của ISBD nhưng chi tiết và sâu sắc hơn, đi sâu xét nhiều trường hợp tỉ mỉ, khắc phục được những khó khăn cụ thể. AACR ưu tiên lấy các thông tin trong trang tiêu đề để đưa vào mô tả và đưa ra các quy tắc mô tả đối với các tư liệu không phải dạng sách báo. Tuy nhiên qua thực tiễn biên mục AACR1 cũng còn thể hiện nhiều nhược điểm.

Năm 1978, với sự hợp tác của các thư viện quốc gia của Anh, Mỹ và Canada, quy tắc mô tả AACR2 được xuất bản và nhanh chóng được áp dụng. AACR2 đã khắc phục được những nhược điểm của quy tắc AACR1, tạo thuận lợi cho việc sử dụng mục lục để tìm tư liệu và cải thiện hợp tác quốc tế trong lĩnh vực biên mục.

2.2. Biên mục tự động

Quá trình ứng dụng tin học trong việc lưu trữ và khai thác tài liệu gắn liền với quá trình thực hiện tự động hoá công tác biên mục, một khâu quan trọng trong dây chuyền xử lý thông tin.

Trong tiếng Anh, có các thuật ngữ tương đương sau dùng để diễn tả khái niệm biên mục tự động:

Automated cataloguing

Computerised cataloguing

Machine readable cataloguing

Thực chất của biên mục tự động là sử dụng một phần mềm tư liệu hoặc sử dụng phân hệ biên mục của phần mềm quản trị thư viện để tạo lập các biểu ghi cho một CSDL thư mục và tạo ra các mục lục thích hợp.

Trong biên mục tự động việc tạo lập biểu ghi thường là xử lý tiền máy và nhập dữ liệu, do con người thực hiện. Người ta nhập dữ liệu qua các khổ mẫu hiển thị trên màn hình theo kiểu xử lý văn bản. Còn việc tổ chức và sắp xếp biểu ghi và biên soạn các mục lục thì do máy tính thực hiện. Máy tính có thể in các phiếu mục lục và chế bản cho các ấn phẩm thư mục. Đó là sản phẩm đầu ra của CSDL thư mục.

Sự ra đời của biên mục tự động cho phép kiểm soát thư mục toàn cầu. Trong biên mục tự động khổ mẫu hay format có ý nghĩa quan trọng hàng đầu.

Quy tắc mô tả thư mục theo tiêu chuẩn quốc tế ISBD chưa hoàn toàn thích nghi với cách xử lý tin học đối với các mô tả thư mục. Các phương pháp xử lý tin học đòi hỏi dữ liệu phải được cấu trúc hoá. Để máy tính có thể nhận biết được các dữ liệu thư mục, các chỉ dẫn không những phải tuân theo những quy tắc chặt chẽ mà còn phải được trình bày một cách chính xác theo một khổ mẫu thống nhất. Ví dụ như kiểu ký tự được sử dụng, độ dài của các trường dữ liệu, thứ tự sắp xếp v.v... Các quy tắc và cách trình bày đó tạo thành một khổ mẫu (format).

Khổ mẫu là hình thức trình bày dữ liệu trên biểu ghi được cấu trúc hoá. Các dữ liệu trên biểu ghi được sắp xếp thành các trường, trường con, kết hợp với các mã số và các chỉ thị để điều hành sự sắp xếp này sao cho có thể nhận biết, trình bày và tìm kiếm dữ liệu bằng máy tính.

Mã số (codes) và chỉ thị (indicators) là các dấu hiệu để nhận biết và điều khiển cách bố trí các trường và trường con, do các

chương trình quản trị CSDL quy ước khi xây dựng. Khi một biểu ghi thư mục đã được đánh dấu một cách chính xác và được lưu trữ dưới dạng một tệp dữ liệu trên máy tính thì chương trình quản trị CSDL sẽ đánh dấu và tạo khuôn dạng cho các thông tin này để in ra thành một bản thư mục, một phiếu mục lục hay hiển thị trên màn hình. Các chương trình này còn cung cấp công cụ tìm kiếm các thông tin thoả mãn yêu cầu đặt ra, dựa trên các điểm truy nhập nằm trong các trường của biểu ghi.

Mỗi hệ quản trị CSDL văn bản thường có 3 loại khổ mẫu:

a- Khổ mẫu hiển thị và in ấn:

- Hiển thị trên màn hình
- In ra các phiếu mô tả

Các phiếu mô tả được sắp xếp theo một tiêu chuẩn chung, với một hệ thống các ký hiệu và dấu phân cách.

b- Khổ mẫu nhập tin:

Khổ mẫu nhập tin là khổ mẫu đầu vào, dùng để nhập các biểu ghi trong CSDL. Có sự khác nhau trong thiết kế, cách trình bày các trường và trường con.

c- Khổ mẫu trao đổi:

Khổ mẫu trao đổi ra đời dựa trên cơ sở các khổ mẫu theo tiêu chuẩn ISO.2709, quy định chia biểu ghi thành các trường, trường con, dùng để biên mục và trao đổi thông tin trên mạng. Format trao đổi bao gồm các yếu tố chung, không bị ngăn cách bởi các phần mềm ứng dụng khác nhau.

Có 3 format trao đổi được dùng hiện nay là:

- CCF (Commun Communication Format - Format trao đổi chung) do UNESCO xây dựng.

- UNIMARC do IFLA và UNESCO xây dựng.
- INTERMARC của Pháp.

3. KHỔ MẪU BIÊN MỤC ĐỌC MÁY MARC

Năm 1966 lần đầu tiên Thư viện Quốc hội Mỹ xuất bản tập quy tắc của khổ mẫu MARC (Machine Readable Cataloging - Biên mục đọc được bằng máy). Khổ mẫu MARC là một mô tả có cấu trúc, dành riêng cho các dữ liệu thư mục được đưa vào máy tính điện tử. Nó là khổ mẫu cho phép máy tính lưu giữ và truy xuất thông tin.

Cấu trúc của khổ mẫu MARC là cấu trúc biểu ghi, trong đó các dữ liệu thư mục được sắp xếp trong các trường, có độ dài xác định, được mã hoá và trình bày theo một quy định chặt chẽ. Khổ mẫu MARC sử dụng các chữ số, chữ cái, các ký hiệu ngắn gọn đặt ngay trong biểu ghi thư mục để đánh dấu và nhận biết các loại thông tin khác nhau trong mỗi biểu ghi.

Mỗi biểu ghi của khổ mẫu MARC bao gồm các trường (fields). Ngoài các trường dành cho các yếu tố mô tả thư mục theo AACR2 như : nhan đề, thông tin về trách nhiệm, thông tin về xuất bản, từng thư, đặc trưng số lượng, phụ chú, tóm tắt, v.v... còn có các trường dành cho đề mục chủ đề, ký hiệu phân loại của Thư viện Quốc hội Mỹ, ký hiệu phân loại thập tiến Dewey,... Các trường này lại có thể chia nhỏ thành các trường con (subfields). Vì tên của trường thường khá dài, nên trong biểu ghi MARC mỗi trường được biểu diễn bằng một nhãn gồm 3 chữ số.

Để tiện trình bày, người ta tập hợp các nhãn trường thành từng nhóm. Các trường có nhãn bắt đầu bằng số "0" thuộc nhóm trường "0XX" là các trường điều khiển, các chỉ số nhận dạng và

phân loại. Các trường có nhãn bắt đầu bằng số “1” thuộc nhóm trường “1XX” là các trường tiêu đề chính (tên cá nhân, tên tập thể, ...), các trường có nhãn bắt đầu bằng số “2” thuộc nhóm trường “2XX” là các trường mô tả (nhân đề và thông tin trách nhiệm, thông tin về xuất bản, lần xuất bản,...), v.v...

Các nhãn trường thường được sử dụng là:

- | | |
|-----|---|
| 010 | Số kiểm tra của Thư viện Quốc hội Mỹ. |
| 020 | Chỉ số sách quốc tế (ISBN) |
| 022 | Chỉ số ấn phẩm định kỳ quốc tế (ISSN) |
| 050 | Chỉ số phân loại của Thư viện Quốc hội Mỹ (LCC) |
| 082 | Chỉ số phân loại thập phân Dewey (DDC) |
| 100 | Tiêu đề chính, tên cá nhân (tác giả) |
| 245 | Nhan đề và thông tin về trách nhiệm |
| 250 | Thông tin về lần xuất bản |
| 260 | Thông tin về xuất bản, phát hành |
| 300 | Mô tả vật lý (đặc trưng số lượng) |
| 440 | Thông tin về từng thư |
| 520 | Chú giải hay tóm tắt |
| 650 | Điểm truy nhập chủ đề, đề mục chủ đề |
| 700 | Điểm truy nhập bổ sung, tên cá nhân |

Danh sách tất cả các nhãn trường được in trong 2 tập *MARC 21 Format for Bibliographic Data* do Thư viện Quốc hội Mỹ xuất bản. Qua thống kê người ta thấy rằng chỉ có khoảng 10% số nhãn được sử dụng thường xuyên trong các biểu ghi MARC, 90% còn lại ít khi được sử dụng.

Ngoài nhãn trường, trong biểu ghi MARC một số trường còn được xác định bằng các *chỉ thị (indicators)*. Chỉ thị được mã hoá bằng hai chữ số (từ 0 đến 9) đi theo sau nhãn trường. Có trường chỉ dùng chữ số thứ nhất hoặc thứ hai, có trường dùng cả hai. Trong các trường không cần chỉ thị người ta bỏ trống.

Ví dụ:

245	14	\$a The emperor's new clothes / \$c adapted from Hans Christian Andersen and illustrated by Janet Stevens
-----	----	---

Trong ví dụ trên, 245 là nhãn trường nhan đề và thông tin về trách nhiệm, chỉ thị thứ nhất "1" có nghĩa là trong phiếu mục lục phải in từ "Title" vào trước nhan đề. Còn chỉ thị thứ hai "4" có nghĩa là bỏ qua 4 ký tự đầu của nhan đề (*T, h, e* và *dấu trống*). Khi đó nhan đề "*The emperor's new clothes*" sẽ được ghi ra là "*emperor's new clothes*".

Trong biểu ghi MARC, trường con được nhận biết bởi *mã trường con (subfield codes)*, đó là một ký tự bằng chữ in thường, đặt sau *dấu ngăn cách \$ (delimiter)*.

Ví dụ:

300	\$a 675p. ; \$b ill. ; \$c 24cm
-----	---------------------------------

Trong đó 300 là nhãn trường mô tả vật lý của cuốn sách, bao gồm các trường con \$a (Số trang), \$b (Thông tin minh hoạ), \$c (Khổ, cỡ). Trường này không cần chỉ thị.

Ví dụ:

260	\$a Newyork : \$b Chelsea House, \$c 1986
-----	---

Trong đó 260 là nhân trường xuất bản, bao gồm các trường con: \$a (Nơi xuất bản), \$b (Nhà xuất bản), \$c (Năm xuất bản).

Các nhân trường, chỉ thị và mã trường con là các dấu hiệu để nhận biết và điều khiển cách bố trí các trường và trường con, do các chương trình quản trị CSDL quy ước khi xây dựng. Khi một biểu ghi thư mục đã được đánh dấu một cách chính xác và được lưu trữ dưới dạng một tệp dữ liệu trên máy tính thì chương trình quản trị CSDL sẽ đánh dấu và tạo khuôn dạng cho các thông tin này để in ra thành một bản thư mục, một phiếu mục lục hay hiển thị trên màn hình. Các chương trình này còn cung cấp công cụ tìm kiếm các thông tin thoả mãn yêu cầu đặt ra, dựa trên các điểm truy nhập nằm trong các trường của biểu ghi MARC.

Ngoài các trường dữ liệu thư mục là phần chính của biểu ghi thư mục, mỗi biểu ghi MARC còn có các thành phần cố định sau, xuất hiện ở đầu mỗi biểu ghi:

a- *Đầu biểu (Leader)*. Đầu biểu gồm 24 ký tự đầu tiên của biểu ghi, ứng với 24 vị trí, mỗi vị trí này được gán cho một ý nghĩa xác định và thể hiện bằng một mã (chữ in thường, chữ số hoặc khoảng trống) cho ta biết các thông tin về trạng thái và các thuộc tính của biểu ghi như: độ dài biểu ghi, loại hình tài liệu, cấp thư mục, mức độ mã hoá, quy tắc mô tả được sử dụng (ISBD, AACR),... Nhiều thông tin trong đầu biểu ghi là để dành cho máy tính sử dụng để nhận dạng biểu ghi.

Ví dụ: Giả sử một biểu ghi MARC có đầu biểu:

01401cam/2200265/a/4500

thì ý nghĩa cơ bản của nó là:

Vị trí	Ký tự	ý nghĩa ký tự của vị trí
0- 4	01041	Độ dài biểu ghi: 1041 ký tự
5	c	Trạng thái biểu ghi: đã sửa chữa
6	a	Dạng tài liệu: văn bản in
7	m	Cấp thư mục: sách chuyên khảo
...		
18	a	Quy tắc mô tả được sử dụng: AACR2
...		

b- Danh mục (Directory). Tiếp theo đầu biểu là một loạt các tiểu dẫn (bằng các chữ số) có độ dài xác định gọi là danh mục. Danh mục này cho biết các nhân trường có trong biểu ghi, độ dài của trường và vị trí bắt đầu của trường trong biểu ghi. Danh mục được xây dựng bởi máy tính từ biểu ghi thư mục, dựa trên các thông tin đã nhập.

Ví dụ: Đoạn tiểu dẫn sau đây trong danh mục

.. 245003600354250001200390260003700402 . . .

có nghĩa là: trong biểu ghi có trường với nhân 245, có độ dài là 36 ký tự và bắt đầu từ vị trí thứ 354. Tiếp theo là trường có nhân 250, có độ dài 12 ký tự và bắt đầu ở vị trí thứ 390 ($36+354=390$). Tiếp theo nữa là trường có nhân 260, có độ dài 12 ký tự, bắt đầu từ vị trí 402 ($402=12+390$). Như vậy đoạn tiểu dẫn trên trong danh mục có cấu trúc và ý nghĩa như sau:

Nhãn	Độ dài	Bắt đầu từ vị trí
....		
245	0036	00354
250	0012	00390
260	0037	00402
....		

Tóm lại mỗi biểu ghi MARC phải bao gồm các thành phần cơ bản sau:

- Đầu biểu gồm 24 ký tự.
- Một danh mục các trường dữ liệu mà với mỗi trường dữ liệu phải bao gồm một nhãn với 3 chữ số, độ dài của trường dữ liệu và vị trí của ký tự đầu tiên.
- Các dữ liệu tương ứng với các trường trong danh mục, chứa các dữ liệu thư mục cần xử lý với độ dài thay đổi.

Dưới đây là ví dụ về một biểu ghi MARC:

Leader	00718nam//2200217/a/4500
001	///93154367
003	DLC
005	19951019113246.0
008	940827s1993///at/ac///b///001/0/eng//
010	Sa///93154367
020	\$a1863731695 (pbk.) :\$c\$22.95
043	\$au-at---
050	00 \$aHQ1391.A8\$bS28 1993

082	00	\$a320.994082\$220
100	1	\$aSawer, Marian,\$d1946-
245	12	\$aA woman's place :\$bwoman and politics in Australia /\$cMarian Sawer and Marian Simms
250		\$a2nd ed.
260		\$aSt. Leonards, N.S.W. :bAllen & Unwin, \$c1993
300		\$axiii, 345 p. :\$bill., ports. ;\$c22 cm.
500		\$aPrevious ed.: Sydney : Allen & Unwin,1984.
500		\$aIncludes index
504		\$aBibliography: p. 309-328.
650	0	\$aWomen in politics\$zAustralia.
650	0	\$aWomen legislators\$zAustralia.
700	1	\$aSimms, Marian.

Trong biểu ghi trên, 001 là nhãn trường số kiểm soát, 003 là nhận dạng số kiểm soát, 005 là ngày và thời gian thực hiện giao tác gần nhất, 008 là các thông tin chung được mã hoá về biểu ghi.

Trong biểu ghi MARC, các trường cách nhau bởi dấu phân cách trường. Trong MARC21, dấu phân cách trường ký hiệu là ^. Dấu phân cách trường đặt ở cuối mỗi trường. Người ta dùng ký hiệu \ để đánh dấu kết thúc mỗi biểu ghi.

Khổ mẫu MARC do Thư viện Quốc hội Mỹ xây dựng đầu tiên và sử dụng nên còn gọi là USMARC. Cấu trúc biểu ghi của khổ mẫu MARC tạo ra nhiều khả năng cho máy tính lựa chọn và sắp xếp các dữ liệu thư mục:

- Cho phép người sử dụng dễ dàng truy cập tới các biểu ghi.

- In ra các thông báo sách mới, các ấn phẩm thư mục, các mục lục dưới dạng thức khác nhau, các nhãn trên gáy sách.
- Trao đổi dữ liệu thư mục với các thư viện khác trong nước và trên thế giới.

Khổ mẫu MARC có ý nghĩa quan trọng trong biên mục tự động. Vì vậy các phần mềm tư liệu và phần mềm quản trị thư viện cần phải được xây dựng tuân theo các chuẩn của khổ mẫu MARC.

Sau khi được chỉnh lý vào năm 1968, khổ mẫu MARC là cơ sở cho sự ra đời của một loạt các khổ mẫu quốc gia như: CAN.MARC của Canada, UK.MARC của Anh, AUS.MARC của Úc, INTERMARC của Pháp, IBERMARC của Tây Ban Nha. Mỗi hệ thống biên mục sử dụng format MARC đều có hướng dẫn riêng về hệ thống phân cách và hình thức trình bày các biểu ghi của mình.

Năm 1977, dựa theo tiêu chuẩn ISO-2709 Hiệp hội các Thư viện Quốc tế (IFLA) đã phổ biến khổ mẫu UNIMARC. Mục đích đầu tiên của UNIMARC là tạo thuận lợi cho sự trao đổi quốc tế các dữ liệu thư mục đọc được bằng máy giữa các trung tâm thư mục quốc gia. UNIMARC cũng có thể được sử dụng như một mô hình để phát triển các khổ mẫu trao đổi mới.

UNIMARC xác định các trường, các dấu hiệu và các mã của các trường con, gán cho một biểu ghi thư mục dưới dạng đọc được bằng máy. UNIMARC xử lý tất cả các loại hình tài liệu như: sách, ấn phẩm định kỳ, tài liệu đồ biểu, âm nhạc, phim ảnh, tài liệu nghe nhìn...

UNIMARC là một áp dụng riêng của tiêu chuẩn ISO 2709 - 1981. Đó là chuẩn quốc tế dùng để xác định cấu trúc của các biểu ghi chứa các dữ liệu thư mục. Chuẩn này quy định mỗi

biểu ghi thư mục phải bao gồm các thành phần giống như các thành phần của biểu ghi MARC, tức là:

- a) Đầu biểu ghi gồm 24 ký tự.
- b) Một danh mục các trường dữ liệu.
- c) Các dữ liệu tương ứng với các trường trong danh mục.

Các khối trường dữ liệu của UNIMARC

0xx- Khối thông tin nhận dạng, dùng để nhận dạng tài liệu và nhận dạng biểu ghi, bao gồm các trường: mã số thư mục quốc gia, mã số lưu chiếu quốc gia, mã số xuất bản, chỉ số ISBN, ISSN, mã số biểu ghi, ...

1xx- Khối thông tin mã hoá, dùng 3 ký tự để chỉ mã nước xuất bản hay mã ngôn ngữ của tài liệu (theo ISO), ví dụ: Vie (Việt Nam), Eng (Anh); mã các loại hình tài liệu như sách, tạp chí, tài liệu đồ biểu, tài liệu nghe nhìn, hình ảnh, đồ thị, âm thanh,...

2xx- Khối thông tin mô tả thư mục: bao gồm các trường, trường con. Về cơ bản giống như ISBD, trừ phụ chú. Ví dụ: nhan đề, thông tin về trách nhiệm, thông tin về lần xuất bản, loại hình tài liệu, đặc trưng số liệu (collation), từng thư (collection - sách bộ),...

3xx- Khối phụ chú: phụ chú về mô tả thư mục, về nhan đề và tác giả, về nguồn gốc tài liệu và về bản dịch,...

4xx- Khối tiểu dẫn liên kết: bộ sưu tập, xuất bản gốc, đồng xuất bản, nhan đề bản dịch,...

5xx- Khối nhan đề có liên quan: nhan đề chính, nhan đề sắp xếp, nhan đề song song, nhan đề ngoài bìa, nhan đề đầu trang, nhan đề cũ (ấn phẩm định kỳ),...

6xx- Khối các điểm tiếp cận thông tin chủ đề: tác giả, nhan đề, nơi xuất bản, ngày tháng, mã vùng địa lý, ký hiệu phân loại, từ khoá,...

7xx- Khối thông tin về trách nhiệm nội dung: tác giả chính, đồng tác giả, tác giả tập thể,...

8xx- Khối thông tin về các dữ liệu quốc tế: mã nước, cơ quan biên mục.

9xx- Khối thông tin về các dữ liệu sử dụng ở quốc gia và nội bộ: ký hiệu kho, số lượng tài liệu trong kho, người xử lý, người kiểm tra ,...

Năm 1984 khổ mẫu trao đổi chung CCF (Common Communication Format) đã được biên soạn nhằm đáp ứng việc trao đổi dữ liệu thư mục giữa các cơ quan sử dụng format MARC. Sản phẩm từ các nguồn khác nhau dần trở nên tương hợp giúp các thư viện có thể trao đổi biểu ghi cho nhau. Việc cho mượn giữa các thư viện cũng thuận lợi hơn nhờ hệ thống mục lục liên hợp đã bảo đảm tính nhất quán trong biên mục. Ngoài ra quá trình nối mạng cũng dễ dàng hơn. Người ta có thể truy nhập các CSDL khác nhau qua việc sử dụng giao thức chuẩn.

Chương 5

MÔ HÌNH DỮ LIỆU VÀ VẤN ĐỀ THIẾT KẾ CƠ SỞ DỮ LIỆU

1. CÁC MÔ HÌNH DỮ LIỆU

Cơ sở dữ liệu là tập hợp các dữ liệu được quản lý trên máy tính điện tử theo một cơ chế thống nhất, nhằm giúp cho việc truy nhập và xử lý dữ liệu được dễ dàng và nhanh chóng. CSDL phải được cấu trúc sao cho có thể dễ dàng lưu trữ, sửa đổi và tìm kiếm dữ liệu kết hợp với nhiều quá trình xử lý dữ liệu khác nhau.

Để xây dựng các CSDL thoả mãn những yêu cầu đặt ra, người ta phải thiết kế những mô hình để bố trí dữ liệu về mặt logic theo một quan điểm hệ thống nhất định. Đó là các mô hình dữ liệu.

1.1. Một số khái niệm cơ bản

Mô hình dữ liệu (Data model) là một tập hợp các khái niệm dùng để mô tả dữ liệu và các thao tác xử lý dữ liệu.

Khi mô hình dữ liệu mô tả khái niệm từ một thực tế cho trước, người ta có mô hình dữ liệu khái niệm (gọi tắt là mô hình khái niệm).

Những khái niệm trong mô hình dữ liệu được xây dựng bằng cách trừu tượng hoá thực tế và được mô tả thông qua các thể hiện ngôn ngữ và đồ hoạ. Đó là các cú pháp và các ký hiệu đồ hoạ, được xác định như một bộ phận của mô hình dữ liệu.

Trừu tượng hoá (Abstraction) là quá trình tư duy, ở đó người ta tập trung vào các tính chất đặc trưng của đối tượng mà quên đi các khác biệt của chúng. Nói cách khác, người ta áp dụng trừu tượng hoá khi ta tập trung vào các tính chất mà ta coi là thiết yếu của đối tượng và loại bỏ đi các tính chất thứ yếu khác.

Ví dụ: Khái niệm xe đạp có thể coi là kết quả của quá trình trừu tượng, mà ở đó ta đã loại bỏ đi các bộ phận chi tiết trong cấu trúc của xe đạp (xích, bánh xe, bàn đạp, khung,...) và tất cả những khác biệt có thể có giữa các chiếc xe đạp khác nhau. Người ta thường kết hợp một tên gọi với một đối tượng đã được trừu tượng hoá.

Người ta dùng mô hình khái niệm để thiết kế CSDL. Ở đây mô hình khái niệm cung cấp cho ta những quy tắc cấu trúc dữ liệu. Từ một mô hình dữ liệu cụ thể người ta xây dựng nên lược đồ.

Lược đồ (Schema) là một tập hợp các thể hiện đồ hoạ và ngôn ngữ cho phép mô tả cấu trúc của dữ liệu và mối quan hệ giữa chúng.

Lược đồ có thể coi như là một cái khung của CSDL, nó bất biến với thời gian.

Mỗi lược đồ có nhiều thể hiện ứng với các trạng thái khác nhau của CSDL tại các thời điểm khác nhau.

Thể hiện (Instance) của lược đồ là tập hợp động thay đổi theo thời gian của các dữ liệu, cho bởi cấu trúc dữ liệu do lược đồ quy định.

Ví dụ: Cần xây dựng lược đồ cho CSDL quản lý xe, mô tả hiện thực là: người sở hữu những chiếc xe.

Trừu xuất thực tế ta được mô hình dữ liệu là các biểu ghi và các trường dữ liệu. Ở đây có 3 biểu ghi (người, xe, sở hữu) và 8 trường (Họ tên, năm sinh, địa chỉ, số chứng minh, số xe, màu xe, màu, ngày đăng ký).

Lược đồ gồm hai loại biểu ghi sau:

- Loại đối tượng được mô hình hoá bằng hai loại biểu ghi là NGƯỜI và XE.

- Loại quan hệ được mô hình hoá bằng biểu ghi SỞ HỮU.

Lược đồ khái niệm của CSDL Quản lý xe:

NGƯỜI

HỌ TÊN	NĂM SINH	ĐỊA CHỈ	SỐ CM
--------	----------	---------	-------

XE

SỐ XE	MÁC XE	MÀU
-------	--------	-----

SỞ HỮU

SỐ CM	SỐ XE	NGÀY ĐK
-------	-------	---------

trong đó:

SỐ CM = Số chứng minh

NGÀY ĐK = Ngày đăng ký

Dưới đây là thể hiện của CSDL Quản lý xe tại một thời điểm:

NGƯỜI

HỌ TÊN	NĂM SINH	ĐỊA CHỈ	SỐ CM
Trần Văn Minh	1972	43 Hàng Ngang	H-112876
Vũ Thuý Hiền	1968	107 Bà Triệu	B-109872
Lê Hoài Nam	1974	24 Quán Sứ	T-234518
Trịnh Văn An	1965	18 Ngô Văn Sở	H-140054
Vũ Huy Hùng	1969	4 Hàng Đào	A-247715
Ngô Hồng Vân	1970	14 Hàng Khoai	A-431087
Trần Thu Hà	1973	55 Hàng Chiếu	N-510186

XE

SỐ XE	MÁC XE	MÀU
F2-5460	Honda	Đỏ
F2-5461	Suzuki	Đen
F2-5462	Honda	Xanh thẫm
F2-5463	Honda	Đỏ
F2-5464	Vespa	Xanh nhạt
F2-5465	Suzuki	Đen
F2-5466	Honda	Xanh thẫm

SỞ HỮU

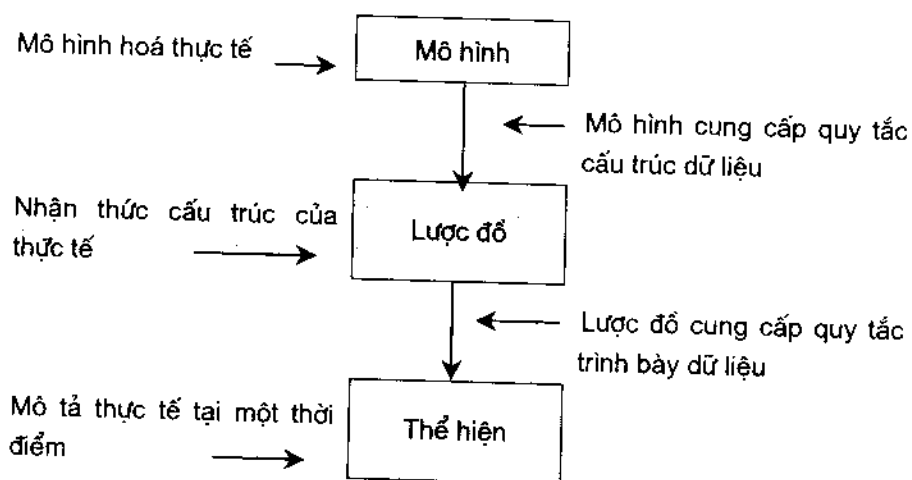
SỐ CM	SỐ XE	NGÀY ĐK
H-112876	F2-5460	23-2-1995
B-109872	F2-5461	23-2-1995
T-234518	F2-5462	24-2-1995
H-140054	F2-5463	24-2-1995
A-247715	F2-5464	24-2-1995
A-431087	F2-5465	25-2-1995
N-510186	F2-5466	25-2-1995

Tóm lại:

Mô hình dữ liệu cung cấp cho ta quy tắc cấu trúc dữ liệu. Cấu trúc đó được thể hiện bằng lược đồ.

Lược đồ cung cấp cách trình bày dữ liệu. Lược đồ có thể coi là nhận thức cấu trúc của thực tế.

Một trình bày thực tế bằng các giá trị của CSDL tại một thời điểm gọi là thể hiện của CSDL. Vậy thể hiện có thể coi là một mô tả thực tế tại một thời điểm.



Hình 6. Quan hệ giữa mô hình, lược đồ, thể hiện

1.2. các loại mô hình dữ liệu

Có hai loại mô hình dữ liệu là mô hình khái niệm và mô hình logic.

Mô hình khái niệm cung cấp cho ta những quy tắc cấu trúc dữ liệu bằng tập hợp các khái niệm từ thực tế của thế giới thực. Mô hình khái niệm là hình thức mô tả thực tế sao cho dễ hiểu và dễ giải thích. Đó là một công cụ để trình bày thực tế ở trình độ trừu tượng cao.

Trong các mô hình khái niệm thì mô hình thực thể quan hệ (Entity Relationship Model) là mô hình có nhiều ưu điểm được dùng phổ biến nhất hiện nay.

Mô hình logic cung cấp những mô tả dữ liệu sao cho chúng có thể xử lý bằng máy tính điện tử. Có ba loại mô hình logic cơ bản là:

- Mô hình mạng (Network Data Model)

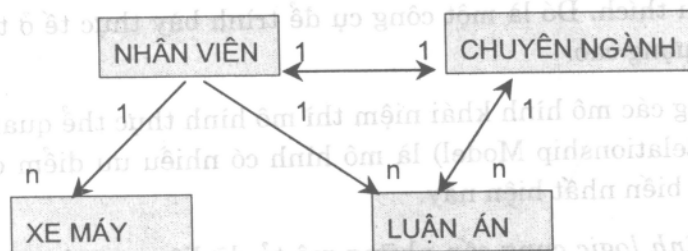
- Mô hình phân cấp (Hierarchical Data Model)
- Mô hình quan hệ (Relational Data Model)

Trong ba loại mô hình trên, mô hình quan hệ có nhiều ưu điểm và được quan tâm hơn cả vì nó được xây dựng trên một cơ sở toán học chặt chẽ, đó là lý thuyết về các quan hệ. Nó có hình ảnh trực quan gần gũi với quan niệm thông thường của người sử dụng. Các ngôn ngữ thao tác trên các CSDL quan hệ dễ sử dụng và có hiệu quả cao.

Mô hình mạng

Mô hình mạng được xây dựng trên cơ sở các tập dữ liệu và các quan hệ dạng 1-1 hoặc 1- n giữa các tập. Cấu trúc mạng cho phép thể hiện mối quan hệ liên kết giữa dữ liệu này với một nhóm dữ liệu khác.

Ví dụ: Một CSDL quản lý chuyên môn của các nhân viên cùng với phương tiện đi lại bằng xe máy của họ. CSDL này có các tập cùng với các quan hệ giữa chúng được biểu diễn bằng sơ đồ ở Hình 7.



Hình 7. Mô hình mạng

Để tạo lập một CSDL theo mô hình mạng cần các thao tác sau:

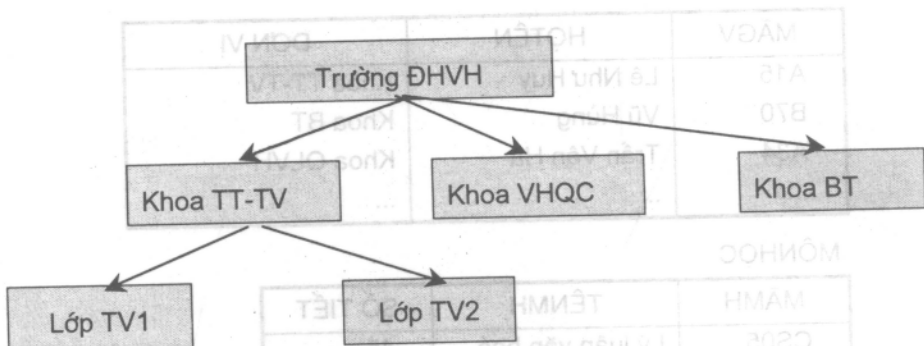
- Tạo lập các tệp: bao gồm việc khai báo tên tệp và các thuộc tính của tệp.

- Thiết lập quan hệ giữa hai tệp: bao gồm việc khai báo tên quan hệ, tên tệp nguồn, tên tệp đích và kiểu quan hệ.

Mô hình phân cấp

Mô hình phân cấp là trường hợp riêng của mô hình mạng, trong đó khái niệm tệp vẫn giữ nguyên, còn khái niệm quan hệ được giới hạn ở kiểu quan hệ phân cấp, ở đó giữa hai tệp có không quá một quan hệ và quan hệ này tuân theo trật tự trên dưới. Trong mô hình phân cấp các dữ liệu được tổ chức dưới dạng cấu trúc cây.

Một mô hình phân cấp thường gặp trong máy tính là mô hình quản lý thư mục và các tệp. Đặc điểm nổi bật trong các thủ tục truy nhập tới một đối tượng trong mô hình phân cấp là đường dẫn. Đó là đường đi từ gốc đến phần tử cần xét trong cây phân cấp.



Hình 8. Mô hình phân cấp

Ví dụ: Một CSDL quản lý chất lượng đào tạo của một trường đại học có thể có cấu trúc phân cấp như ở hình vẽ 8.

Mô hình quan hệ

Mô hình quan hệ lần đầu tiên được E.Codd - một nhân viên của hãng IBM đề xuất năm 1970. Mô hình quan hệ được xây dựng dựa trên hai ngành của toán học là lý thuyết tập hợp và logic mệnh đề. Tên gọi "mô hình quan hệ" xuất phát từ thuật ngữ "quan hệ" là một bộ phận của lý thuyết tập hợp.

Trong mô hình quan hệ dữ liệu được lưu trữ trong các *quan hệ*, đó là một bảng hai chiều mà mỗi dòng là một *bộ* hay *biểu ghi*, mỗi cột là một *thuộc tính* hay một *trường*. Thứ tự của các biểu ghi hay các trường trong bảng hoàn toàn không ảnh hưởng đến cấu trúc vật lý của bảng và mỗi biểu ghi được xác định bởi một số hiệu lưu trong một trường riêng. Hai đặc trưng này cho phép dữ liệu tồn tại độc lập, không phụ thuộc vào cách lưu trữ nó trong máy tính.

Ví dụ: Một CSDL quản lý việc giảng dạy của giảng viên có thể bao gồm ba quan hệ sau:

GIẢNGVIÊN

MÃGV	HỌTÊN	ĐƠN VỊ
A15	Lê Như Huy	Khoa TT-TV
B70	Vũ Hùng	Khoa BT
A34	Trần Văn Hà	Khoa QLVH
...	...	...

MÔN HỌC

MÃMH	TÊN MH	SỐ TIẾT
CS05	Lý luận văn hoá	45
CN12	Tin học tư liệu	60
CS25	Bảo tồn di tích	60
CN18	Biên mục mô tả	45
...	...	...

PHÂN CÔNG

MÃGV	MÃMH	LỚP	NGÀYKĐ	NGÀYKT
A15	CN12	Lớp TT-TV30	1-3-95	30-3-95
A15	CN18	Lớp TT-TV29	1-4-95	30-4--95
B70	CS25	Lớp BT16	10-4-95	16-5-95
A34	CS05	Lớp QLVH15	1-3-92	12-4-94
A34	CS05	Lớp QLVH16	15-5-93	27-6-94
...	...	...	...	...

Mô hình quan hệ có những ưu điểm sau:

- Có cơ sở toán học chặt chẽ, cho phép áp dụng rộng rãi các công cụ toán học và logic.
- Khá tự nhiên, gần với quan niệm thông thường của người sử dụng.
- Ngôn ngữ thao tác dữ liệu trong sáng và có khả năng tổ hợp cao.
- Dễ bảo đảm tính an toàn dữ liệu.
- Dễ bảo đảm tính độc lập dữ liệu.
- Dễ truy nhập tới các đơn vị dữ liệu.

Ta sẽ trình bày chi tiết về mô hình quan hệ ở chương sau.

Các hệ quản trị CSDL khác nhau được phát triển dựa trên các mô hình dữ liệu khác nhau. Mỗi loại hệ quản trị CSDL đều được xây dựng dựa trên một mô hình dữ liệu xác định. Do những ưu điểm nổi bật như đã nêu ở trên, nhiều hệ quản trị CSDL nổi tiếng như Foxpro, Access, Oracle đều được xây dựng trên mô hình quan hệ.

2. KHÁI NIỆM CHUNG VỀ THIẾT KẾ CƠ SỞ DỮ LIỆU

2.1. Thiết kế CSDL trong vòng đời của hệ thống thông tin

Trong mấy chục năm gần đây chúng ta đã chứng kiến sự phát triển mạnh mẽ về số lượng cũng như mức độ quan trọng trong việc ứng dụng CSDL. Các CSDL là thành phần cơ bản của hệ thống thông tin. Thiết kế CSDL trở thành một hoạt động thông dụng, được thực hiện không những bởi các cán bộ chuyên môn mà cả các cán bộ không chuyên.

Vào cuối những năm 60, khi các CSDL đầu tiên ra đời, các nhà thiết kế CSDL xoay xở như thợ thủ công với các công cụ rất thô sơ; họ thường sử dụng các sơ đồ khối và cấu trúc biểu ghi và thiết kế CSDL thường bị lẫn với cài đặt CSDL. Ngày nay tình hình này đã thay đổi, các phương pháp và các mô hình thiết kế CSDL đã phát triển song song với quá trình công nghệ hệ thống CSDL. Công nghệ CSDL đã có nền tảng lý thuyết, bao gồm lý thuyết quan hệ về dữ liệu, xử lý và tối ưu hoá câu hỏi, kiểm soát trùng lặp, quản lý giao tác và phục hồi sai sót, v.v...

Khi công nghệ CSDL tiến lên, công nghệ thiết kế và kỹ thuật cũng phát triển. Người ta đạt được nhiều sự nhất trí, chẳng hạn, chia quá trình thiết kế CSDL thành các pha, đặt mục tiêu chính cho mỗi pha và các kỹ thuật đạt được các mục tiêu đó.

Thiết kế CSDL là một hoạt động thiết yếu trong việc phát triển hệ thống thông tin và được đặt trong vòng đời của hệ thống thông tin.

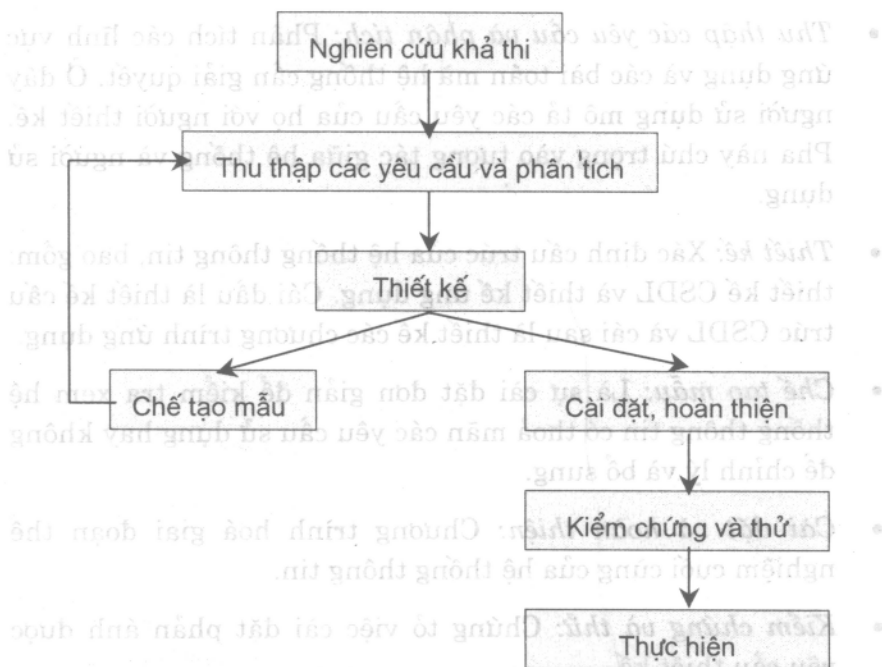
Vòng đời của hệ thống thông tin bao gồm:

- *Nghiên cứu tính khả thi*: Xác định hiệu quả của các phương án khác nhau trong việc thiết kế hệ thống thông tin.

- *Thu thập các yêu cầu và phân tích:* Phân tích các lĩnh vực ứng dụng và các bài toán mà hệ thống cần giải quyết. Ở đây người sử dụng mô tả các yêu cầu của họ với người thiết kế. Pha này chú trọng vào tương tác giữa hệ thống và người sử dụng.
- *Thiết kế:* Xác định cấu trúc của hệ thống thông tin, bao gồm: thiết kế CSDL và thiết kế ứng dụng. Cái đầu là thiết kế cấu trúc CSDL và cái sau là thiết kế các chương trình ứng dụng.
- *Chế tạo mẫu:* Là sự cài đặt đơn giản để kiểm tra xem hệ thống thông tin có thoả mãn các yêu cầu sử dụng hay không để chỉnh lý và bổ sung.
- *Cài đặt và hoàn thiện:* Chương trình hoá giai đoạn thể nghiệm cuối cùng của hệ thống thông tin.
- *Kiểm chứng và thử:* Chứng tỏ việc cài đặt phản ánh được yêu cầu thiết kế.
- *Thực hiện:* Đưa dữ liệu vào, vận hành sự hoạt động của hệ thống và kết thúc khi hệ thống trở nên lạc hậu, cần phải thay thế.

Như vậy thiết kế CSDL là một khâu quan trọng trong vòng đời của hệ thống thông tin. Đó là quá trình xây dựng cấu trúc cho CSDL.

Có thể nói quá trình thiết kế CSDL là quá trình xuất phát từ thế giới thực, thông qua phân tích chức năng, rút ra các dữ liệu và các quy tắc vận hành, sau đó phân tích dữ liệu để xây dựng nên mô hình dữ liệu rồi tiến hành cài đặt vật lý để tạo ra CSDL.



Hình 9. Vòng đời của hệ thống thông tin

2.2. Trừu tượng hoá trong thiết kế khái niệm CSDL

Như đã trình bày ở trên, trừu tượng hoá là quá trình tư duy, ở đó người ta tập trung vào các tính chất đặc trưng của đối tượng mà quên đi các khác biệt của chúng. Người ta áp dụng trừu tượng hoá khi cần phải tập trung vào các tính chất mà ta coi là thiết yếu của đối tượng và loại bỏ đi các tính chất thứ yếu khác.

Có ba loại trừu tượng được dùng trong thiết kế CSDL là:

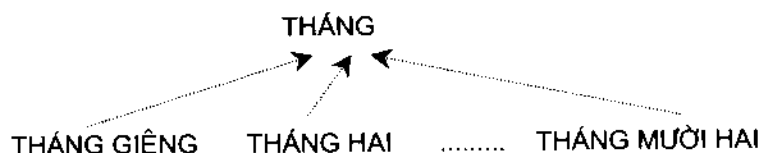
- Trừu tượng phân loại
- Trừu tượng gộp lớn
- Trừu tượng tổng quát hoá

Trừu tượng phân loại

Trừu tượng phân loại được dùng để xác định một khái niệm như một loại các khái niệm của thế giới thực được đặc trưng bởi những tính chất chung.

Chẳng hạn, khái niệm *xe đạp* là loại mà các thành viên của nó là tất cả các xe đạp (xe đạp nam, xe đạp nữ, xe đạp màu đỏ, xe đạp của Liên,...), khái niệm *tháng* là loại mà các thành phần của nó là tháng giêng, tháng hai,..., tháng chạp. Khi nói về tháng, ta đã trừu tượng từ các nét riêng của mỗi tháng (ví dụ số ngày) và ta nhấn mạnh vào các khía cạnh chung của các tháng: đó là tập hợp các ngày mà ta có thể xác định được ngày đầu và ngày cuối và có kích thước xấp xỉ nhau (từ 28 đến 31 ngày).

Sự phân loại được mô tả như cây một mức, trong đó nút gốc là lớp và lá là các phần tử của lớp.

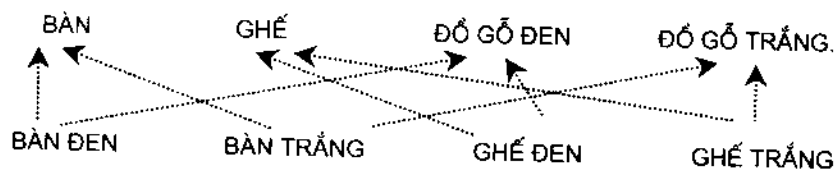


Hình 10. Ví dụ về phân loại

Cùng một tập hợp đối tượng của thế giới thực có thể phân loại theo nhiều cách khác nhau. Chẳng hạn, ta xét tập hợp các đồ vật:

{GHẾ ĐEN, BÀN ĐEN, GHẾ TRẮNG, BÀN TRẮNG}

Ta có thể phân loại các đồ vật trên thành BÀN và GHẾ hay nếu ta quan tâm đến màu sắc thì ta phân loại chúng thành ĐỒ GỖ ĐEN và ĐỒ GỖ TRẮNG.



Hình 11. Nhiều cách phân loại đối với cùng một tập hợp đối tượng

Trừu tượng gộp lớn

Trừu tượng gộp lớn xác định một lớp mới từ một tập hợp các lớp khác, các lớp này được coi là các bộ phận cấu thành.

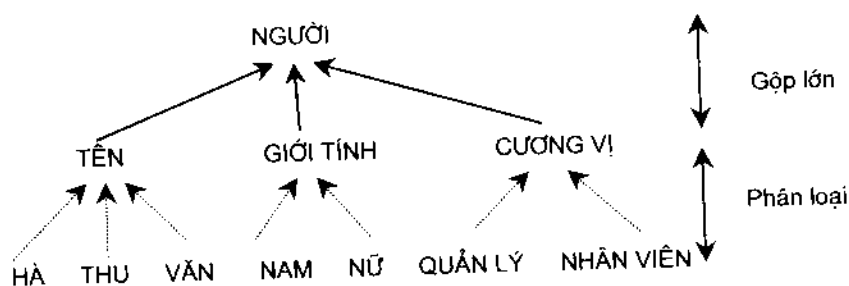
Người ta dùng loại trừu tượng này để thu được một lớp lớn hơn từ các lớp nhỏ. Chẳng hạn, từ các lớp vành, bàn đạp, tay lái... ta có thể nhận được lớp xe đạp. Tương tự lớp người được trừu tượng hoá gộp lớn từ các lớp tên, giới tính và lương.

Trừu tượng gộp lớn được thể hiện bằng cây một mức, trong đó mỗi nút ứng với một lớp, gốc thể hiện lớp được tạo ra như là gộp lớn của các lớp thành phần tại nút lá.



Hình 12. Ví dụ về trừu tượng gộp lớn

Phân loại và gộp lớn là hai trừu tượng cơ bản được dùng để xây dựng cấu trúc dữ liệu trong CSDL và nhiều ngôn ngữ lập trình. Phân loại là quá trình được dùng khi xuất phát từ các *thông tin* mà người ta xác định lớp *trường dữ liệu* hay *thuộc tính*. Gộp lớn là quá trình trong đó người gộp các *trường dữ liệu* có liên quan vào một nhóm thành các *mẫu biểu ghi*.



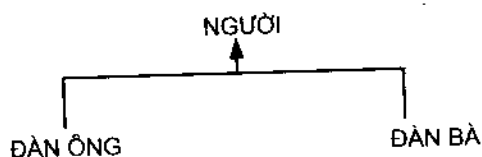
Hình 13. Phân loại và gộp lớn là cơ chế để xây dựng cấu trúc dữ liệu

Trừu tượng tổng quát hoá

Trừu tượng tổng quát hoá xác định các quan hệ tập con giữa các phần tử của hai hay nhiều loại.

Ví dụ: Lớp XE là tổng quát hoá của lớp XE ĐẠP vì mỗi XE ĐẠP là một XE. Cũng vậy lớp NGƯỜI là tổng quát hoá của các lớp con ĐÀN ÔNG và ĐÀN BÀ.

Tổng quát hoá được biểu diễn bằng cây một mức, trong đó tất cả các nút là các lớp mà gốc ứng với lớp tổng quát và các nút lá ứng với các lớp con.



Hình 14. Ví dụ về tổng quát hoá

Ba sự trừu tượng trên là độc lập với nhau và mỗi trừu tượng đều bảo đảm cơ chế của quá trình cấu trúc hoá thông tin. Thật vậy: phân loại tương ứng với các thành viên của tập hợp (IS-MEMBER-OF), gộp lớn tương ứng với các bộ phận của tập hợp (IS-PART-OF), tổng quát hoá tương ứng với quan hệ tập con (IS-A).

2.3. Các pha trong thiết kế CSDL

Thiết kế CSDL là một quá trình phức hợp, bao gồm các mức độ khác nhau. Đó là thiết kế CSDL ở các mức khái niệm, logic và cài đặt vật lý.

Thiết kế khái niệm

Thiết kế khái niệm nhấn mạnh vào mô hình hoá thế giới thực. Mục đích của thiết kế khái niệm là mô tả nội dung thông tin của CSDL dựa trên mô hình khái niệm. Đó là một tập hợp các khái niệm từ thực tế của thế giới thực. Các khái niệm này được mô tả bằng ngôn ngữ và các biểu diễn đồ thị.

Sản phẩm của thiết kế khái niệm là mô hình khái niệm. Mô hình khái niệm là mô tả ở trình độ cao cấu trúc của CSDL. Nó độc lập với phần mềm quản trị CSDL được sử dụng.

Mô hình khái niệm cung cấp quy tắc cấu trúc dữ liệu và là một ngôn ngữ dùng để mô tả lược đồ khái niệm của CSDL. Lược đồ khái niệm cung cấp những quy tắc trình bày dữ liệu và cũng không phụ thuộc vào các phần mềm quản trị CSDL được sử dụng để quản lý CSDL.

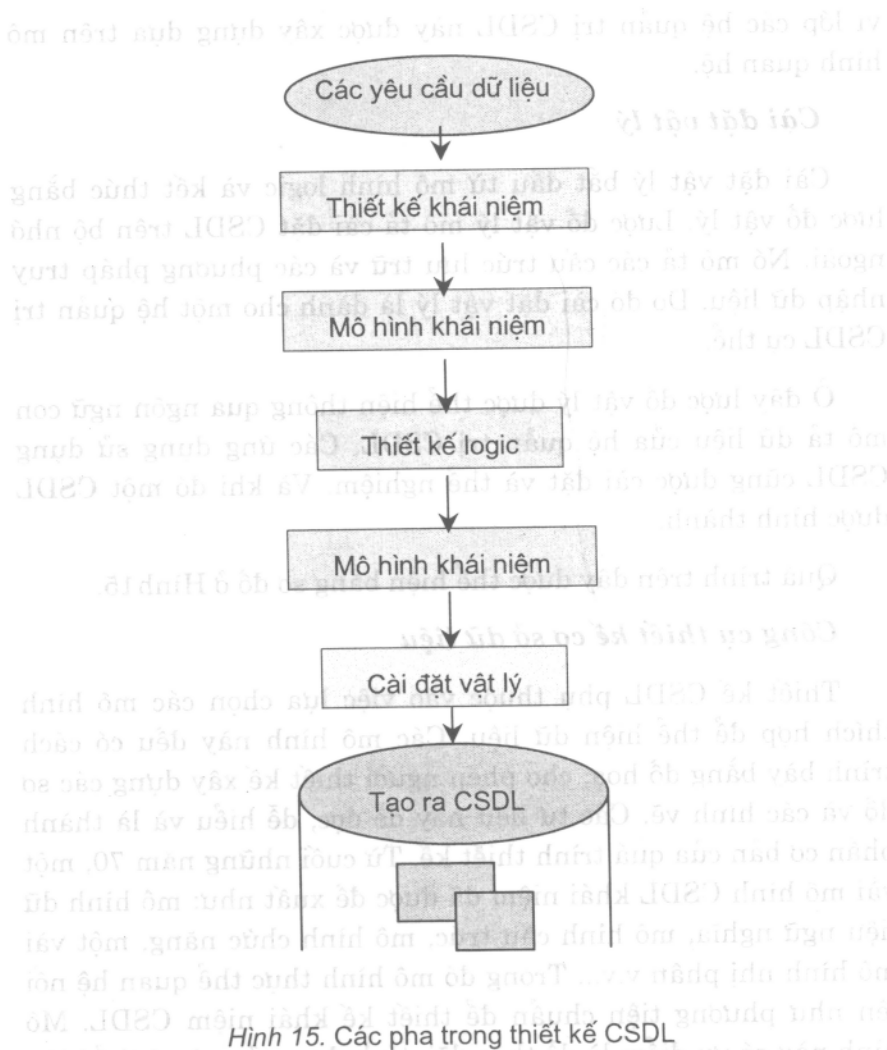
Thiết kế logic

Thiết kế logic bắt đầu từ mô hình khái niệm và cho ra kết quả là mô hình logic.

Mô hình logic cung cấp một mô tả dữ liệu sao cho chúng có thể xử lý bằng máy tính điện tử, tức là mô tả cấu trúc của CSDL mà hệ quản trị CSDL quản lý.

Có ba mô hình logic là:

- Mô hình mạng
- Mô hình phân cấp
- Mô hình quan hệ



Hình 15. Các pha trong thiết kế CSDL

Thiết kế logic phụ thuộc vào lớp các hệ quản trị CSDL sử dụng mô hình logic mà người thiết kế lựa chọn, nhưng không phụ thuộc vào một hệ quản trị CSDL cụ thể nào. Ví dụ: Khi dùng hệ quản trị CSDL Foxpro, Access hoặc Oracle để xây dựng và quản lý CSDL, ta phải thiết kế logic theo mô hình quan hệ,

vì lớp các hệ quản trị CSDL này được xây dựng dựa trên mô hình quan hệ.

Cài đặt vật lý

Cài đặt vật lý bắt đầu từ mô hình logic và kết thúc bằng lược đồ vật lý. Lược đồ vật lý mô tả cài đặt CSDL trên bộ nhớ ngoài. Nó mô tả các cấu trúc lưu trữ và các phương pháp truy nhập dữ liệu. Do đó cài đặt vật lý là dành cho một hệ quản trị CSDL cụ thể.

Ở đây lược đồ vật lý được thể hiện thông qua ngôn ngữ con mô tả dữ liệu của hệ quản trị CSDL. Các ứng dụng sử dụng CSDL cũng được cài đặt và thử nghiệm. Và khi đó một CSDL được hình thành.

Quá trình trên đây được thể hiện bằng sơ đồ ở Hình 15.

Công cụ thiết kế cơ sở dữ liệu

Thiết kế CSDL phụ thuộc vào việc lựa chọn các mô hình thích hợp để thể hiện dữ liệu. Các mô hình này đều có cách trình bày bằng đồ họa, cho phép người thiết kế xây dựng các sơ đồ và các hình vẽ. Các tư liệu này dễ đọc, dễ hiểu và là thành phần cơ bản của quá trình thiết kế. Từ cuối những năm 70, một vài mô hình CSDL khái niệm đã được đề xuất như: mô hình dữ liệu ngữ nghĩa, mô hình cấu trúc, mô hình chức năng, một vài mô hình nhị phân v.v... Trong đó mô hình thực thể quan hệ nổi lên như phương tiện chuẩn để thiết kế khái niệm CSDL. Mô hình này có ưu điểm là dễ theo dõi, tính đơn giản và có thể hiện đồ họa rõ ràng.

Còn trong các mô hình logic thì mô hình quan hệ thể hiện những ưu điểm nổi bật. Nhiều hệ quản trị CSDL nổi tiếng hiện nay như Foxpro, Oracle, Access... đều sử dụng mô hình này.

Chương 6

MÔ HÌNH THỰC THỂ QUAN HỆ VÀ MÔ HÌNH QUAN HỆ

Người ta dùng mô hình dữ liệu để thiết kế CSDL. Có thể nói quá trình thiết kế CSDL là quá trình xuất phát từ thế giới thực, thông qua phân tích chức năng, rút ra các dữ liệu và các quy tắc vận hành, sau đó phân tích dữ liệu để xây dựng nên mô hình dữ liệu, rồi tiến hành cài đặt vật lý để tạo ra CSDL.

Trong chương này ta sẽ đi sâu giới thiệu hai mô hình dữ liệu tiêu biểu được sử dụng nhiều nhất hiện nay để thiết kế khái niệm và thiết kế logic CSDL. Đó là:

- Mô hình thực thể quan hệ (Entity Relationship Model).
- Mô hình quan hệ (Relational Data Model).

1. MÔ HÌNH THỰC THỂ QUAN HỆ

Mô hình thực thể quan hệ, viết tắt là mô hình ER, được Peter Chen giới thiệu năm 1976. Mô hình ER khá đơn giản, được xây dựng trên một số ít khái niệm cơ bản và có biểu diễn đồ thị rõ ràng, dễ hiểu.

1.1. Các khái niệm cơ bản

a. Thực thể (Entity)

Các thực thể thể hiện lớp các đối tượng của thế giới thực.

Ví dụ: Thực thể người, thực thể tài liệu, thực thể từ khoá,...

b. Quan hệ (Relationship)

Các quan hệ thể hiện mối liên hệ của hai hay nhiều thực thể.

Ví dụ:

- Quan hệ "sinh ra tại" liên kết hai thực thể người và thành phố (quan hệ hai chiều).

- Quan hệ "giảng" liên kết các thực thể giáo trình, ngày, phòng học (quan hệ nhiều chiều).

c. Thuộc tính (Attribut)

Các thuộc tính thể hiện các tính chất cơ bản của các thực thể hoặc các quan hệ.

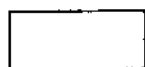
Tất cả các thông tin phát triển đều được thể hiện thông qua các thuộc tính.

Ví dụ:

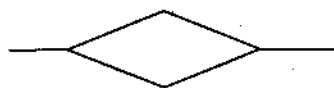
Thực thể người có các thuộc tính sau: tên, số chứng minh, ngày sinh, nghề nghiệp, trình độ,...

Thuộc tính của quan hệ "sinh ra tại" là ngày sinh.

Một số ký hiệu đồ họa dùng trong mô hình ER



Thực thể



Quan hệ



Thuộc tính

Các quan hệ cơ bản

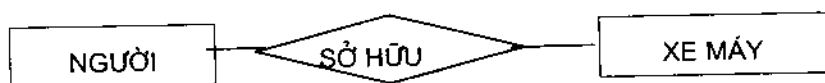
a) Quan hệ một - một (1-1)

Ví dụ:



b) Quan hệ một - nhiều (1- n)

Ví dụ:



c) Quan hệ nhiều - nhiều (n- n)

Ví dụ:



1.2. Phương pháp thiết kế khái niệm cơ sở dữ liệu bằng mô hình thực thể quan hệ

Để thiết kế khái niệm một CSDL bằng mô hình ER, ta tiến hành các bước sau đây:

- 1) **Xác định dữ liệu:** Xác định các dữ liệu về đối tượng phải khảo sát trong CSDL.

- 2) Phân lớp: Dùng phương pháp phân lớp để xác định các thực thể, các quan hệ và các thuộc tính của các thực thể và quan hệ.
- 3) Xây dựng mô hình ER: Sử dụng ký hiệu đồ hoạ để xây dựng mô hình ER cho CSDL.
- 4) Kiểm tra mô hình ER.

Ví dụ sau đây sẽ minh hoạ cho quá trình thiết kế khái niệm một CSDL theo mô hình thực thể quan hệ. Giả sử cần xây dựng một CSDL về các công trình nghiên cứu khoa học của một học viện. Mỗi tài liệu khoa học có một nhan đề và được công bố vào một ngày xác định. Nội dung của tài liệu được mô tả bởi các từ khoá.

Ta gọi tên CSDL này là CTKH.

Bước 1: Xác định dữ liệu: Tập hợp các đối tượng biểu diễn trong CSDL bao gồm:

- NHÂN VIÊN: Tập hợp các cán bộ nghiên cứu.
- PHÂN VIỆN: Tập hợp các phân viện của học viện.
- HỌ TÊN: Tập hợp họ tên của các cán bộ nghiên cứu.
- TÀI LIỆU: Tập hợp các công trình khoa học do học viện xuất bản.
- NHAN ĐỀ: Tập hợp các nhan đề của các công trình khoa học.
- NGÀY THÁNG: Tập hợp các ngày tháng năm xuất bản tài liệu.
- TỪ KHOÁ: Tập hợp các từ chuẩn mô tả nội dung tài liệu.
- THUẬT NGỮ: Tập hợp các thuật ngữ biểu diễn các từ chuẩn.
- LĨNH VỰC: Tập hợp các lĩnh vực chuyên môn.

Các tập hợp đối tượng này có mối liên kết lẫn nhau. Ta cần liệt kê các mối quan hệ này và làm rõ các đặc trưng của chúng.

Ta thấy mỗi cán bộ nghiên cứu có một tên, nhưng nhiều cán bộ nghiên cứu có thể có cùng một tên. Như vậy quan hệ giữa tập hợp **NHÂN VIÊN** và tập hợp **HỌ TÊN** là quan hệ 1- n.

Cũng vậy, một **PHÂN VIỆN** bao gồm n người, nhưng một người chỉ thuộc một **PHÂN VIỆN**, nên quan hệ **PHÂN VIỆN** - **NHÂN VIÊN** là quan hệ 1- n.

Tương tự:

Quan hệ **NHÂN VIÊN** - **TÀI LIỆU** là quan hệ n - n

Quan hệ **TÀI LIỆU** - **NHAN ĐỀ** là quan hệ 1- 1

Quan hệ **TÀI LIỆU** - **TỪ KHOÁ** là quan hệ n - n

Quan hệ **TÀI LIỆU** - **NGÀY THÁNG** là quan hệ n - 1

Quan hệ **TỪ KHOÁ** - **LĨNH VỰC** là quan hệ n - 1

Quan hệ **TỪ KHOÁ** - **THUẬT NGỮ** là quan hệ 1-1

Các mối quan hệ trên giữa các đối tượng có thể biểu diễn bằng sơ đồ Hình 15.

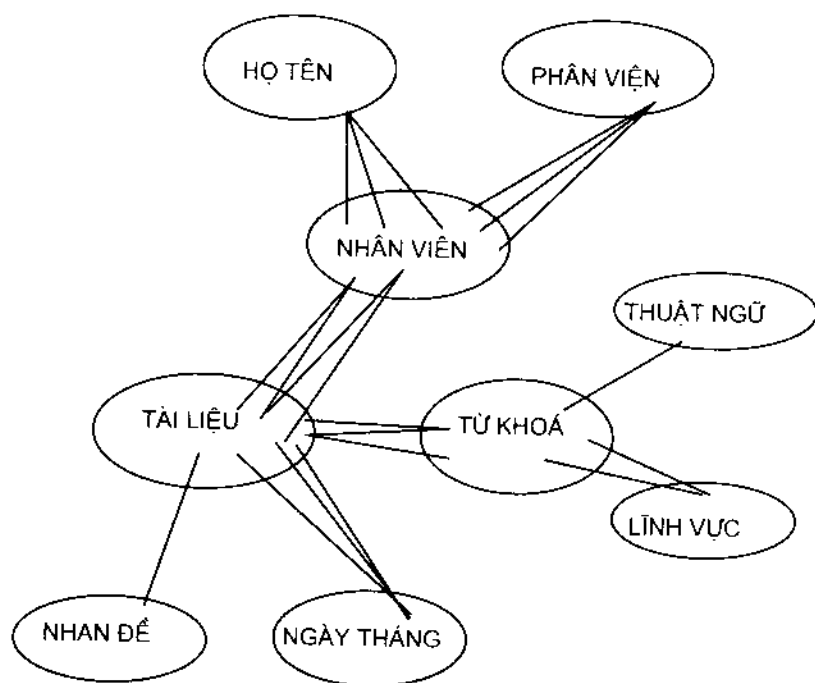
Bước 2: Để xác định các thực thể và các thuộc tính của chúng ta chú ý rằng trong các đối tượng trên đây ta có thể phân biệt hai loại:

- Một loại bao gồm các đối tượng có thể biểu diễn bằng một chuỗi các ký tự như:

NHANĐỀ, NGÀYTHÁNG, HỌTÊN, THUẬTNGỮ, PHÂNVIỆN, LĨNH VỰC

- Loại khác bao gồm các đối tượng mà ta phải gán cho một mã số để xác định chúng một cách rõ ràng như:

NHÂN VIÊN, TÀI LIỆU, TỪ KHOÁ



Hình 16. Quan hệ giữa các đối tượng mô tả trong CSDL

Từ lược đồ trên ta thấy rằng tham gia CSDL trên có ba thực thể cùng các thuộc tính sau:

- Thực thể nhân viên, có các thuộc tính: MANV (mã nhân viên), họ tên, phân viện.

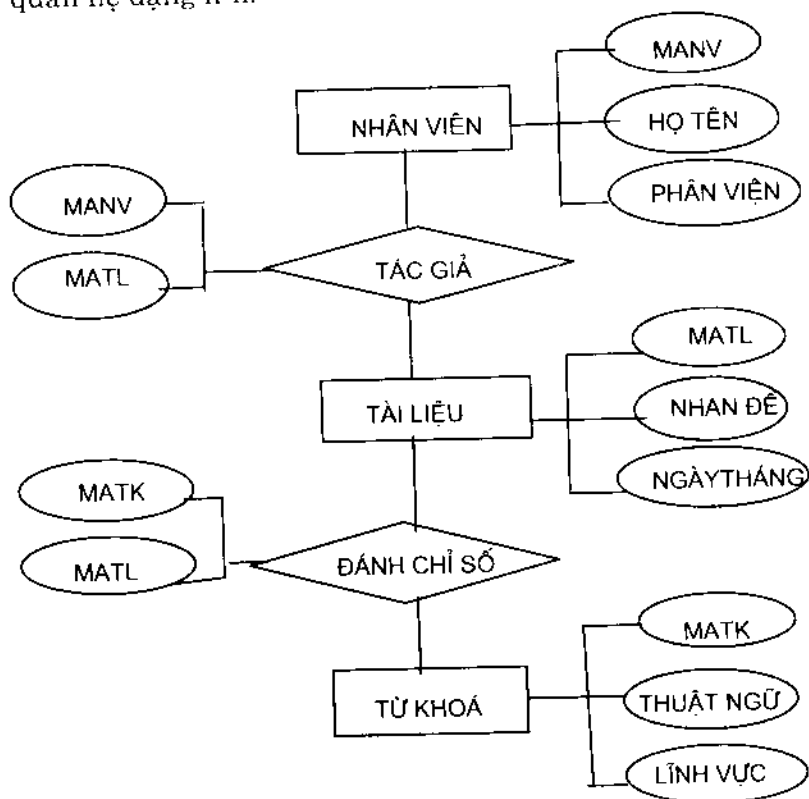
- Thực thể tài liệu, có các thuộc tính: MATL (mã tài liệu), nhan đề, ngày tháng.

- Thực thể từ khoá, có các thuộc tính: MATK (mã từ khoá), thuật ngữ, lĩnh vực.

Vì nhân viên là tác giả của tài liệu và tài liệu được mô tả nội dung bằng các từ khoá, nên giữa các thực thể trên có các quan hệ:

- Quan hệ Tác giả, liên kết hai thực thể nhân viên và tài liệu bằng các thuộc tính MANV và MATL. Quan hệ này là quan hệ dạng n-n.

- Quan hệ Đánh chỉ số, liên kết hai thực thể tài liệu và từ khoá bằng các thuộc tính MATL và MATK. Quan hệ này cũng là quan hệ dạng n-n.



Hình 17. Mô hình ER của CSDL CTKH

Bước 3: Xây dựng mô hình ER cho CSDL CTKH. Sử dụng các ký hiệu đồ họa ta có thể trình bày mô hình ER của CSDL như ở hình 16.

Từ mô hình khái niệm trên ta thấy rằng có thể xây dựng 3 biểu ghi logic cho các thực thể (nhân viên, tài liệu, từ khóa) và 2 biểu ghi logic cho các quan hệ (TÁC GIẢ, ĐÁNH CHỈ SỐ), với 9 trường dữ liệu (MANV, HỌ TÊN, PHÂN VIỆN, MATL, NHAN ĐỀ, NGÀY THÁNG, MATK, THUẬT NGỮ, LĨNH VỰC). Các biểu ghi này có cấu trúc như sau:

NHÂN VIÊN

MANV	HỌ TÊN	PHÂN VIỆN
------	--------	-----------

TÀI LIỆU

MATL	NHAN ĐỀ	NGÀY THÁNG
------	---------	------------

TỪ KHOA

MATK	THUẬT NGỮ	LĨNH VỰC
------	-----------	----------

TÁC GIẢ

MANV	MATL
------	------

ĐÁNH CHỈ SỐ

MATK	MATL
------	------

Tập hợp các biểu ghi trên tạo thành *lược đồ khái niệm* của cơ sở dữ liệu CTKH. Đó là cái khung của CSDL, nó cho ta cách biểu diễn các giá trị trong CSDL và mối quan hệ giữa chúng. Ví dụ dưới đây là một thể hiện của CSDL CTKH tại một thời điểm.

NHÂN VIÊN

MANV	HỌ TÊN	PHÂN VIỆN
NV12	Nguyễn Văn Hải	Ngôn ngữ
NV15	Bùi Văn Nam	Công nghệ thông tin
NV18	Trần Thu Hà	Thông tin-tư liệu

TÀI LIỆU

MATL	NHÂN ĐỀ	NGÀY THÁNG
TL118	Vấn đề chuẩn hoá tiếng Việt trong tin học	16-11-1994
TL123	Mô hình hoá quá trình lưu trữ và tìm kiếm thông tin	25-02-1995

TỪ KHOẢ

MATK	THUẬT NGỮ	LĨNH VỰC
TK112	Chuẩn hoá	Công nghệ thông tin
TK125	Tin học	Công nghệ thông tin
TK134	Tiếng Việt	Ngôn ngữ
TK138	Lưu trữ thông tin	Thông tin học
TK145	Tìm tin	Thông tin học
TK149	Mô hình hoá	Thông tin học

TÁC GIẢ

MANV	MATL
NV15	TL118
NV18	TL123

ĐÁNH CHỈ SỐ

MATK	MATL
TK112	TL118
TK125	TL118
TK134	TL118
TK138	TL123
TK145	TL123
TK149	TL123

Từ các dữ liệu trên ta thấy rằng ở thời điểm đang xét, trong học viện có nghiên cứu viên Trần Thu Hà ở phân viện Thông tin-Tư liệu, là tác giả của tài liệu nhan đề "Mô hình hoá quá trình lưu trữ và tìm kiếm thông tin", công bố ngày 25-02-1995 và nội dung được mô tả bởi các từ khoá: Lưu trữ thông tin, tìm tin, mô hình hoá. Đây là đề tài thuộc lĩnh vực Thông tin học. Ta cũng có những thông tin tương tự với nghiên cứu viên Bùi Văn Nam. Còn nghiên cứu viên Nguyễn Văn Hải chưa có kết quả nghiên cứu nào.

Thông qua ví dụ trên đây, chúng ta đã làm quen với phương pháp thiết kế khái niệm CSDL bằng mô hình thực thể quan hệ. Sản phẩm của thiết kế khái niệm là mô hình khái niệm. Từ mô

hình khái niệm người ta tiếp tục xây dựng mô hình logic cho CSDL. Mô hình logic cung cấp một mô tả dữ liệu sao cho chúng có thể xử lý bằng máy tính điện tử, tức là mô tả cấu trúc của CSDL mà hệ quản trị CSDL quản lý. Từ mô hình logic đã xây dựng ta có thể cài đặt CSDL trên bộ nhớ của máy tính điện tử bằng cách sử dụng một hệ quản trị CSDL thích hợp. Và khi đó một CSDL được hình thành. Có thể nói thiết kế CSDL là quá trình xây dựng cấu trúc cho CSDL. Đó là một khâu quan trọng trong vòng đời của mọi hệ thống thông tin.

2. MÔ HÌNH QUAN HỆ

Mô hình quan hệ là mô hình dữ liệu lần đầu tiên được E.Codd - một nhân viên của hãng IBM - đề xuất năm 1970. Ý tưởng của E.Codd là sử dụng lý thuyết toán học của các quan hệ để cấu trúc hóa các dữ liệu. Từ đó ý tưởng này là đối tượng của nhiều nghiên cứu lý thuyết.

2.1. Các khái niệm cơ bản

Mô hình quan hệ được xây dựng trên cơ sở lý thuyết quan hệ của toán học. Trong toán học người ta định nghĩa quan hệ là tập con của tích Đề các của các tập hợp.

Giả sử $D_1, D_2, \dots, D_n$ là các tập hợp các đối tượng. Tích Đề các của n tập này, ký hiệu $D_1 \times D_2 \times \dots \times D_n$ là tập các bộ- n $(d_1, d_2, \dots, d_n)$ sao cho $d_1 \in D_1, d_2 \in D_2, \dots, d_n \in D_n$.

Một quan hệ R trên $D_1, D_2, \dots, D_n$ là một tập con của tích Đề các $D_1 \times D_2 \times \dots \times D_n$.

Ví dụ:

$D_1 = \{ \text{Vũ Hoài Nam, Nguyễn Văn Hải, Đặng Quốc Hùng, Lê Thu Lan} \}$

$$D_2 = \{ 12-4-1965, 25-8-1962, 7-10-1970, 16-5-1968 \}$$

$$D_3 = \{ \text{Thái Bình, Hà Bắc, Hà Nội} \}$$

$$D_4 = \{ \text{Ngôn ngữ học, Tin học, Thông tin học} \}$$

$R = \{ (\text{Đặng Quốc Hùng}, 25-8-1962, \text{Hà Bắc, Thông tin học}), (\text{Lê Thu Lan}, 16-5-1968, \text{Hà Nội, Ngôn ngữ học}), (\text{Vũ Hoài Nam}, 7-10-1970, \text{Thái Bình, Thông tin học}), (\text{Nguyễn Văn Hải}, 12-4-1965, \text{Hà Bắc, Tin học}) \}$

Quan hệ trên có thể trình bày một cách thuận tiện dưới dạng một bảng. Ở đó giá trị của cùng một tập hợp được ghi trong một cột và mỗi tình huống của quan hệ được ghi trên một hàng. Nếu gọi quan hệ này là quan hệ **GIẢNG VIÊN** thì ta có thể trình bày như sau:

GIẢNG VIÊN

HỌ TÊN	NGÀY SINH	QUÊ QUÁN	CHUYÊN MÔN
Đặng Quốc Hùng	25-8-1962	Hà Bắc	Thông tin học
Lê Thu Lan	16-5-1968	Hà Nội	Ngôn ngữ học
Vũ Hoài Nam	7-10-1970	Thái Bình	Thông tin học
Nguyễn Văn Hải	12-4-1965	Hà Bắc	Tin học

Một hệ thông tin xử lý các đối tượng và các sự kiện của một thế giới thực gọi là thực thể. Một thực thể thường được mô tả bằng các thuộc tính (attribut). Các thông tin liên quan đến một thực thể được biểu diễn dưới dạng các giá trị về lượng hoặc về chất của các thuộc tính của chúng. Tập hợp các giá trị của một thuộc tính được gọi là miền (domaine).

Ví dụ: Thực thể giảng viên có thể bao gồm các thuộc tính: số chứng minh, họ tên, ngày sinh, quê quán, chuyên môn, mức lương...

Trong các thuộc tính của một thực thể có thể có một hay một vài thuộc tính gắn liền với sự tồn tại của thực thể. Những thuộc tính đó gọi là *thuộc tính khóa*.

Ví dụ: Trong thực thể giảng viên các thuộc tính số chứng minh, họ tên, ngày sinh là thuộc tính khóa.

Bây giờ ta có thể định nghĩa quan hệ một cách hình thức như sau:

Định nghĩa: Giả sử $R = \{ A_1, A_2, \dots, A_n \}$ là tập hợp hữu hạn các thuộc tính của thực thể, mỗi thuộc tính A_i có một miền trị tương ứng là $\text{dom}(A_i)$. Quan hệ r trên tập thuộc tính R là tập con của tích Đề các $\text{dom}(A_1) \times \text{dom}(A_2) \times \dots \times \text{dom}(A_n)$:

$$r \subset \text{dom}(A_1) \times \text{dom}(A_2) \times \dots \times \text{dom}(A_n)$$

Khi đó ta ký hiệu $r(R)$ hoặc $r(A_1, A_2, \dots, A_n)$.

Các phần tử $t_1, t_2, \dots, t_p \in r(A_1, A_2, \dots, A_n)$ được gọi là các bộ.

Ví dụ: Quan hệ

BANDOC (SỐTHẺ, HỌTÊN, NĂMSINH, NGHỀNGHIỆP, CƠQUAN)

là một quan hệ giữa các thuộc tính: số thẻ, họ tên, năm sinh, nghề nghiệp, cơ quan.

Một bộ t của quan hệ **BANDOC** có thể viết dưới dạng:

$$t = (001, \text{Đỗ Cao Thắng}, 1965, \text{Bác sĩ}, \text{Bộ Năng lượng})$$

Về mặt trực quan quan hệ trên có thể biểu diễn bằng một bảng hai chiều, ở đó các cột là các thuộc tính, các dòng là các bộ.

BẠN ĐỌC

SỐ THẺ	HỌ TÊN	NĂM SINH	NGHỀ NGHIỆP	CƠ QUAN
001	Đỗ Cao Thắng	1965	Bác sĩ	Bộ Năng lượng
002	Nguyễn Ngọc Hà	1960	Giáo viên	ĐH Văn hoá
003	Lê Thu Trang	1954	Kỹ sư	Bộ Nông nghiệp
...	...	...	...	...

Các ký hiệu: Trong một mô hình quan hệ các thuộc tính thường được ký hiệu là A, B, C,... Tập các thuộc tính được ký hiệu là X, Y, Z... Các thuộc tính trong một tập được liệt kê như một xâu ký tự. Thay cho cách viết $X = \{A, B, C\}$ ta viết $X = ABC$. Các bộ được ký hiệu là t, u, v... Giả sử t là một bộ của quan hệ $r(R)$ và $X \subset R$. Ta ký hiệu $t(X) = t \mid_x$ là hạn chế của t trên X.

Cơ sở dữ liệu quan hệ là tập hợp các quan hệ biến thiên theo thời gian. Ở đây quan hệ là hình thức mô tả dữ liệu.

Để tạo lập một CSDL quan hệ ta phải khai báo tên các quan hệ và các thuộc tính cùng với kiểu của chúng, đồng thời quy định về các ràng buộc dữ liệu giữa các thuộc tính.

Lấy lại ví dụ ở mục 1, với mỗi loại biểu ghi NHÂN VIÊN, TÀI LIỆU, TỪ KHÓA ta có thể thiết lập một lược đồ quan hệ tương ứng. Chẳng hạn ta có thể coi NHÂN VIÊN là quan hệ giữa các thuộc tính MANV, HỌ TÊN, PHÂN VIỆN và ta có quan hệ:

NHÂN VIÊN (MANV, HỌ TÊN, PHÂN VIỆN)

Tương tự ta có các quan hệ:

TÀI LIỆU (MATL, NHAN ĐỀ, NGÀY THÁNG)

TỪ KHÓA (MATK, THUẬT NGỮ, LĨNH VỰC)

Các mối liên kết n- n giữa NHÂN VIÊN -TÀI LIỆU và TỪ KHÓA - TÀI LIỆU cũng có thể biểu diễn bằng các quan hệ tác giả và Đánh chỉ số như sau:

TÁC GIẢ (MANV, MATL)

ĐÁNH CHỈ SỐ(MATK, MATL)

Ở đây quan hệ TÁC GIẢ (MANV, MATL) có ý nghĩa là: nhà nghiên cứu có mã số MANV là một trong các tác giả của tài liệu có mã số MATL.

Còn quan hệ ĐÁNH CHỈ SỐ(MATK, MATL) có nghĩa là: từ khóa có mã số MATK mô tả nội dung của tài liệu có mã số MATL.

Như vậy CSDL CTKH được xây dựng từ năm quan hệ sau:

NHÂN VIÊN (MANV, HỌTÊN, PHÂNVIỆN)

TÀI LIỆU (MATL, NHAN ĐỀ, NGÀYTHÁNG)

TỪ KHÓA (MATK, THUẬTNGỮ, LĨNHVỰC)

TÁC GIẢ (MANV, MATL)

ĐÁNH CHỈ SỐ (MATK, MATL)

Khi thiết kế CSDL bằng mô hình quan hệ, ta tránh đưa tất cả các thuộc tính của đối tượng cần quản lý vào một quan hệ duy nhất vì khi đó quan hệ sẽ chồng chéo và sẽ gây khó khăn cho việc xử lý dữ liệu như sẽ được phân tích trong ví dụ sau đây.

Giả sử một CSDL có tên là CÔNGTY lưu trữ thông tin về công việc của các nhân viên trong một công ty kinh doanh được xây dựng từ quan hệ sau:

MÃNV	HỌTÊN	ĐỊA CHỈ	MÃCV
A15	Lê Huy	12 Hàng Đào	F5
A15	Lê Huy	12 Hàng Đào	D7
B70	Vũ Hùng	54 Hàng Bông	F5
C34	Trần Thu Nga	28 Hàng Gai	S25X
C34	Trần Thu Nga	28 Hàng Gai	S26Z
...	...	...	...

TÊN CV	ĐƠN VỊ	NGÀY KD	NGÀY KT
Bán hàng	Kinh doanh	1-9-92	30-9-93
Bảo vệ	Kinh doanh	1-10-93	30-9-95
Bán hàng	Kinh doanh	1-10-93	30-9-95
Thư ký	Tổ chức	1-3-92	30-4-93
Thủ quỹ	Tài vụ	1-5-93	30-4-94
...	...	...	...

Ta nhận thấy quan hệ trên đây khá rườm rà. Thông tin về cá nhân Lê Huy và Trần Thu Nga lặp lại vì họ được giao hai công việc khác nhau ở hai thời gian khác nhau. Thông tin về công việc bán hàng cũng lặp lại vì công việc đó cần nhiều người đảm nhận. Một khó khăn khác ta gặp phải với quan hệ trên là khi ta phải xoá bỏ các thông tin từ CSDL. Ví dụ, giả sử Lê Huy là nhân viên duy nhất đảm nhiệm công việc có mã số D7. Nếu anh ta rời khỏi công ty và bị xoá khỏi CSDL thì ta sẽ bị mất thông tin về công việc D7.

Nguyên nhân của những điều phiền phức trên đây là do ta đã đưa các thông tin của nhiều thực thể vào một quan hệ duy nhất. Thật vậy, quan hệ trên đã chứa các thông tin về NHÂN VIÊN (mã nhân viên - MANV, họ tên - HỌTÊN, địa chỉ - ĐỊA CHỈ), các thông tin về CÔNG VIỆC (mã công việc - MACV, tên công việc- TÊNCV, đơn vị quản lý công việc - ĐƠNVI) và cả những thông tin về mối quan hệ giữa NHÂN VIÊN và CÔNG VIỆC (ngày khởi đầu công việc - NGÀYKD, ngày kết thúc công việc - NGÀYKT).

Để tránh những phiền phức trên đây ta cần thiết kế lại mô hình logic của CSDL bằng cách tách quan hệ trên thành ba quan hệ sau:

NHÂN VIÊN

MÃNV	HỌTÊN	ĐỊA CHỈ
A15	Lê Huy	12 Hàng Đào
B70	Vũ Hùng	54 Hàng Bông
C34	Trần Thu Nga	28 Hàng Gai
...	...	...

CÔNG VIỆC

MÃCV	TÊNCV	ĐƠNVI
F5	Bán hàng	Kinh doanh
D7	Bảo vệ	Kinh doanh
S25X	Thư ký	Tổ chức
S26Z	Thủ quỹ	Tài vụ
...	...	...

PHÂN CÔNG

MÃNV	MÃCV	NGÀYKĐ	NGÀYKT
A15	F5	1-9-92	30-9-93
A15	D7	1-10-93	30-9-95
B70	F5	1-10-93	30-9-95
C34	S25X	1-3-92	30-4-93
C34	S26Z	1-5-93	30-4-94
...	...	...	...

Như vậy, CSDL CÔNGTY bao gồm ba quan hệ chứa các thông tin về nhân viên thông qua quan hệ NHÂN VIÊN, các thông tin về công việc thông qua quan hệ CÔNG VIỆC và mối quan hệ giữa NHÂN VIÊN và CÔNG VIỆC được thể hiện bởi quan hệ PHÂN CÔNG. Khi cần các thông tin bổ sung ta chỉ cần kết hợp các quan hệ thích hợp bằng phép kết nối của đại số quan hệ mà ta sẽ trình bày ở mục dưới đây.

2.2. Đại số quan hệ

Đại số quan hệ là một trong những ngôn ngữ truy nhập dữ liệu. Đại số quan hệ bao gồm các phép toán trên các quan hệ của một CSDL cho trước. Nhờ các phép toán này mà ta có thể trích dữ liệu từ một hay một vài quan hệ cho trước của CSDL để tạo thành các quan hệ mới, nhằm trả lời cho một câu hỏi tìm kiếm thông tin.

Ở đây ta chỉ xét những phép toán thường được sử dụng nhiều nhất để tra cứu và cập nhật dữ liệu trong một CSDL quan hệ.

Để minh họa cho việc ứng dụng các phép toán sẽ được trình bày dưới đây, trước hết ta xét một ví dụ về một CSDL quan hệ hình thành từ thực tiễn công tác thư viện. Giả sử một CSDL có tên là THƯ VIỆN được tạo thành từ ba quan hệ sau:

BẠN ĐỌC (ST, HT, NS, NN, CQ)

SÁCH (MS, TG, TS, NXB, NĂM)

MƯỢN (ST, MS, NM, NT)

trong đó các thuộc tính được ký hiệu như sau:

ST = Số thẻ đọc

TG = Tác giả

HT = Họ tên

TS = Tên sách

NS = Năm sinh

NXB = Nhà xuất bản

NN = Nghề nghiệp

NĂM = Năm xuất bản

CQ = Cơ quan công tác

NM = Ngày mượn

MS = Mã sách

NT = Ngày trả

Giả thiết rằng các quan hệ của CSDL nói trên thỏa mãn các ràng buộc sau:

- Mỗi bạn đọc chỉ có một thẻ đọc xác định.
- Mỗi quyển sách chỉ có một mã số.
- Mỗi bạn đọc mượn sách phải trả sách vào một ngày xác định.

Sau đây là các thể hiện của ba quan hệ nói trên tại một thời điểm:

BANDOC

ST	HT	NS	NN	CQ
001	Đỗ Cao Thắng	1965	Bác sĩ	Bộ Năng lượng
002	Nguyễn Ngọc Hà	1960	Giáo viên	Đại học Văn hoá
003	Lê Thu Trang	1954	Kỹ sư	Bộ Nông nghiệp
004	Vũ Quế Anh	1962	Bác sĩ	Viện Đông y
005	Trần Văn Thành	1965	Kỹ sư	Bộ Xây dựng
006	Hoàng Quốc Long	1967	Giáo viên	Đại học Văn hoá
007	Nguyễn Duy Hiền	1961	Công nhân	Nhà máy dệt 8-3
008	Trần Hải Châu	1968	Giáo viên	Đại học Sư phạm
009	Vũ Thu Uyên	1950	Kỹ sư	Bộ Xây dựng
100	Chu Phương Diệp	1956	Công nhân	Nhà máy dệt 8-3
...	...	...	...	...

SÁCH

MS	TS	TG	NXB	NĂM
1	Foxbase	Lê Quang Anh	Đại học	1991
2	Cơ sở dữ liệu	C.J. Date	Thống kê	1986
3	Ngôn ngữ C	Phạm Văn Ất	Đại học	1991
4	Tin học phổ thông	Nguyễn Chí Công	Giáo dục	1990
5	Foxpro ứng dụng	Trương Văn Vân	Khoa học kỹ thuật	1991
6	Word 6 for Windows	Bùi văn Thanh	Giáo dục	1994

7	Lập trình hệ thống	Niklaus Wirth	Khoa học kỹ thuật	1986
8	Kỹ thuật tin học	J.Ristori	Khoa học kỹ thuật	1990
...	...	...	...	...

MUON

ST	MS	NM	NT
003	1	12-3-1995	22-3-1995
007	5	12-3-1995	22-3-1995
004	2	14-3-1995	24-3-1995
004	3	14-3-1995	24-3-1995
009	7	15-3-1995	25-3-1995
001	6	16-3-1995	26-3-1995
...	...	...	...

1. Phép chọn (Selection)

Giả sử t là một bộ trong quan hệ $r(R)$ và E là một biểu thức logic phát biểu trên tập các thuộc tính của quan hệ r . Ta nói bộ t thỏa mãn biểu thức E , ký hiệu $t(E)$, nếu khi thay mỗi thuộc tính A trong E bằng giá trị $t(A)$ ta được một biểu thức logic mệnh đề đúng. E được gọi là biểu thức chọn hay điều kiện chọn.

Ví dụ: Bộ

$t = (001, \text{Đỗ Cao Thắng}, 1965, \text{Bác sĩ}, \text{Bộ Năng lượng})$

trong quan hệ **BAN ĐỌC** thỏa mãn điều kiện chọn:

$$E = NS < 1966 \wedge NN = \text{Bác sĩ}.$$

Định nghĩa: Cho quan hệ $r(R)$ và điều kiện chọn E trên R . Phép chọn quan hệ r theo điều kiện E , ký hiệu là $r(E)$ là phép toán cho ta quan hệ $r'(R)$ bao gồm các bộ của $r(R)$ thỏa mãn điều kiện chọn E .

$$r(E) = r'(R) = \{ t \in r \mid t(E) \}$$

Đặc biệt nếu điều kiện chọn $A = a$, ta có:

$$r(A = a) = r'(R) = \{ t \in r \mid t(A) = a \}$$

Ví dụ: Cho thông tin về những bạn đọc là bác sĩ sinh trước năm 1966 có trong quan hệ BẠN ĐỌC.

Gọi p là quan hệ kết quả ta có:

$$p = \text{BẠN ĐỌC} (NS < 1966 \wedge NN = \text{Bác sĩ}).$$

p

ST	HT	NS	NN	CQ
001	Đỗ Cao Thắng	1965	Bác sĩ	Bộ Năng lượng
004	Vũ Quế Anh	1962	Bác sĩ	Viện Đông y

2. Phép chiếu (Projection)

Định nghĩa: Cho quan hệ $r(R)$ và tập thuộc tính $X \subseteq R$. Phép chiếu quan hệ r trên X , ký hiệu là $r[X]$ là phép toán cho ta quan hệ $r'(X)$ với các bộ là hạn chế trên X của các bộ của r .

$$r[X] = r'(X) = \{ t(X) \mid t \in r \}$$

Ta có thể xác định $r'(X)$ theo hai bước sau:

- Loại bỏ các thuộc tính không có trong X của r .

- Lược bớt các bộ giống nhau, chỉ giữ lại một bộ trong số các bộ giống nhau.

Ví dụ: Cho biết danh mục nghề nghiệp của các bạn đọc đăng ký đọc ở thư viện.

Gọi q là quan hệ kết quả, ta có:

$$q = \text{BAN ĐOC} [\text{NN}]$$

q

NN
Bác sĩ
Giáo viên
Kỹ sư
Công nhân

Ví dụ: Cho biết tên những quyển sách của Nhà xuất bản Giáo dục.

Gọi r là quan hệ kết quả, ta có:

$$r = \text{SACH (NXB = Giáo dục)} [\text{TS}]$$

Biểu thức trên bao gồm hai phép toán quan hệ:

- Trước hết chọn từ quan hệ SACH các bộ có NXB = Giáo dục, ta được:

SACH (NXB = Giáo dục)

MS	TS	TG	NXB	NĂM
4	Tin học phổ thông	Nguyễn Chí Công	Giáo dục	1990
6	Word 6 for Windows	Bùi Văn Thanh	Giáo dục	1994

- Sau đó chiếu trên TS, ta được kết quả sau:

r	
TS	
Tin học phổ thông	
Word 6 for Windows	

3. Phép kết nối (Join)

Định nghĩa: Cho quan hệ $r(R)$ và $s(S)$. Phép kết nối tự nhiên (Natural join) hai quan hệ r và s , ký hiệu là $r * s$ là phép toán cho ta quan hệ $q(RS)$ với tập thuộc tính RS và với các bộ được xác định như sau:

$$r * s = q(RS) = \{ t \mid t(R) \in r \wedge t(S) \in s \}$$

Như vậy phép kết nối tự nhiên kết hợp hai quan hệ theo thuộc tính chung của chúng, tức là thuộc tính trong $R \cap S$.

Ví dụ: Để biết thông tin về tình hình mượn sách của bạn đọc ta kết nối hai quan hệ BẠN ĐỌC và MUỐN theo thuộc tính chung ST (số thẻ đọc). Mỗi bộ u của quan hệ BẠN ĐỌC được nối với bộ v của quan hệ MUỐN theo điều kiện $u(ST) = v(ST)$.

Gọi p là quan hệ kết quả, ta viết:

$$p = \text{BẠN ĐỌC} * \text{MUỐN}$$

p

ST	HT	NS	NN
001	Đỗ Cao Thắng	1965	Bác sĩ
003	Lê Thu Trang	1954	Kỹ sư
004	Vũ Quế Anh	1962	Bác sĩ
004	Vũ Quế Anh	1962	Bác sĩ
007	Nguyễn Duy Hiến	1961	Công nhân
009	Vũ Thu Uyên	1950	Kỹ sư

CQ	MS	NM	NT
Bộ Năng lượng	6	16-3-1995	26-3-1995
Bộ Nông nghiệp	1	12-3-1995	22-3-1995
Viện Đông y	2	14-3-1995	24-3-1995
Viện Đông y	3	14-3-1995	24-3-1995
Nhà máy dệt 8-3	5	12-3-1995	22-3-1995
Bộ Xây dựng	7	15-3-1995	25-3-1995

Từ kết quả trên ta có thể tìm được những thông tin chi tiết hơn về tình hình mượn sách của bạn đọc.

Ví dụ: Cho biết họ tên và cơ quan công tác của các bạn đọc phải trả sách trước ngày 25 tháng 3 năm 1995.

Gọi q là quan hệ kết quả, ta có:

$$q = \text{BAN ĐOC} * \text{MUON} (\text{NT} < 25-3-1995) [\text{HT}, \text{CQ}]$$

q

HT	CQ
Lê Thu Trang	Bộ Nông nghiệp
Vũ Quế Anh	Viện Đông y
Nguyễn Duy Hiên	Nhà máy dệt 8-3

2.3. Tối ưu hoá câu hỏi

Một câu hỏi tìm kiếm thông tin trong CSDL thường được phát biểu dưới dạng một biểu thức quan hệ T. Đó là một biểu thức được xây dựng từ các phép toán quan hệ với toán hạng là các quan hệ trong một CSDL cho trước.

Ví dụ: Cho biết tên bạn đọc là kỹ sư và tên quyển sách mà họ mượn phải trả trước ngày 25-3-1995.

Biểu thức quan hệ biểu diễn câu hỏi trên có thể là:

$T = \text{BẠNĐỌC} * \text{SÁCH} * \text{MƯỢN} (NN = \text{Kỹ sư} \wedge NT < 25-3-1995) [HT, TS]$

Một hệ quản trị CSDL trong trường hợp được cài đặt phép toán quan hệ thường được trang bị các chức năng sau:

- Kiểm tra cú pháp của biểu thức hỏi T.
- Tổ chức tối ưu hóa việc thực hiện T sao cho tốn ít miền nhớ và thời gian thực hiện ngắn nhất.

Ta biết rằng các quan hệ càng nhỏ thì chi phí thời gian để thực hiện các phép toán càng thấp và càng tiết kiệm được miền nhớ. Quan hệ có thể nhỏ theo hai kích thước: hẹp ngang (ít thuộc tính) và ngắn (ít bộ).

Trong các phép toán quan hệ thì phép chiếu và phép chọn có khả năng làm giảm kích thước ngang và dọc của quan hệ còn phép kết nối thường làm tăng kích thước của quan hệ. Vì thế muốn tối ưu hóa câu hỏi, tức là thực hiện câu hỏi với thời gian ngắn nhất và tốn ít miền nhớ nhất, ta cần tìm cách chuyển đổi vị trí các phép toán sao cho có thể thực hiện phép chọn và phép chiếu ở mức sớm nhất.

Quá trình thực hiện biểu thức T trên đây được thể hiện dưới dạng cây, gọi là cây thực hiện. Cây thực hiện được mô tả ở Hình 17.

Các phép toán trong biểu thức quan hệ T được thực hiện theo trình tự sau:

(1)- Kết nối hai quan hệ BẠNĐỌC và SÁCH được quan hệ T_1 .

(2)- Kết nối T_1 với quan hệ MUỖN ta được quan hệ T_2

(3)- Chọn các bộ trong T_2 thỏa biểu thức chọn:

$$E = (NN = \text{Kỹ sư} \wedge NT < 25-3-1995)$$

ta được quan hệ T_3

(4)- Chiếu T_3 trên các thuộc tính HT và TS để tìm họ tên và tên sách. Kết quả cuối cùng phải được gán trở lại cho biểu thức ban đầu T.

Có thể viết thứ tự thực hiện đó như sau:

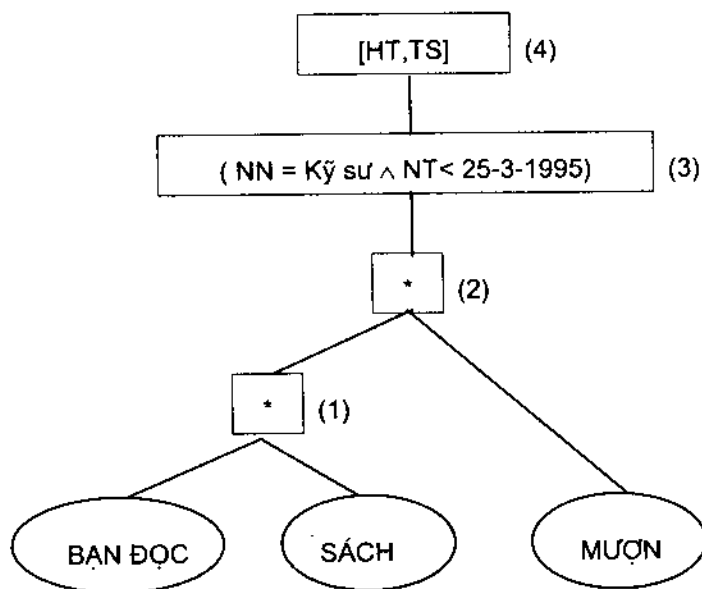
$$(1)- T_1 = \text{BẠNĐỌC} * \text{SÁCH}$$

$$(2)- T_2 = T_1 * \text{MUỖN}$$

$$(3)- T_3 = T_2 (NN = \text{Kỹ sư} \wedge NT < 25-3-1995)$$

$$(4)- T = T_3 [\text{HT}, \text{TS}].$$

Quá trình đó có thể trình bày bằng sơ đồ ở hình 17, gọi là cây biểu thức.



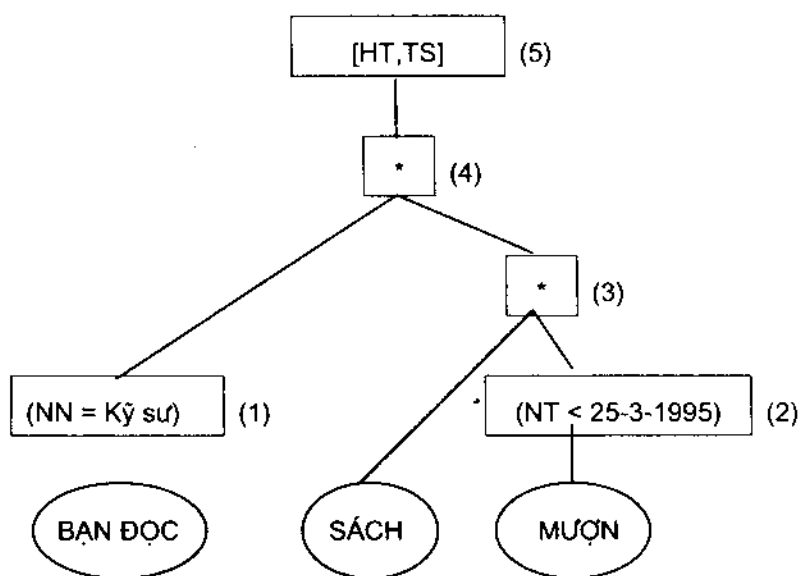
Hình 18. Cây biểu thức được thực hiện từ lá đến gốc

Để ý rằng NN là thuộc tính của quan hệ BẠN ĐỌC, NT là thuộc tính của quan hệ MUỖN nên có thể đưa điều kiện chọn (NN = Kỹ sư) vào quan hệ BẠN ĐỌC, còn điều kiện (NT < 25-3-1995) có thể đưa vào quan hệ MUỖN. Khi đó cây biểu thức được mô tả như hình 18.

Biểu thức quan hệ nói trên bây giờ có dạng:

$T_1 = \text{BẠN ĐỌC} (\text{NN} = \text{Kỹ sư}) * (\text{SÁCH} * \text{MUỖN} (\text{NT} < 25-3-1995)) [\text{HT}, \text{TS}]$.

Đó là biểu thức quan hệ tối ưu của câu hỏi tìm tin trên.



Hình 19. Cây biểu thức đã tối ưu hoá

Trình tự thực hiện biểu thức quan hệ trên là:

- (1)- $T_2 = \text{BẠN ĐỌC} (NN = \text{Kỹ sư})$
- (2)- $T_3 = \text{MƯỢN} (NT < 25-3-1995)$
- (3)- $T_4 = \text{SÁCH} * T_3$
- (4)- $T_5 = T_2 * T_4$
- (5)- $T_1 = T_5 [HT, TS]$

Với CSDL đang xét, trình tự thực hiện đó được thể hiện cụ thể như sau:

(1)- Chọn từ quan hệ **BẠN ĐỌC** những bộ thỏa mãn điều kiện chọn ($NN = \text{Kỹ sư}$), ta được quan hệ:

$T_2 = \text{BANDOC (NN = Kỹ sư)}$

ST	HT	NS	NN	CQ
003	Lê Thu Trang	1954	Kỹ sư	Bộ Nông nghiệp
005	Trần Văn Thành	1965	Kỹ sư	Bộ Xây dựng
009	Vũ Thu Uyên	1950	Kỹ sư	Bộ Xây dựng

(2)- Chọn từ quan hệ MUON những bộ thỏa mãn điều kiện ($NT < 25-3-1995$), ta được quan hệ:

$T_3 = \text{MUON (NT < 25-3-1995)}$

ST	MS	NM	NT
003	1	12-3-1995	22-3-1995
007	5	12-3-1995	22-3-1995
004	2	14-3-1995	24-3-1995
004	3	14-3-1995	24-3-1995

(3)- Kết nối quan hệ SÁCH với quan hệ T_3 , theo thuộc tính chung $MS = \{1,5,2,3\}$, ta được quan hệ:

$T_4 = \text{SACH} * T_3$

MS	TS	TG	NXB
1	Foxbase	Nguyễn Quang Anh	Đại học
5	Foxpro ứng dụng	Trương Văn Vân	Khoa học kỹ thuật
2	Cơ sở dữ liệu	C.J. Date	Thống kê
3	Ngôn ngữ C	Phạm Văn Ất	Đại học

NĂM	ST	NM	NT
1991	003	12-3-1995	22-3-1995
1991	007	12-3-1995	22-3-1995
1986	004	14-3-1995	24-3-1995
1991	004	14-3-1995	24-3-1995

(4)- Kết nối quan hệ T_2 với T_4 theo thuộc tính chung (ST = 003), ta được quan hệ $T_5 = T_2 * T_4$, quan hệ này chỉ có một bộ, ứng với bạn đọc có số thẻ là 003, mượn cuốn sách có MS là 1, ngày trả là 22-3-1995.

(5)- Thực hiện phép chiếu quan hệ T_5 trên các thuộc tính HT và TS để tìm họ tên và tên sách, ta được kết quả:

$$T_1 = T_5[HT, TS]$$

HT	TS
Lê Thu Trang	Foxbase

Vậy bạn đọc Lê Thu Trang là kỹ sư, mượn cuốn sách Foxbase, phải trả sách trước ngày 25-3-1995. Đó là câu trả lời cho yêu cầu tìm tin đặt ra.

Chương 7

HỆ QUẢN TRỊ CƠ SỞ DỮ LIỆU CDS/ISIS FOR WINDOWS

1. TỔNG QUAN VỀ HỆ THỐNG LƯU TRỮ VÀ TÌM KIẾM THÔNG TIN CDS-ISIS FOR WINDOWS

CDS/ISIS (Computer Documentation System - Integrated Set of Information System) là phần mềm tư liệu do UNESCO phát triển và phổ biến từ năm 1985. CDS/ISIS quản lý các CSDL dạng văn bản có cấu trúc. Nó có khả năng quản lý các dữ liệu có độ dài biến động, quản lý các thông tin có tính chất lặp (trường lặp), quản lý các thông tin có nhiều phần tử khác nhau nhưng cùng ý nghĩa (trường con). Nó không quản lý số liệu và không cho phép thực hiện các phép tính và thống kê.

CDS/ISIS là một trong những phần mềm tư liệu được sử dụng rộng rãi trên thế giới đặc biệt là trong các nước đang phát triển, trong đó có Việt Nam. CDS/ISIS đã được Trung tâm Thông tin - Tư liệu Khoa học và Công nghệ Quốc gia Việt hoá để sử dụng với bộ mã chuẩn quốc gia.

CDS/ISIS for Windows, gọi tắt là WinISIS, là sự phát triển của chương trình CDS/ISIS chạy trên hệ điều hành DOS. Phần trình bày dưới đây giới thiệu version 1.31 của bộ chương trình này do UNESCO phát hành năm 1999.

1.1. Chức năng của WinISIS

WinISIS có các chức năng sau đây:

- Tạo lập và sửa đổi cấu trúc của CSDL.
- Nhập và sửa đổi các biểu ghi.
- Trình bày biểu ghi hay một phần biểu ghi theo yêu cầu.
- Tự động xây dựng và duy trì các tệp đảo cho phép truy nhập nhanh CSDL.
- Tìm kiếm các biểu ghi theo nhiều dấu hiệu khác nhau thông qua ngôn ngữ tìm rất mềm dẻo và linh hoạt.
- In và sắp xếp các kết quả tìm theo trật tự tùy ý.
- Trao đổi dữ liệu.
- Phát triển các ứng dụng đặc thù nhờ phương tiện lập trình tích hợp.

Giới hạn khả năng của phần mềm WinISIS 1.31 được thể hiện trong những đặc điểm sau:

1- Số lượng các CSDL quản lý được	không hạn chế
2- Số lượng biểu ghi tối đa trong 1 CSDL	16 triệu (với bộ nhớ 500MB)
3- Kích thước tối đa của một biểu ghi	32.000 ký tự
4- Kích thước tối đa của một trường	32.000 ký tự
5- Số trường tối đa trong một CSDL	200 trường
6- Độ lớn tối đa của một format trình bày	10.000 ký tự
7- Số dòng tối đa trong FST	600 dòng

So với CDS.ISIS chạy trên DOS, WinISIS 1.31 có những ưu điểm sau:

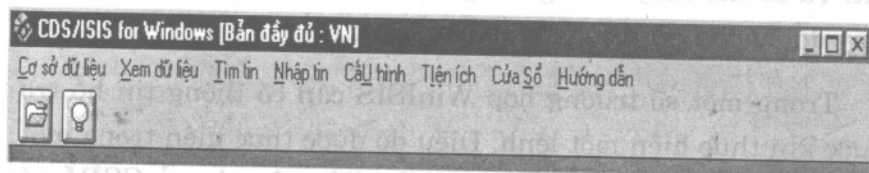
- Việc xây dựng cấu trúc của CSDL thuận tiện và đơn giản hơn nhờ có các phương tiện trợ giúp để tạo biểu mẫu nhập tin và format hiện hình.
- Việc tìm tin đơn giản thuận tiện hơn nhờ giao diện kiểu cửa sổ và các phương tiện tìm được thực hiện trên các nút chức năng.
- Hiệu quả tìm tin tăng rõ rệt nhờ các toán tử tìm thể hiện trên các nút; cửa sổ từ điển có thể đặt ở vị trí bất kỳ trên màn hình và có thể kéo thuật ngữ tìm từ từ điển vào hộp thuật ngữ tìm; lệnh tìm có thể lưu lại cho các lần tìm sau.
- Có hai kiểu tìm tin: tìm tin ở trình độ cao và tìm tin có trợ giúp.
- Có khả năng kết nối một CSDL này với các CSDL khác để mở rộng các chức năng quản lý, tạo ra hệ thống CSDL tựa quan hệ (Pseudo - Relational Databases).

1.2. Nguyên tắc hoạt động của WinISIS

Nguyên tắc hoạt động của WinISIS thể hiện ở những đặc điểm sau đây:

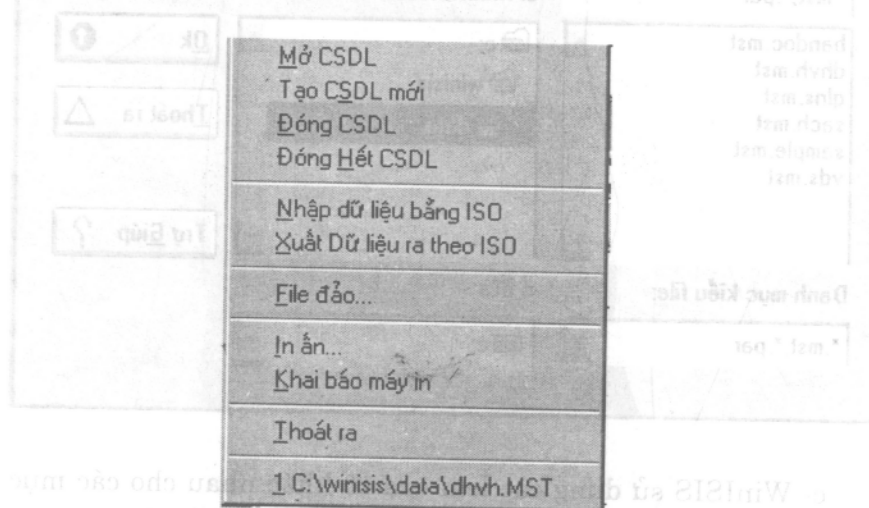
a- Thành phần cơ bản của WinISIS là hệ thống các *menu*. Đó là danh sách các chức năng hay các lệnh mà từ đó người sử dụng có thể chọn bằng cách dùng chuột hay bàn phím để chọn các chức năng hay các lệnh tương ứng trên menu. Menu chính được đưa ra dưới dạng thanh menu, ở dòng trên cùng của cửa sổ chính của hệ thống WinISIS, ngay dưới thanh tiêu đề. Menu

chính của WinISIS gồm các menu con, thể hiện trên một thanh menu như sau:



Để chọn một chức năng, trước hết ta kích chuột vào menu tương ứng, khi đó một danh sách các lệnh được phơi ra. Muốn chọn lệnh nào ta chỉ việc nháy đúp chuột vào lệnh ấy hoặc gõ một lệnh tương ứng trên bàn phím.

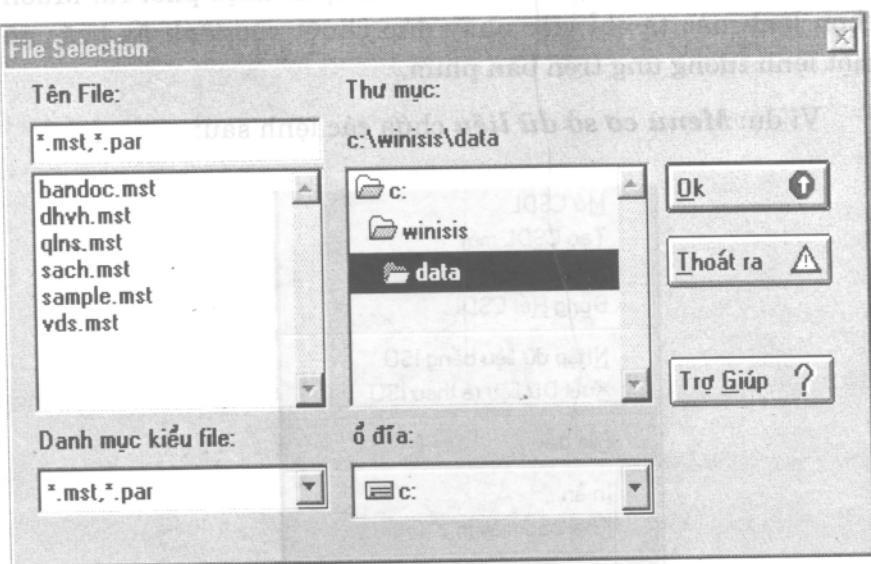
Ví dụ: **Menu cơ sở dữ liệu** chứa các lệnh sau:



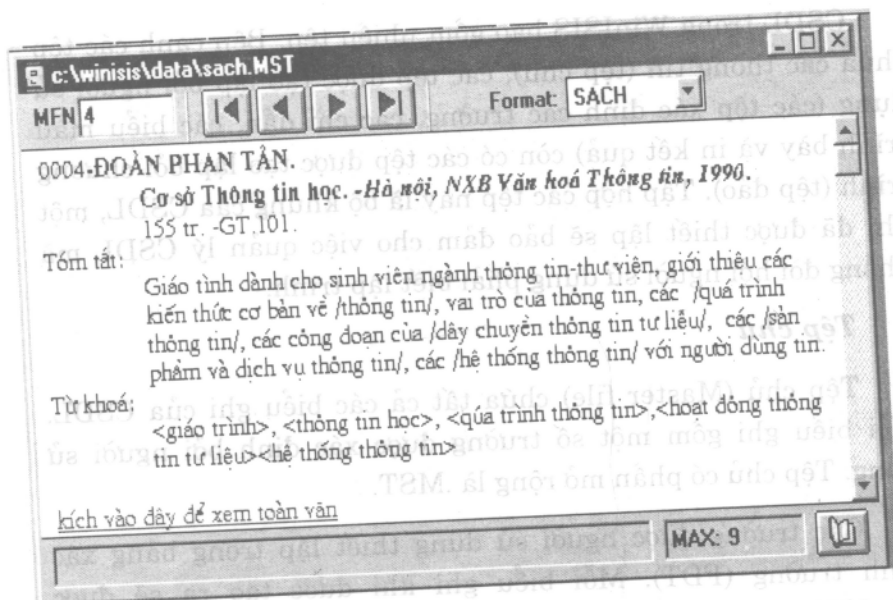
b- Tất cả các chương trình đều làm việc trong chế độ hội thoại và có khả năng đối thoại trên nhiều ngôn ngữ. Nhờ đó ta có thể chọn ngôn ngữ để thể hiện các menu và thông báo hệ thống. Mỗi lần chạy chương trình WinISIS sẽ đưa ra menu bằng

ngôn ngữ định trước, do người sử dụng thiết lập ban đầu khi cài đặt. Ta có thể thay đổi ngôn ngữ đối thoại bằng cách chọn lệnh thay đổi ngôn ngữ trong Menu cấu hình.

Trong một số trường hợp WinISIS cần có thông tin hỗ trợ trước khi thực hiện một lệnh. Điều đó được thực hiện trong một cửa sổ riêng, gọi là hộp thoại. Ví dụ **hộp thoại mở CSDL** có dạng như sau:



c- WinISIS sử dụng các loại cửa sổ khác nhau cho các mục đích khác nhau. Có cửa sổ CSDL để trình bày các biểu ghi, cửa sổ nhập tin để cập nhật dữ liệu, cửa sổ tìm để thực hiện các thao tác tìm tin. Các cửa sổ này được tự động đưa ra trên màn hình tương ứng với chức năng được chọn. Ví dụ: Cửa sổ cơ sở dữ liệu có dạng sau:



1.3. Hệ thống các tệp của cơ sở dữ liệu trong WinISIS

Cấu trúc chung của một tệp của CSDL

Các tệp dữ liệu của CSDL trong WinISIS bao gồm các biểu ghi (Records). Mỗi biểu ghi tương ứng với một đơn vị thông tin. Ví dụ: Một bản mô tả thư mục của một tài liệu, một bản sơ yếu lý lịch của một nhân viên ...

Mỗi biểu ghi bao gồm các trường (Fields). Mỗi trường lưu trữ một phần tử dữ liệu. Ví dụ: Trong một biểu ghi thư mục có các trường: tác giả, nhan đề của tài liệu, nơi xuất bản, nhà xuất bản, năm xuất bản, v.v...

Một đặc điểm của WinISIS là chúng được thiết kế đặc biệt để quản lý các trường có độ dài không cố định.

CSDL trong WinISIS bao gồm nhiều tệp. Bên cạnh các tệp chứa các thông tin (tệp chủ), các tệp được tạo lập bởi người sử dụng (các tệp xác định các trường, các chỉ dẫn, các biểu mẫu trình bày và in kết quả) còn có các tệp được tạo lập bởi chương trình (tệp đảo). Tập hợp các tệp này là bộ khung của CSDL, một khi đã được thiết lập sẽ bảo đảm cho việc quản lý CSDL mà không đòi hỏi người sử dụng phải biết lập trình.

Tệp chủ

Tệp chủ (Master file) chứa tất cả các biểu ghi của CSDL. Mỗi biểu ghi gồm một số trường được xác định bởi người sử dụng. Tệp chủ có phần mở rộng là .MST.

Các trường được người sử dụng thiết lập trong bảng xác định trường (FDT). Mỗi biểu ghi khi được tạo ra sẽ được WinISIS tự động gán cho một số hiệu, gọi là số hiệu biểu ghi tệp chủ (MFN).

Để truy nhập nhanh tới từng biểu ghi trong tệp chủ WinISIS lập ra một tệp đặc biệt có liên quan đến tệp chủ, gọi là tệp tham khảo chéo (Cross reference file) mà thực chất là tệp chứa địa chỉ của các biểu ghi trong tệp chủ. Tệp này có phần mở rộng là .XRF.

Tệp đảo

Để truy nhập nhanh tới các thông tin cần tìm, WinISIS sử dụng các tệp đảo (inverted file). Tệp đảo chứa tất cả các thuật ngữ có thể dùng như những điểm truy nhập thông tin trong CSDL. Mỗi thuật ngữ là một chỉ dẫn đến các biểu ghi trong tệp chủ mà từ đó thuật ngữ được lấy ra. Tập hợp tất cả các điểm truy nhập của CSDL được gọi là *từ điển*. Có thể coi tệp đảo là tệp chỉ dẫn về nội dung của tệp chủ.

Ví dụ, có 4 biểu ghi (với MFN 18, 26, 874, 1205) chứa từ khoá tin học. Cấu trúc logic của biểu ghi tương ứng trong tệp đảo sẽ là:

Tin học 18 26 874 1205

Ở đây tin học là điểm tiếp cận thông tin. Vì mỗi từ khoá có số các biểu ghi khác nhau được đánh chỉ số bởi từ khoá đó nên biểu ghi logic trong tệp đảo có độ dài thay đổi.

Tệp đảo được WinISIS tự động tạo ra. Người sử dụng chỉ cần chọn ra các trường mà từ đó dữ liệu chỉ dẫn được lấy ra để đưa vào tệp đảo. Việc xác định các phần tử tìm cho một CSDL được thực hiện bằng *Bảng chọn trường* (FST), trong đó có chứa các trường cần đảo và kỹ thuật đánh chỉ số cho từng trường.

Khác với các hệ thống tìm xây dựng trên cơ sở sử dụng nhiều tệp đảo riêng rẽ cho mỗi một trường tìm, WinISIS chỉ dùng một tệp đảo duy nhất cho cả CSDL. Tuy nhiên để tăng nhanh tốc độ tìm, tệp đảo này được phân ra thành 6 tệp nhỏ, có phần mở rộng là .L01, .L02, .N01, .N02, .IFP và .CNT.

Có một kiểu tệp không bắt buộc, có liên quan tới tệp đảo, đó là tệp ANY. Tệp ANY cho phép nối một thuật ngữ với các thuật ngữ khác, tùy theo yêu cầu của người sử dụng khi tìm tin. Tệp này có phần mở rộng là .ANY.

Các tệp xác định cấu trúc CSDL

Một CSDL trong WinISIS được xác định bởi các thành phần sau, mỗi thành phần được lưu trữ trong một tệp riêng biệt:

Bảng xác định trường (FDT- Field Definition Table): FDT xác định các trường có thể có trong các biểu ghi của CSDL và

các đặc tính của chúng (nhãn, tên, kiểu, trường lập, trường con...). Tập này là duy nhất với một CSDL, có phần mở rộng là .FDT.

Bảng chọn trường (FST- Field Selection Table): FST chứa danh mục các trường là chỉ dẫn tìm tin trong tệp đảo. Tập này là duy nhất với một CSDL và có phần mở rộng là .FST.

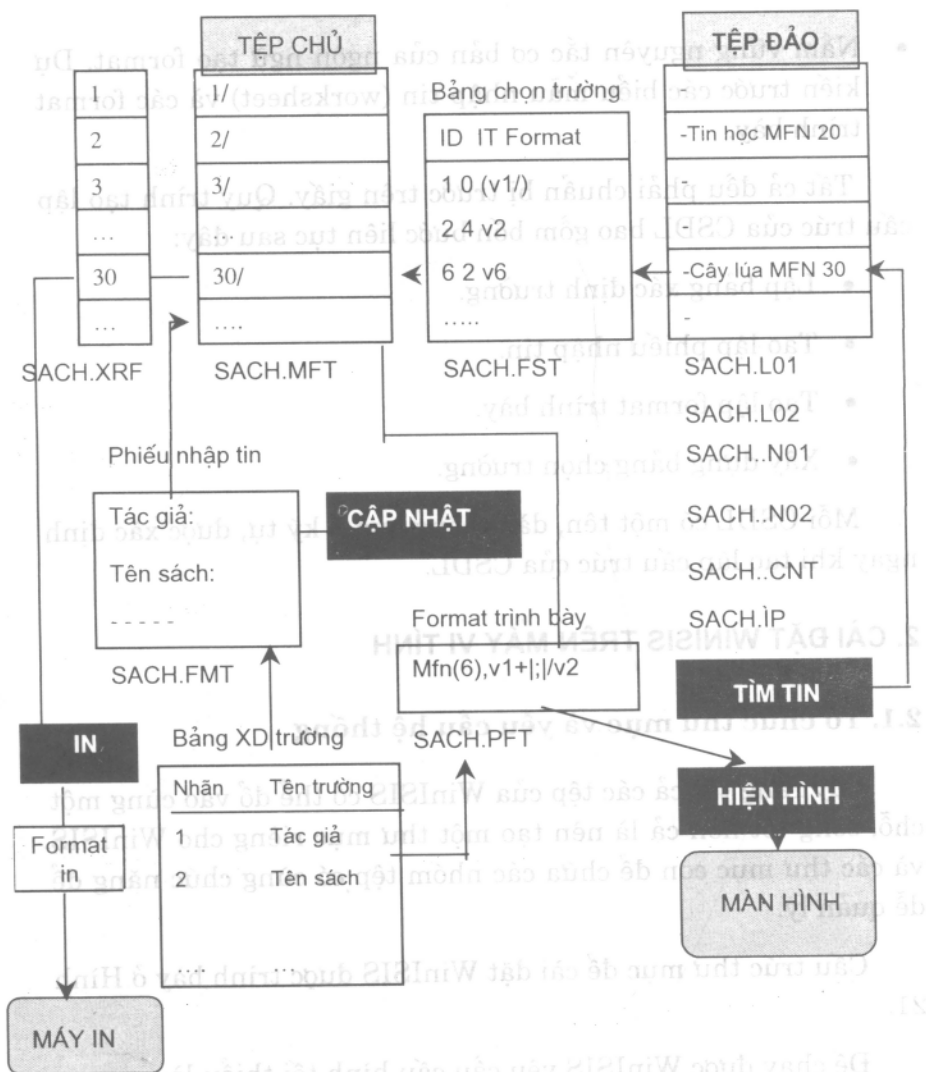
Worksheet nhập dữ liệu: là một hoặc nhiều mẫu màn hình dùng để tạo và cập nhật các biểu ghi của CSDL. Trong thực tế mỗi trang màn hình của worksheet tạo thành một tệp riêng biệt (phần mở rộng là .FMT).

Format trình bày: xác định cách biểu diễn kết quả tìm tin trên màn hình hoặc in các sản phẩm ở đầu ra. WinISIS có một ngôn ngữ tạo format đủ mạnh, cho phép trình bày nội dung biểu ghi theo ý muốn. Một CSDL có thể có nhiều Format trình bày (phần mở rộng là PFT).

Sơ đồ Hình 20 trình bày mối liên hệ giữa các tệp khác nhau của CSDL có tên là SACH.

Để tạo lập CSDL bằng CDS/ISIS người sử dụng trước hết cần nắm vững các thông tin cần xử lý:

- Lập danh mục các trường dữ liệu, xác định nhãn, tên, kiểu dữ liệu (chữ, số, hỗn hợp), kiểu trường (trường đơn, trường lập, trường con).
- Chọn các trường dùng để đảo (đánh chỉ số) và kỹ thuật đảo (có ký số 0,1,2,3,4,5,6,7,8).



Hình 20. Quan hệ giữa các tệp của CSDL trong WinISIS

- Nắm vững nguyên tắc cơ bản của ngôn ngữ tạo format. Dự kiến trước các biểu mẫu nhập tin (worksheet) và các format trình bày.

Tất cả đều phải chuẩn bị trước trên giấy. Quy trình tạo lập cấu trúc của CSDL bao gồm bốn bước liên tục sau đây:

- Lập bảng xác định trường.
- Tạo lập phiếu nhập tin.
- Tạo lập format trình bày.
- Xây dựng bảng chọn trường.

Mỗi CSDL có một tên, dài không quá 6 ký tự, được xác định ngay khi tạo lập cấu trúc của CSDL.

2. CÀI ĐẶT WINISIS TRÊN MÁY VI TÍNH

2.1. Tổ chức thư mục và yêu cầu hệ thống

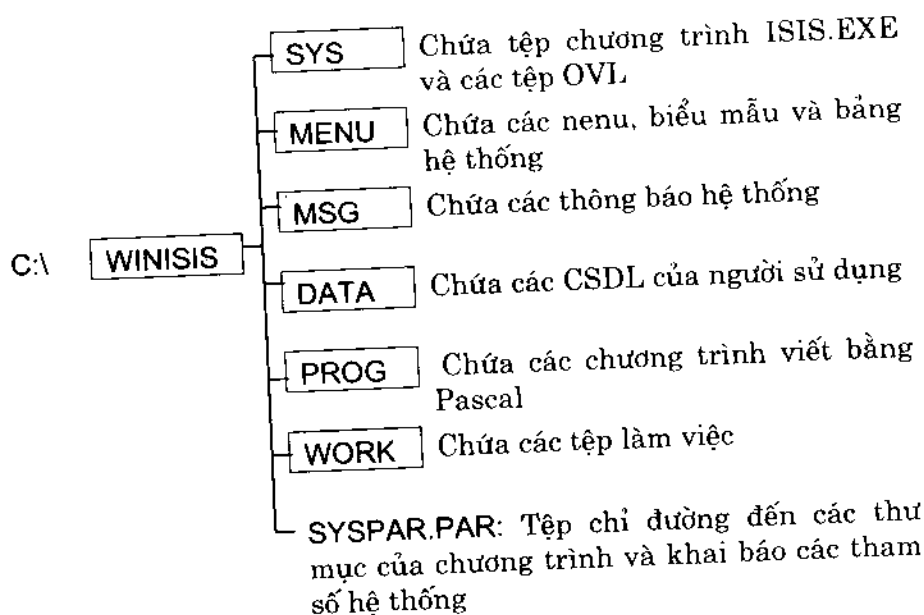
Tuy rằng tất cả các tệp của WinISIS có thể đổ vào cùng một chỗ, song tốt hơn cả là nên tạo một thư mục riêng cho WinISIS và các thư mục con để chứa các nhóm tệp có cùng chức năng để dễ quản lý.

Cấu trúc thư mục để cài đặt WinISIS được trình bày ở Hình 21.

Để chạy được WinISIS yêu cầu cấu hình tối thiểu là:

- CPU: Bộ xử lý 486 40MHz trở lên (tốt hơn là Pentium 100 MHz trở lên).
- RAM: Tối thiểu 8 MB, nếu có 16 MB thì càng tốt.

- 01 ổ đĩa mềm.
- 01 ổ đĩa cứng (tối thiểu còn 4 MB trống).
- 01 màn hình loại VGA với độ phân giải tối thiểu 640×480.
- 01 máy in tùy chọn.
- Hệ điều hành Windows 3.1 trở lên, tốt nhất là Windows 9x trở lên.
- Bộ mã tiếng Việt là bộ mã theo tiêu chuẩn Việt Nam TCVN 5712-1993 (Bộ mã ABC). Bộ gõ có thể dùng bộ gõ ABC hay Vietkey.



Hình 21. Tổ chức thư mục của chương trình WinISIS

2.2. Nội dung của tệp syspar.par

Tệp syspar.par đặt trong thư mục chính của WinISIS, là một tệp ASCII chứa các tham số hệ thống. Mỗi tham số bắt đầu ở dòng mới và có format chung như sau:

n = giá trị

trong đó:

n: Số hiệu của tham số

giá trị: Giá trị tương ứng của tham số

Mục giá trị phải đặt tiếp ngay sau dấu =, không có khoảng trống. Tên thư mục con trong tham số phải kết thúc bằng dấu "\".

Ví dụ: 2=\Winisis\Menu\

Các tham số không khai báo hay các tham số có giá trị là các khoảng trống sẽ được gán bằng các giá trị ngầm định của hệ thống. Vì thế trong tệp SYSPAR.PAR chỉ nên chứa các tham số mà ta cần thay đổi. Các tham số có thể sắp theo thứ tự tùy ý. Các tham số đặt sau dấu chấm phẩy ";" chỉ có ý nghĩa chú giải, không có hiệu lực.

Dưới đây là một số tham số hệ thống chủ yếu.

Tham số 0: Chuyển hướng tệp SYSPAR.PAR sang một tên đĩa hay một thư mục khác.

Ví dụ: 0=c:\Winisis\mypar\

Tham số 1: Đường dẫn tới thư mục chứa chương trình WinISIS Pascal.

Tham số 2: Đường dẫn tới thư mục chứa menu.

Tham số 3: Đường dẫn tới thư mục chứa các thông báo hệ thống.

Tham số 4: Đường dẫn tới thư mục chứa các tệp làm việc tạm thời.

Tham số 5: Đường dẫn tới thư mục chứa các tệp CSDL.

Tham số 6: Tên CSDL định trước.

Tham số 8: Dấu phân cách trường lập. Dấu định trước là %. Có thể thay bằng dấu khác, ví dụ:

8=&

Tham số 101: Ngôn ngữ định trước, EN là tiếng Anh, VN là tiếng Việt. Ví dụ:

101=VN

Tham số 102: Menu định trước, DF là menu đầy đủ, SH là menu ngắn gọn hơn. Ví dụ:

102=DF

Tham số 103: Hiện hình chi tiết các quá trình tìm tin. Nếu đặt giá trị của tham số này là 1 thì hệ thống sẽ đưa ra cửa sổ trình bày chi tiết quá trình tìm tin. Nếu đặt là 0 thì cửa sổ này sẽ không xuất hiện. Ví dụ:

103=1

Tham số 104: Trình bày từ điển. Đặt là 0 sẽ hiện đầy đủ từ điển. Nếu đặt là 1 thì chỉ các thuật ngữ trong trường được chọn mới xuất hiện. Ví dụ:

104=1

Tham số 105: Hiện hình chi tiết kết quả tìm tin. Khi sử dụng chế độ tìm tin ở trình độ cao và đặt giá trị của tham số này là 1 thì hệ thống sẽ đưa ra cửa sổ trình bày các kết quả tìm tin. Nếu đặt là 0 thì cửa sổ này sẽ không xuất hiện. Ví dụ 105=1.

Tham số 109: Font chữ dùng để hiệu đính trường. Ví dụ:

109=.VnTime,24

có nghĩa là trong trường nhập dữ liệu sẽ có font chữ VnTime với cỡ chữ 12.

Tham số 110: Font chữ, cỡ chữ, kiểu chữ, màu của chữ dùng cho format hiện hình. Ví dụ:

110=fonts((nil,.VnTime)),cols((0,0,0),(255,0,0),(0,255,0),(0,0,255)),cl0,f0,fs26

Tham số 111: Font chữ, cỡ chữ dùng để hiệu đính format. Ví dụ:

111=.VnTime,24

Tham số 120: Hiện hình thanh công cụ. Đặt là 1, thanh công cụ sẽ hiện ra dưới thanh menu. Đặt là 0, thanh công cụ sẽ không hiện ra. Ví dụ:

120=1

Tham số 121: Hiện hình trợ giúp thanh công cụ. Đặt là 1 thì sẽ hiện thông báo trợ giúp mỗi khi con trỏ được đặt vào nút công cụ. Đặt là 0 thì thông báo sẽ không hiện ra. Ví dụ:

121=1

Tham số 123: Hiện hình các trường rỗng. Nếu đặt là 1, trên biểu mẫu hiệu đính trường sẽ hiện đầy đủ các trường có và

không có dữ liệu. Nếu đặt là 0 thì trên biểu mẫu hiệu đính trường chỉ hiện các trường có dữ liệu. Ví dụ:

123=1

Tham số 130: Cập nhật tự động tệp đảo. Đặt là 1, hệ thống sẽ tự động cập nhật tệp đảo mỗi khi đóng cửa sổ nhập dữ liệu. Đặt là 0 thì hệ thống sẽ không tự động cập nhật tệp đảo. Ví dụ:

130=1

Ta có thể tạo ra tệp syspar.par bằng một chương trình soạn thảo bất kỳ. Ví dụ : Một tệp syspar.par có thể có nội dung như sau:

1=c:\winisis\prog\	(đường dẫn tới chương trình)
2=C:\winisis\menu\	(đường dẫn tới menu)
3=C:\winisis\msg\	(đường dẫn tới các thông báo)
4=C:\winisis\work\	(đường dẫn tới các tệp làm việc tạm thời)
5=c:\winisis\data\	(đường dẫn tới CSDL)
101=VN	(Tiếng Việt)
102=DF	(menu ngầm định là MNENDF)
103=1	(đưa ra cửa sổ trình bày quá trình tìm kiếm)
104=1	(trình bày từ điển của trường được chọn)
105=1	(tự động trình bày kết quả tìm kiếm)
106=txt_txt.tab	(bảng chuyển đổi ký tự từ DOS sang Windows)
107=Txt_txt.tab	(bảng chuyển đổi ký tự từ Windows sang DOS)

109=.VnTime, 24 (font chữ dùng để hiệu đính trường)

111=.VnTime, 24 (font chữ dùng để hiệu đính format)

110=fonts((nil,.VnTime) (nil,.VnTimeH)),cl0,f0,fs26 (font, màu, cỡ chữ)

; Highlight settings

; Auto I/F Update

120=1 (hiện hình thanh công cụ)

121=1 (hiện hình thanh trợ giúp)

123=1 (hiện hình trường rỗng)

130=1 (tự động cập nhật tệp đảo)

142=3,15 (làm cho các thuật ngữ tìm có màu sáng ở kết quả tìm)

902=0

981=(null)

122=1, 0, 0, 640, 480 (trạng thái của cửa sổ chính)

Khai báo các tham số của tệp SYSPAR.PAR bằng chương trình WinISIS

- Chọn menu **Cấu hình** trên thanh Menu
- Chọn mục **Tạo tệp hệ thống SYSPAR.PAR**

Trên màn hình hiện ra trang menu khai báo các tham số hệ thống như sau:

Thiết lập CDS/ISIS

Hệ thống	Lựa chọn	Hiển thị
Directories		
Chương trình:	c:\winisis\prog\	
Thực đơn:	C:\winisis\menu\	
Thông báo:	C:\winisis\msg\	
CSDL:	c:\winisis\data\	
File KQ:	C:\winisis\work\	
Kiểu mẫu/ảnh:		
Khu vực		
Ngôn ngữ ban đầu:		VN
Profile ban đầu:		DF
Font chữ Windows:		Nhỏ
CSDL mặc định: ☒ Mở không sử dụng thanh công Vị trí Bên trên: 0 Trái: 0 Rộng: 0 Cao: 0		
Thoát ra		Ok

Để cài đặt tiếng Việt, trong mục **Default Language** ta đặt : VN

Kích vào mục **Lựa chọn**, ta sẽ thấy hộp thoại sau:

Thiết lập CDS/ISIS

Hệ thống	Lựa chọn	Hiển thị
Tìm tin		
☐ Cho xem tiến triển của lệnh tìm ☐ Hiện Từ điển đầy đủ ☒ Tự động hiện KQ Tìm tin ☒ Đánh dấu từ tìm tin Màu chữ nền: 13 Màu nền: 8		
☒ Cập nhật tự động File đảo ☒ Hiện thị thanh công cụ ☒ Hiện thị thanh trạng thái		
Dữ liệu		
☒ Hiện kết quả Default ☐ Thanh công cụ kiểu khác (Alternate) ☐ Hiện kết quả theo bảng		
Bản chuyển - TAB (Đọc): txt_txt.tab Bảng chuyển - TAB (Ghi): txt_txt.tab		
Thoát ra		Ok

Trong mục *Dữ liệu* ta phải điền vào mục *Bảng chuyển-TAB* các tham số:

txt_txt.tab

Trong mục *Tìm tin* ta đánh dấu vào các tham số cần chọn, ví dụ: Tự động hiện kết quả tìm tin, đánh dấu từ tìm tin, cập nhật tự động file đảo, v.v...

Sau đó chọn mục *Hiển thị*, trên màn hình xuất hiện hộp thoại:

The screenshot shows a Windows-style dialog box titled "Thiết lập CDS/ISIS". It has three tabs: "Hệ thống", "Lựa chọn", and "Hiển thị". The "Hiển thị" tab is active. Inside the dialog, there are several settings:

- On the left, a group box contains three checkboxes:
 - ☐ Hiện thị từ phải sang trái
 - ☒ Hiện cả trường rỗng
 - ☐ Hộp thông báo lỗi ISIS-Pascal
- On the right, there are two numeric input fields:
 - "Hiện thị có Zoom:" with a value of 120
 - "Độ dài từ tối đa:" with a value of 50
- Below these, there are two text input fields for "Dấu phân cách In Index:" and "Dấu phân cách trường lặp:", both currently empty.
- A section titled "Các giá trị Format defaults:" contains a text box with the value: `fonts((nil,.VnCourier),(nil,.VnTime),(nil,.VnTimeH)),cl0,f0,fs26`.
- Below that, a "Fonts:" section has two input fields:
 - "Font nhập tin:" with the value ".VnTime"
 - "Font hiệu đính Format in:" with the value ".VnTime"
- At the bottom, there are two buttons: "Thoát ra" (Exit) and "Ok".

Trong ô *formatting default* phải nhập font chữ tối thiểu:

fonts((nil,.VnTime),(nil,.VnTimeH)),cl0,f0,fs26

Kích chuột vào chữ A và chọn font chữ để hiển thị tiếng Việt khi nhập tin và hiệu đính format in. Có thể chọn tham số hiển thị trường rỗng. Cuối cùng kích chuột vào nút OK để lưu các tham số này vào SYSPAR.PAR.

2.3. Thủ tục cài đặt

Chương trình cài đặt được tổ chức trên 3 đĩa mềm 1,44MB, gồm hai đĩa chương trình có số hiệu #1 và #2, một đĩa chứa thực đơn và thông báo tiếng Việt.

Để tiến hành cài đặt WinISIS ta tiến hành các bước sau:

(1) Cài đặt chương trình WinISIS với menu và thông báo bằng tiếng Anh

Đặt đĩa #1 vào ổ A. Chạy chương trình cài đặt bằng cách nháy đúp chuột vào file SETUP.EXE trong ổ A hay tại cửa sổ mục "RUN" của nút START gõ vào a:setup rồi bấm OK.

Chương trình SETUP sẽ tự động cài đặt chương trình, tạo các thư mục cần thiết và tạo ra tệp tham biến hệ thống SYSPAR.PAR cho bản tiếng Anh. Ta nên trả lời **"continues"** cho tất cả các câu hỏi của chương trình SETUP. Riêng câu hỏi về ngôn ngữ thì thay dòng **"FR as French and SP as Spanish"** bằng chữ **EN** (cho tiếng Anh).

Khi hết đĩa #1, theo thông báo của hệ thống ta đặt tiếp đĩa #2, rồi bấm OK. Kết thúc việc cài đặt, hệ thống sẽ thông báo:

Setup succesfully

Đối với trường hợp WinISIS chạy trong môi trường Windows 95 trở lên, có thể phải cài đặt thêm một số file phụ, bằng cách copy các tệp CTL3DV2.DLL trong thư mục \WINISIS\CTLDV2\ sang thư mục \WINISIS\ và thư mục \WINDOW\SYSTEM\ của máy đang cài đặt.

(2) Chạy thử chương trình

Chạy thử chương trình bằng cách thực hiện các lệnh:

START → Programs → CDS_ISIS for Windows → CDS_ISIS for Windows

Nếu chương trình cài đặt bình thường ta sẽ mở được WinISIS với menu bằng tiếng Anh.

(3) Cài đặt thực đơn và thông báo tiếng Việt

Các tệp thực đơn (menu) và thông báo (message) tiếng Việt được lưu giữ trong đĩa mềm #3, trong hai thư mục MENUVN và MSGVN. Để chương trình dùng được thực đơn và thông báo tiếng Việt ta cần sao chép các tệp đó bằng lệnh COPY trên DOS như sau:

```
COPY A:\MENUVN\*. * C:\WINISIS\MENU
```

```
COPY A:\MSGVN\*. * C:\WINISIS\MSG
```

Sau đó phải khai báo bổ sung các tham số trong tệp SYSPAR.PAR theo nội dung đã trình bày ở trên bằng chương trình soạn thảo bất kỳ (ví dụ, dùng chức năng F4- Edit của NC).

Trong một số trường hợp, sau khi cài đặt ta vẫn không đọc được menu và thông báo bằng tiếng Việt. Nguyên nhân vì Windows của máy bạn không sử dụng font hệ thống của bộ gõ ABC. Để khắc phục bạn phải sao chép các tệp SSERIFE.FON, VGASYS.FON, VGAIFIX.FON từ thư mục C:\WINDOWS\ABC sang thư mục C:\WINDOWS\SYSTEM.

Bây giờ bạn có thể khởi tạo WinISIS như phần trên. Các menu và thông báo sẽ hiện ra bằng tiếng Việt.

2.4. Cài đặt trên mạng cục bộ (LAN)

Version này của WinISIS có thể chạy trên mạng cục bộ. Để hỗ trợ phương tiện này, ta phải đặt tham số 14 của tệp SYSPAR.PAR là 1:

$$14=1$$

Khi đó WinISIS cho phép nhiều người sử dụng nhập dữ liệu vào CSDL cùng một lúc.

Khi tham số 14=0 có nghĩa là chế độ một người dùng.

Để cài đặt WinISIS trên mạng có hai cách:

- Cài đặt một bản duy nhất vào máy chủ, nhiều người sử dụng chung phần mềm.
- Cài đặt phần mềm trên từng máy trạm, còn dữ liệu thì dùng chung.

Khi WinISIS khởi động, nó sẽ tìm trong thư mục hiện tại hai loại tệp: SYSPAR.PAR và WINISIS.DAT. Sau đó nó đọc tệp SYSPAR.PAR để định vị các tệp hệ thống khác như các menu và thông báo. Khi ta cài WinISIS lên máy chủ phải bảo đảm thư mục Windows hiện tại (thường là \WINISIS) thực sự là thư mục của mạng chứa tệp WISIS.EXE.

Để xác lập khai báo ổ đĩa logic kết nối với mạng trong Windows 95 hoặc Windows NT, ta sử dụng lệnh:

Windows explorer → Tools → Map Network Drive

Khi đó Windows sẽ gán tên ổ (D, E hoặc F...) để kết nối mạng. Cần phải bảo đảm ký tự mà Windows gán cho ổ logic luôn luôn cố định.

Để WinISIS tìm đúng các tham biến của hệ thống, ta cần khai báo các tham số trong SYSPAR.PAR theo phương thức địa chỉ tương đối mà không dùng địa chỉ tuyệt đối. Riêng mục 4= cần phải ghi rõ ổ đĩa C: để bảo đảm mỗi người sử dụng có thư mục làm việc riêng rẽ. Ví dụ, nếu ta đặt dữ liệu vào thư mục \WINISIS\DATA, ta nên khai báo các tham số của SYSPAR.PAR như sau:

1=PROG\

2=MENU\

3=MSG\

4=C:\

5=DATA\

2.5. Khởi động WinISIS

Để gọi chương trình WinISIS, ta có thể thực hiện theo hai cách:

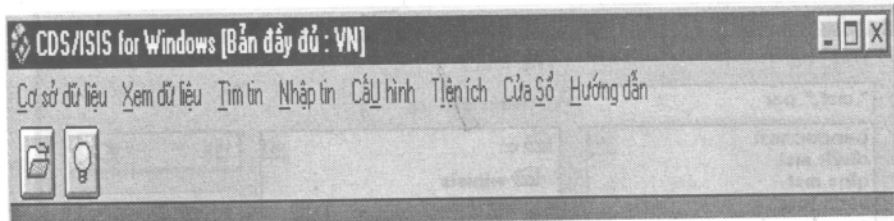
1) Thực hiện dãy các lệnh sau:

START → Programs → CDS_ISIS for Windows → CDS_ISIS for Windows

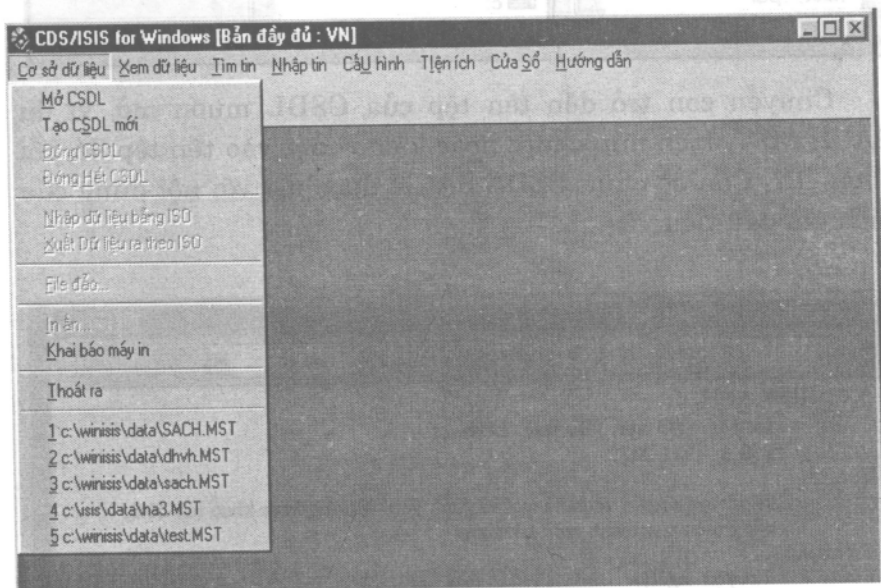
2) Nháy đúp vào biểu tượng của WinISIS trên nền màn hình Desktop:



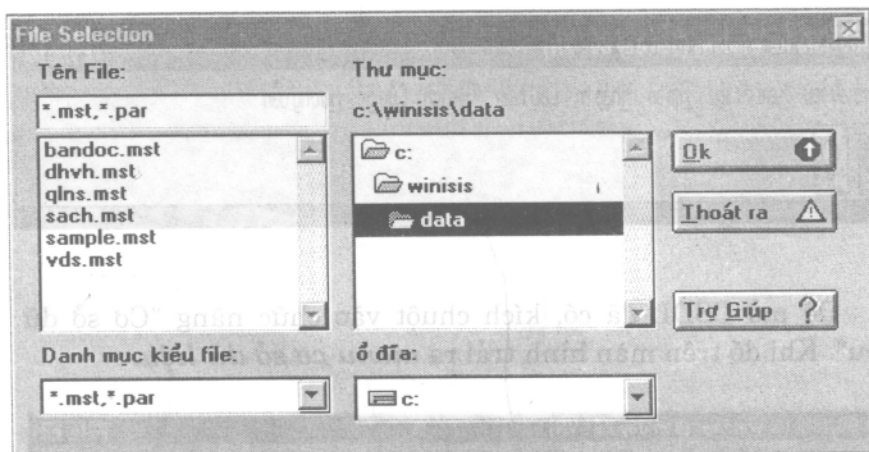
Xuất hiện cửa sổ chính của WinISIS. Cửa sổ này có dạng:



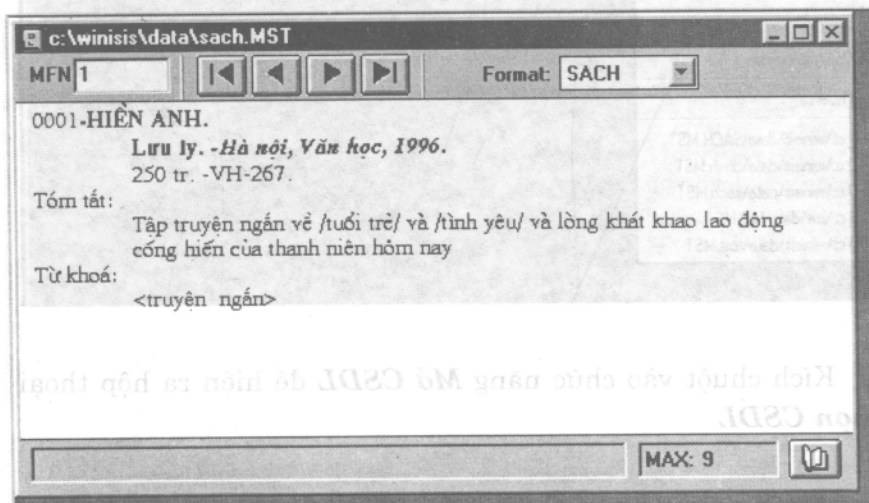
Để mở CSDL đã có, kích chuột vào chức năng "Cơ sở dữ liệu". Khi đó trên màn hình trái ra **menu cơ sở dữ liệu**.



Kích chuột vào chức năng **Mở CSDL** để hiện ra hộp thoại **Chọn CSDL**



Chuyển con trỏ đến tên tệp của CSDL muốn mở, ví dụ SACH.MST, kích đúp chuột hoặc kích chuột vào tên tệp đó, rồi chọn OK. Cửa sổ chứa CSDL SACH được mở với nội dung của biểu ghi đầu tiên.



Chương trình WinISIS cho phép lưu tên của 5 CSDL đã từng được mở trước đó. Ta có thể mở một trong những CSDL đó bằng cách chọn tên CSDL tương ứng, có trong danh sách các CSDL đã mở ở cuối thực đơn, rồi kích đúp chuột vào đó.

3. NGÔN NGỮ TẠO FORMAT

Ngôn ngữ tạo format dùng để mô tả:

- **Format hiện hình:** sử dụng để trình bày biểu ghi trên màn hình hoặc máy in (sau đây gọi là format in).
- **Format chọn lọc dữ liệu:** sử dụng trong FST để xác định dữ liệu đưa vào tập đảo.

Một format bao gồm một dãy các lệnh được WinISIS thực hiện theo thứ tự từ trái sang phải.

Ví dụ: Để có kết quả đầu ra thể hiện như cách trình bày biểu ghi 0001 ở trên, format hiện hình của CSDL SACH là:

```
mfn(4),mdl,{f26,b,'-',v1+|:|}/{m(900,0),b,v2,i,'-',v3}/ {m(900,0),
v4,'-', v5}/mpl,{'Tóm tắt:',m(900,0),v6}/ {'Từ khoá: ', m(900,0),v7} /#
if p(v8) then mpl, link(('Kích vào đây để xem toàn văn'),'CMD
winword ',v8) fi/
```

Các lệnh và các hàm cơ bản của ngôn ngữ tạo format

1. Các lệnh hiển thị nội dung một trường, trường con

Để hiển thị nội dung một trường hoặc trường con của một biểu ghi người ta dùng lệnh bao gồm chữ V theo sau là nhãn trường hoặc nhãn trường kèm theo ký hiệu của trường con:

Vtt

Vtt^x

trong đó tt là nhãn trường, x là ký hiệu của trường con. Ví dụ:

FORMAT	ĐẦU RA
V2	Lưu ly
V3^b	Văn học
V3	^aHà Nội^bVăn học^c1996

2. Lệnh hiển thị nội dung trường lặp

Khi hiển thị nội dung trường lặp, dấu % dùng để ngăn cách các tình huống của trường phải được thay bằng một dấu chấm câu để có cách trình bày đúng. Có nhiều cách trình bày. Dưới đây là một số lệnh đơn giản:

- Vtt+|;|** Ký tự giữa hai dấu gạch đứng (ở đây là dấu chấm phẩy) sẽ được lặp lại trong mỗi tình huống.
- (Vtt/)** Hiển thị mỗi tình huống trên một dòng; dấu ngoặc đơn là ký hiệu yêu cầu WinISIS thực hiện lại việc xuống dòng sau mỗi tình huống.

Ví dụ: Nếu trường tác giả (V1) là trường lặp, với dữ liệu nhập là: Bảng Việt%Lưu Quang Vũ thì format và đầu ra tương ứng là:

FORMAT	ĐẦU RA
V1+ ;	Bảng Việt; Lưu Quang Vũ
(V1/)	Bảng Việt Lưu Quang Vũ

3. Lệnh hiển thị số hiệu biểu ghi tệp chủ MFN

Để hiển thị số hiệu biểu ghi tệp chủ với x chữ số ta dùng lệnh:

MFN(x)

Ví dụ: MFN(3) = 004 MFN(4) = 0004

4. Lệnh hiển thị các nhan đề, các ký tự trước nội dung của trường

"xxxxxx" = Các ký tự trong dấu ngoặc kép sẽ hiển thị nếu trường sau nó không rỗng

'xxxxxx' = Các ký tự trong dấu ngoặc đơn sẽ hiển thị trong mọi trường hợp

Ví dụ:

FORMAT	ĐẦU RA
"Từ khoá:", V7	Từ khoá: <giáo trình> <thông tin> <thông tin học>
V4, '-', V5	155 tr - GT.101

5. Trình bày trang

+ Các lệnh xuống dòng:

Xuống một dòng (vô điều kiện)

/ Xuống dòng mới (nếu dòng trước không rỗng)

+ Lệnh xác định độ lùi:

M(indent, findent)

trong đó:

indent = độ lùi vào của đoạn văn bản tính từ lề trái.

findent = độ lùi vào của dòng đầu tiên của đoạn văn.

Cả hai tham số được tính theo đơn vị *twip* (567 twip bằng 1 cm)

+ *Lệnh xác định điểm dừng (Tabulation):*

TAB(giá trị)

trong đó:

giá trị = vị trí của ký tự tiếp theo được đo bằng twip tính từ lề trái.

6. Các lệnh định dạng ký tự

Các lệnh này liên quan đến việc thể hiện các đặc tính của ký tự (kiểu font, kích cỡ, màu và dạng)

LỆNH	CHỨC NĂNG
b	đậm
i	ngghiêng
ul	gạch chân
fn	font (n là số hiệu font chữ, f26 = font chữ in hoa)
fsn	cỡ font chữ (n là kích thước tính theo nửa giá trị font, fs24 = font cỡ 12 point)
cln	màu (n là số hiệu của màu, được xác định trong bảng màu)

Một số màu cơ bản được cho trong bảng sau:

MÀU	LỆNH
Đen (Black)	cl0
Nâu (Brown)	cl3
Trắng (White)	cl10
Xám (Gray)	cl11
Đỏ (Red)	cl13
Xanh lá cây (Green)	cl14
Vàng (Yellow)	cl15
Xanh da trời (Blue)	cl16
Tím (Violet)	cl17
Xanh da trời sáng (Light blue)	cl18

Các lệnh trình bày trang và định dạng có hiệu lực ngay với các trường đứng sau nó. Ta có thể điều chỉnh hiệu lực của các lệnh này bằng dấu { }, nghĩa là chỉ có những trường nằm trong hai dấu đó mới chịu tác dụng của lệnh.

Ví dụ: Trong CSDL SACH với format:

mfn(4),mdl,{f26,b,'-',v1+|:|}/ {m(1000,0),b,V2,i,'-',V3}

thì kết quả đầu ra sẽ là:

0001-HIẾN ANH.

Lưu ly - Hà Nội, Văn học, 1996.

Kết quả trên cũng có thể được thực hiện bởi format sau:

`mfn(4),mdl, {f26, b,'-', v1+|;|} / {tab(1000), b, V2, i, '-', V3}`

7. Chế độ trình bày dữ liệu

WinISIS trình bày dữ liệu theo ba chế độ nhờ một lệnh tạo thành bởi chữ đầu tiên là chữ M, theo sau là hai chữ nữa:

Chữ thứ nhất sau M:

P = Chế độ kiểm định: các trường được in ra như khi cập nhật, tức là bao gồm cả các dấu ngăn cách (^a, < >, / /, ...) nhưng dấu % thì bị loại bỏ nhưng không được thay thế.

H = Chế độ tiêu đề: loại bỏ dấu < > giữa các từ khoá (nhưng không bỏ dấu / /) và thay thế chúng bằng dấu chấm phẩy, các dấu ngăn cách trường con được thay thế bằng dấu phẩy.

D = Chế độ dữ liệu: tương tự như chế độ tiêu đề, ngoài ra cuối một trường được tự động thêm vào một dấu chấm, tiếp theo là hai khoảng trống.

Chữ thứ hai sau M:

U: Dữ liệu được chuyển thành chữ hoa.

L: Dữ liệu giữ nguyên không đổi.

Nếu lệnh chọn chế độ trình bày không được chỉ ra thì WinISIS sử dụng chế độ MPL như là chế độ mặc định.

Ví dụ:

FORMAT	ĐẦU RA
mpl,V3	^a Hà Nội ^b Văn học ^c 1995
mdl,V3	Hà Nội,Văn học,1995.
mhu,v3	HÀ NỘI,VĂN HỌC,1995
mpl,V7	<giáo trình> <thông tin> <thông tin học>
mdl,V7	giáo trình;thông tin;thông tin học.

8. Các hàm

a. Hàm số

VAL(format): trả lại giá trị bằng số của đối số cho bởi *format*

Ví dụ: Nếu $V3^c = 1995$ thì $val(V3^c) = 1975$

$val('15.76') = 15.76$ và $val('water') = 0$

LR((format), từ đến)

Hàm LR tìm trong tệp đảo thuật ngữ xác định bởi *format* và trở về tất cả các định vị của thuật ngữ, tức là trở về các số hiệu MFN của biểu ghi tệp chủ mà từ đó thuật ngữ được lấy ra. Hàm này thường được sử dụng kèm theo hàm ký tự REF như sẽ trình bày dưới đây.

b. Hàm ký tự

REF(biểu thức, format)

Hàm REF cho phép trích dữ liệu từ các biểu ghi luân phiên trong tệp chủ. Tham số thứ nhất là một biểu thức số cho MFN

của các biểu ghi luân phiên đã được chọn, còn tham số thứ hai là format được áp dụng cho các biểu ghi này.

Ví dụ: Với format DHVH.PFT của cơ sở dữ liệu DHVH:

```
mfn(3), '-', mdl, {f26,b,cl17, V1+;|} /{m(900,0), b, V2,i ,'-', V3} /  
{m(900,0), V4,'-',V5}/ "Từ khóa:", v6/# mpl, {b,ul, cl2, 'Cùng một  
tác giả:'} / ref(lr((V1)), {m(900,0), '-', V2})/##
```

thì đầu ra của một biểu ghi sẽ là:

001-Tế Hanh.

Hai nửa yêu thương. -Hà Nội, Văn học, 1963.

97 tr. -V21.T.250H.

Từ khóa:<Tho>

Cùng một tác giả:

- Hai nửa yêu thương
- Theo nhịp tháng ngày
- Khúc ca mới

Trong lệnh

$\text{ref}(\text{lr}((V1)), \{m(900,0)\}, '-', V2/)$

hàm $\text{lr}((V1))$ sẽ cung cấp tất cả các số hiệu MFN của các biểu ghi có tác giả là Tế Hanh ($V1=\text{Tế Hanh}$) và cho hiển thị tất cả tên sách của tác giả đó (các giá trị của trường $V2$).

c. Hàm logic

$P(Vtt)$, $P(Vtt^x)$: nhận giá trị đúng nếu trường có nhân tt hoặc tt^x trong dấu ngoặc có nội dung.

$A(Vtt)$, $A(Vtt^x)$: nhận giá trị đúng nếu trường có nhân tt hoặc tt^x trong dấu ngoặc không có nội dung.

Các hàm logic trên thường được sử dụng với **lệnh IF**, có cấu trúc như sau:

IF điều kiện THEN format-1 ELSE format-2 FI

Ví dụ: CSDL SACH có trường 10 là trường từng thư, trường 2 là trường tên sách, trường 4 là trường số trang. Giả sử trong biểu ghi số 12 cuốn sách “Môi trường châu Á - Thái Bình Dương” là ấn phẩm Tổng luận Khoa học - Kỹ thuật - Kinh tế, một từng thư. Còn biểu ghi số 13 là cuốn sách “Một số cơ sở ban đầu về Internet”, 214 trang, không phải là từng thư. Khi đó, với format:

$mfn(4)/mdl,tab(900),v2, '-',v4,if p(v10) then mpl,'-','(,v10, ')' fi$
thì kết quả đầu ra của biểu ghi 12 sẽ là:

0012

Môi trường châu Á - Thái Bình Dương. - 96 p. - (Tổng
luận Khoa học - Kỹ thuật - Kinh tế)

Còn đầu ra của biểu ghi 13 sẽ là:

0013

Một số cơ sở ban đầu về Internet. - 214 p.

Dòng lệnh

$if p(v10) then mpl, '-', '(,v10, ')' fi$

có nghĩa là: nội dung trường 10 chỉ được hiện ra và đặt trong dấu ngoặc đơn, nếu trường 10 không rỗng. Vì vậy ở biểu ghi số 13, vị trí này bỏ trống.

Ví dụ: Cơ sở dữ liệu QLNS có trường 1 là trường họ và tên, trường 9 là trường "Đảng", có hai trường con. Trường con a cho biết nhân viên trong biểu ghi có là đảng viên hay không, trường con b cho biết ngày kết nạp nếu là đảng viên.

Trong biểu ghi 001, anh Lê Hoài Nam là đảng viên, kết nạp ngày 19/05/1989. Khi nhập dữ liệu ở trường 9 ta ghi:

Đảng: ^aĐảng viên^b19/05/1989

Trong biểu ghi 002, chị Phạm Kim Hoa không là đảng viên. Khi nhập dữ liệu ở trường 9 ta để trường này trống.

Với Format:

mfn(3)/"Họ và tên: ",Tab(3000),V1/ if p(v9) then 'Đảng : Đảng viên' ,("v9^b")" else if a(v9) then 'Đảng: không' fi fi/#

thì thông tin ở đầu ra của hai biểu ghi tương ứng sẽ là:

001
Họ và tên: Lê Hoài Nam
Đảng: Đảng viên(19/05/1989)

002
Họ và tên: Phạm Kim Hoa
Đảng: không

9. Các lệnh liên kết siêu văn bản LINK

Lệnh LINK có dạng:

LINK((mô tả), hành động)

trong đó:

Mô tả: là mệnh đề mô tả hành động cho người dùng thực hiện. Mệnh đề này thường có màu xanh, có gạch dưới và phải đặt trong dấu ngoặc.

Hành động: là một lệnh siêu văn bản, viết bằng chữ in hoa, sát ngay sau là một khoảng trống. Khi kích chuột vào mệnh đề mô tả thì lệnh này sẽ được thực hiện.

Dưới đây là một số lệnh chính.

a. Lệnh CMD

Lệnh CMD khởi động chương trình của Windows hay MS/DOS được chỉ ra ngay sau lệnh, tiếp theo là một khoảng trống (Chú ý: trong Windows để tìm đến chương trình, đường đến phải được chỉ ra trong lệnh PATH của tệp AUTOEXEC.BAT). Ví dụ:

if p(v8) then mpl, link(('Kích vào đây để xem toàn văn'),'CMD winword ',v8) fi/

Nếu trường 8 của biểu ghi đang xét là không rỗng thì trên màn hình đầu ra của format của biểu ghi này sẽ hiện dòng thông báo:

Kích vào đây để xem toàn văn

Nếu ta kích chuột vào dòng chữ trên, lệnh CMD sẽ gọi chương trình Word for Windows để đọc tệp văn bản, tên (kèm theo đường dẫn) của tệp văn bản này chứa trong trường 8.

b. Lệnh **FORMAT** tên-format

Lệnh này cho phép thay đổi format hiện hình sang format khác được chỉ ra trong tham số tên-format.

Ví dụ: Trong CSDL mẫu DHVH ngoài format DHVH.PFT còn format PHIEU.PFT. Nếu trong format hiện hình DHVH.PFT có dòng lệnh:

link(('Xem dữ liệu bằng format in phiếu'),'FORMAT phiếu')

thì trên màn hình xuất hiện dòng:

Xem dữ liệu bằng format in phiếu

Nếu kích vào dòng này, WinISIS sẽ lại đưa ra biểu ghi hiện tại nhưng sử dụng format **phiếu**.

c. Lệnh **BROWSE** tên-CSDL [,mfn,tên-format]

Lệnh này cho phép mở CSDL được chỉ ra trong tham số tên-CSDL trong một cửa sổ CSDL mới, đưa ra biểu ghi có số mfn trong tham số *mfn* và sử dụng format trong tham số *tên-format*. Nếu không đưa vào các tham số này thì ngầm định mfn = 1 và tên format là format ngầm định.

Ví dụ: Trong CSDL quản lý bạn đọc, có tên là BANDO, có hai format:

- Format BANDO.PFT trình bày các thông tin về bạn đọc (sơ yếu lý lịch của bạn đọc).

- Format MUON.PFT trình bày các thông tin về tình hình mượn sách của bạn đọc. Sách mượn thì được lưu trong CSDL SACH.

Để biết thông tin chi tiết về cuốn sách mà bạn đã mượn, trong format MUON.PFT bạn hãy đưa thêm vào dòng lệnh:

`link(('Mở cơ sở dữ liệu SACH'), 'BROWSE sach')`

Khi đó trên màn hình xuất hiện dòng:

Mở cơ sở dữ liệu SACH

Kích vào dòng trên, WinISIS sẽ mở CSDL SACH trong một cửa sổ mới và đưa ra biểu ghi số 1, sử dụng format mặc định SACH.PFT.

d. Lệnh TEXBOX format

Lệnh này cho phép trình bày văn bản do thực hiện tham số *format* trong một cửa sổ riêng biệt. Thí dụ cơ bản về lệnh này như sau:

`link(('Mở cửa sổ mới'), 'TEXBOX v6')`

Khi ta kích vào dòng siêu văn bản Mở cửa sổ mới, WinISIS sẽ mở ra một cửa sổ mới chứa nội dung trường 6.

Các kiểu cửa sổ:

a. TEXTBOX: cửa sổ hiện ra như là một cửa sổ độc lập. Người dùng có thể đóng cửa sổ này bằng cách nháy vào nút đóng.

b. TEXTBOXCHILD: như là một cửa sổ con của CSDL. Khi đóng CSDL, WinISIS sẽ tự động đóng các cửa sổ con của chúng.

c. TEXTBOXRCHILD: cửa sổ được khai báo như là một cửa sổ con của biểu ghi. Khi chuyển sang biểu ghi khác hay đóng CSDL, WinISIS sẽ tự động đóng các cửa sổ con của nó.

Tải tệp format:

Ta có thể gắn từ khoá LOAD vào các dạng lệnh TEXTBOX trên, tiếp theo sau là tham số tên- format để tải tệp format vào. Format này sẽ được thực hiện khi kích vào dòng siêu văn bản.

Ví dụ: Trong cơ sở dữ liệu BANDOC có hai format hiện hình:

- Format BANDOC.PFT dùng để trình bày phiếu đăng ký của bạn đọc.
- Format MUON.PFT dùng để trình bày phiếu mượn sách của bạn đọc.

Để xem phiếu mượn sách của bạn đọc đang xét, trong format BANDOC.PFT ta đưa vào dòng lệnh sau:

```
mpl,link(('Cho xem phiếu mượn sách'), 'TEXTBOXRCHILDLOAD  
muon')
```

Khi kích vào dòng siêu văn bản Cho xem phiếu mượn sách, lệnh này sẽ mở cửa sổ con của biểu ghi đang xét, cung cấp thông tin về tình hình mượn sách của bạn đọc đó.

4. BẢNG CHỌN TRƯỜNG

Bảng chọn trường (FST) xác định các tiêu chuẩn lấy ra một hoặc nhiều phần tử dữ liệu từ các biểu ghi tệp chủ để tạo ra đầu vào cho tệp đảo của CSDL. Các thuật ngữ tạo ra bởi FST khi đã được lưu trong tệp đảo sẽ tạo thành từ điển các thuật ngữ của CSDL.

Dưới đây là một ví dụ về FST:

Tên CSDL: TEST		Tên FST: TEST
ID	IT	Format tách dữ liệu
1	0	mhl,(v1/)
2	4	mhl,v2
3	0	"Nơi XB=",v3^a
3	0	"Nhà XB=",V3^b
6	3	mpl,v6
7	2	mpl,v7

FST gồm nhiều dòng, mỗi dòng xác định 3 tham số:

1. Dấu nhận biết trường (ID).
2. Kỹ thuật đảo (IT).
3. Format tách dữ liệu (theo ngôn ngữ tạo format).

Để tách các phần tử dữ liệu bằng FST CDS/ISIS sẽ đọc các biểu ghi tệp chủ và thực hiện với từng biểu ghi và từng dòng FST theo quá trình sau:

1. Thực hiện **format** để tách dữ liệu tương ứng từ biểu ghi.
2. Áp dụng **kỹ thuật đảo** đối với các dữ liệu do format tạo ra.
3. Gán cho mỗi phần tử được tạo ra một **dấu nhận biết trường**.

Các kỹ thuật đảo

a. Kỹ thuật 0: Tạo ra một phần tử từ mỗi dòng được tách ra bởi format. Kỹ thuật này thường áp dụng cho việc đảo toàn bộ trường hoặc trường con.

b. Kỹ thuật 1: Tạo một phần tử từ mỗi trường con hoặc từ từng dòng tách ra từ format. Kỹ thuật này phải kết hợp với chế độ kiểm định (MPL hoặc MPU) trong format tách dữ liệu để CDS/ISIS có thể tìm được dấu phân cách trường con. Kỹ thuật 1 thực chất là kỹ thuật 0 thu gọn.

Ví dụ: Nội dung trường 3 của một biểu ghi:

^aHà Nội^bVăn học^c1995

FST			ĐẦU RA TẠO BỞI FORMAT	CÁC PHẦN TỬ ĐƯỢC TẠO RA
3	1	mpl,v3	^aHà Nội^bVăn học^c1995	Hà Nội Văn học 1995
3	0	mhl,v3^a/ v3^b/v3^c	Hà Nội Văn học 1995	Hà Nội Văn học 1995
3	0	mhl,v3	Hà Nội,Văn học,1995	Hà Nội,Văn học,1995

Ví dụ: Trong CSDL quản lý nhân sự trường ngày sinh có nhân là 2. Dữ liệu nhập vào có dạng DD/MM/YYYY, ví dụ 25/08/1972. Nếu chỉ quan tâm đến năm sinh, ta có thể tách phần năm sinh từ trường này theo FST sau:

FST			ĐẦU RA TẠO BỞI FORMAT	CÁC PHẦN TỬ ĐƯỢC TẠO RA
2	0	mpl, v2*6.4	1972	1972

Trong đó lệnh v2*6.4 cho phép tách ra 4 ký tự đặt ngay sau ký tự thứ 6 của trường 2 để đưa vào tệp đảo, đó là số hiệu của năm 1972.

c. Kỹ thuật 2: Tạo một phần tử từ thuật ngữ nằm trong dấu móc nhọn (<...>). Kỹ thuật này phải kết hợp với chế độ kiểm định vì các chế độ khác sẽ loại bỏ các dấu móc.

d. Kỹ thuật 3: Giống kỹ thuật 2, chỉ khác là các thuật ngữ đặt trong dấu gạch chéo (/.../).

Ví dụ:

FST			ĐẦU RA TẠO BỞI FORMAT	CÁC PHẦN TỬ ĐƯỢC TẠO RA
6	3	mpl,v6	Bài viết về /tự động hoá/ hoạt động /thư viện/	tự động hóa thư viện
7	2	mpl,V7	<chuẩn hoá><tiếng Việt> <tin học>	chuẩn hoá tiếng Việt tin học

e. Kỹ thuật 4: Tạo ra phần tử từ mỗi từ (word) trong văn bản do format tạo ra. Chú ý rằng khi sử dụng kỹ thuật này với trường có dấu phân cách trường con ta cần phải kết hợp với chế độ tiêu đề hoặc dữ liệu (mhl, mdl) trong format tách dữ liệu; nếu không dấu phân cách trường con sẽ bị coi như một phần tử của từ.

FST			ĐẦU RA TẠO BỞI FORMAT	CÁC PHẦN TỬ ĐƯỢC TẠO RA
2	4	mhl,v2	Hồn tôi đôi cánh	Hồn tôi đôi cánh

Các kỹ thuật đảo mới

Từ phiên bản CDS/IIS 3.08 và trong WinISIS bạn có thể sử dụng 4 kỹ thuật đảo mới cho phép xác định tiền tố cho mỗi thuật ngữ được tách ra từ các kỹ thuật đảo 1, 2, 3, 4. Các kỹ

thuật này được đánh số tương ứng là 5, 6, 7, 8. Tiền tố được xác định trong format tách dữ liệu có dạng như sau:

'dp...pd', [format]

trong đó d là dấu phân cách (delimiter) do bạn chọn (nó sẽ không xuất hiện trong tiền tố), p...p là tiền tố hiện đang sử dụng.

Ví dụ:

ID	IT	FORMAT TÁCH DỮ LIỆU
103	5	"Xuất bản=\',mhl,v3
102	8	"Nhan đề có từ:\',mhl,v2
107	6	"Từ khoá:\',mpl,v7
106	7	"Từ khoá:\',mpl,v6

Cụ thể với dòng FST:

102 8 "Nhan đề có từ:\',mhl,v2

và giả sử một biểu ghi có nhan đề:

Hồn tôi đôi cánh

thì kỹ thuật đảo 8 sẽ cho phép đặt tiền tố "Nhan đề có từ" trước từng từ của nhan đề và các phần tử dữ liệu được tạo ra sẽ là:

Nhan đề có từ: Hồn

Nhan đề có từ: tôi

Nhan đề có từ: đôi

Nhan đề có từ: cánh

Khi đó nếu trong biểu thức tìm tin ta viết một trong các dòng trên ta sẽ tìm được cuốn sách có nhan đề "Hồn tôi đôi cánh".

Chương 8

XÂY DỰNG VÀ KHAI THÁC CƠ SỞ DỮ LIỆU BẰNG CHƯƠNG TRÌNH WINISIS

1. XÂY DỰNG CẤU TRÚC MỘT CSDL MỚI BẰNG CHƯƠNG TRÌNH WINISIS

Để xây dựng cấu trúc một CSDL bằng CDS/ISIS người sử dụng trước hết cần chuẩn bị trước trên giấy các thông tin cần xử lý.

Giả sử ta cần xây dựng cơ sở dữ liệu có tên là TEST.

- Bảng xác định trường của CSDL này có 7 trường cùng với các đặc tính của chúng được cho trong bảng dưới đây:

Bảng xác định trường (FDT)				CSDL: TEST
Nhãn	Tên trường	Kiểu	Lặp	Dấu pc/mẫu
1	Tác giả	Alphanumeric	R	
2	Tên sách	Alphanumeric		
3	Xuất bản	Alphanumeric		abc
4	Số trang	Alphanumeric		
5	Ký hiệu	Alphanumeric		

6	Tóm tắt	Alphanumeric		
7	Từ khoá	Alphanumeric		

- Dự kiến trước biểu mẫu nhập tin và format trình bày dữ liệu. Giả sử biểu mẫu nhập tin bao gồm tất cả các trường với thứ tự như trên và format hiện hình dự kiến là:

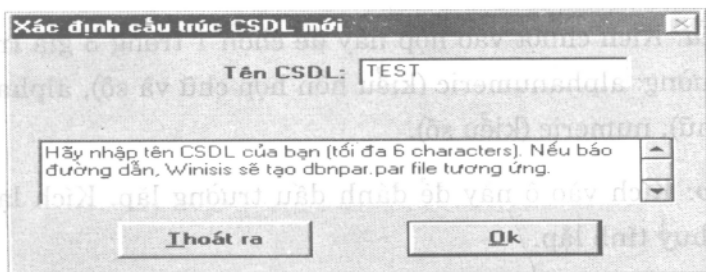
```
mfn(4),mdl,{f26,b,'-',v1+|;|}/{m(900,0),b,v2,i,'-',v3}/m(900,0),v4,'-',v5)/{"Tóm tắt:", m(900,0),v6)/{"Từ khoá:", mpl, m(900,0),v7}/#
```

- Chọn các trường dùng để đảo, kỹ thuật đảo và xác định format tách dữ liệu để lập bảng chọn trường (FST). Giả sử bảng chọn trường được thiết lập như sau:

Tên CSDL: TEST		Tên FST: TEST
ID	IT	Format tách dữ liệu
1	0	mhl,(v1)
2	4	mhl,v2
3	0	mhl,v3^a/v3^b/v3^c
6	3	mpl,v6
7	2	mpl,v7

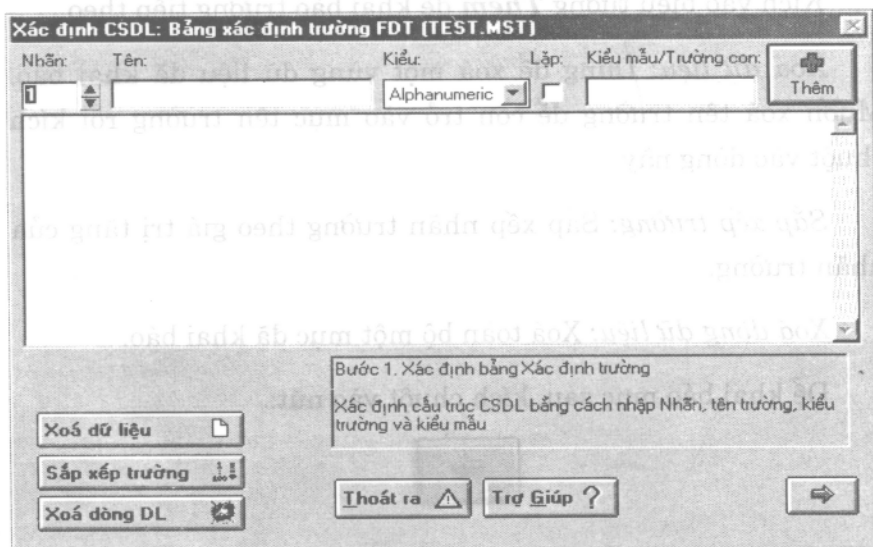
Quy trình tạo lập cấu trúc của CSDL bao gồm các bước liên tục sau đây:

a. Tên CSDL. Trong menu *Cơ sở dữ liệu* chọn mục *Tạo CSDL mới* (kích chuột vào dòng ký tự này). Xuất hiện hộp thoại *Xác định cấu trúc CSDL mới*.



Tại đây ta khai báo tên cơ sở dữ liệu (TEST) vào ô Tên CSDL rồi kích chuột vào nút OK để chấp nhận.

b. Khai báo FDT. Sau khi khai báo tên CSDL sẽ xuất hiện cửa sổ khai báo bảng xác định trường như sau:



Nhãn: nhãn trường là số bất kỳ từ 1 đến 999, hệ thống tự động đưa ra số 1.

Tên: Tên trường cần khai báo.

Kiểu: Kích chuột vào hộp này để chọn 1 trong 3 giá trị của kiểu trường: alphanumeric (kiểu hỗn hợp chữ và số), alphabetic (kiểu chữ), numeric (kiểu số).

Lập: Kích vào ô này để đánh dấu trường lập. Kích lại lần nữa sẽ huỷ tính lập.

Kiểu mẫu / trường con: Khai báo dấu phân cách trường con vào ô này nếu trường có trường con, ví dụ: với trường xuất bản ta khai báo abc, ứng với ba trường con: nơi xuất bản, nhà xuất bản, năm xuất bản.

Kích vào biểu tượng **Thêm** để khai báo trường tiếp theo.

Xoá dữ liệu: Dùng để xoá một vùng dữ liệu đã khai báo. Muốn xoá tên trường để con trở vào mục tên trường rồi kích chuột vào dòng này.

Sắp xếp trường: Sắp xếp nhãn trường theo giá trị tăng của nhãn trường.

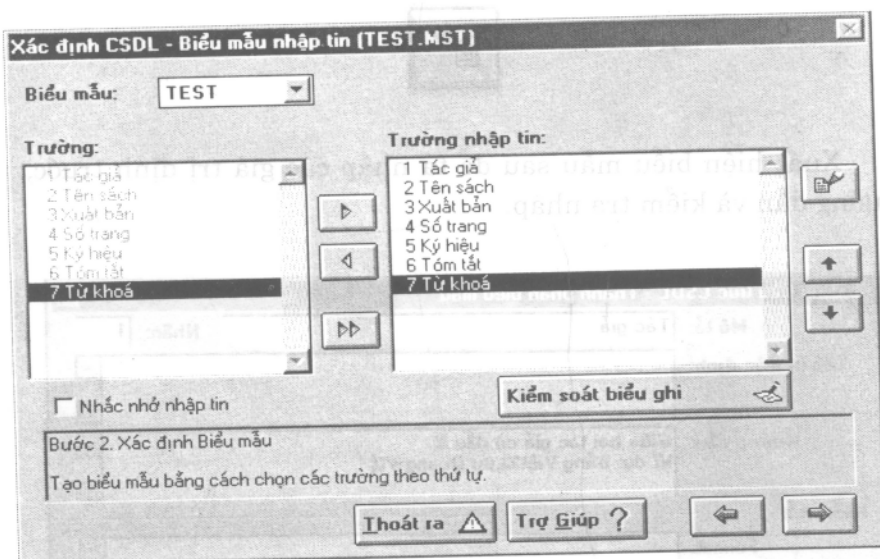
Xoá dòng dữ liệu: Xoá toàn bộ một mục đã khai báo.

Để khai báo mục sau, kích chuột vào nút:



c. Khai báo biểu mẫu nhập dữ liệu (Worksheet)

Khi kích chuột vào nút mũi tên $\Rightarrow$, trên màn hình hiện ra của sổ **Biểu mẫu nhập tin** sau:



Tại đây ta có thể chọn các trường để thiết kế biểu mẫu nhập tin bằng cách kích vào biểu tượng mũi tên ($\Rightarrow$) để chuyển một trường đã được đánh dấu ở hộp **Trường** bên trái sang hộp **Trường nhập tin** ở bên phải. Muốn huỷ bỏ một trường nào đó trong biểu mẫu thì dùng con trỏ đánh dấu trường đó rồi kích vào biểu tượng mũi tên ngược ($\Leftarrow$) để chuyển trường từ hộp phải sang trái.

Muốn đưa tất cả các trường vào biểu mẫu nhập dữ liệu chỉ cần kích vào biểu tượng mũi tên kép:



Để thiết lập thông tin trợ giúp cho việc nhập dữ liệu, ta đánh dấu vào ô trong dòng **Nhắc nhở nhập tin**, sau đó kích vào biểu tượng:



Xuất hiện biểu mẫu sau để ta nhập các giá trị định trước, hướng dẫn và kiểm tra nhập.

Xác định Cấu trúc CSDL - Thành phần biểu mẫu

Mô tả: Tác giả Nhãn: 1

Giá trị mặc định:

Hướng dẫn: Giữa hai tác giả có dấu %
Ví dụ: Bảng Việt%Lưu Quang Vũ

Kiểm tra

Format:

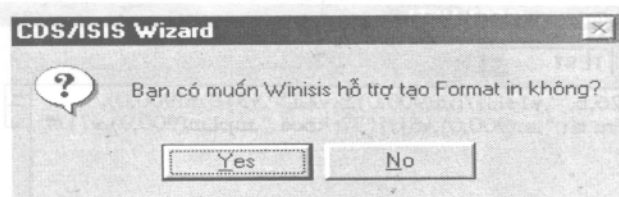
A

Thoát ra Trợ Giúp Ok

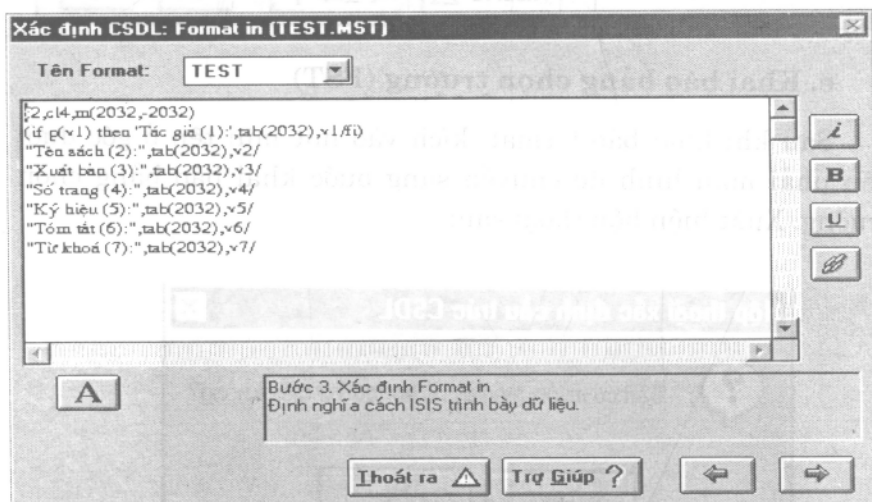
Nhấn OK để lưu.

d. Khai báo format. Sau khi khai báo biểu mẫu nhập dữ liệu, ta nhấn chuột vào biểu tượng mũi tên ($\Rightarrow$) để chuyển sang bước thiết kế format.

Khi đó WinISIS sẽ đưa ra hộp thoại yêu cầu ta trả lời xem có cần hệ thống hỗ trợ việc tạo format hay không.

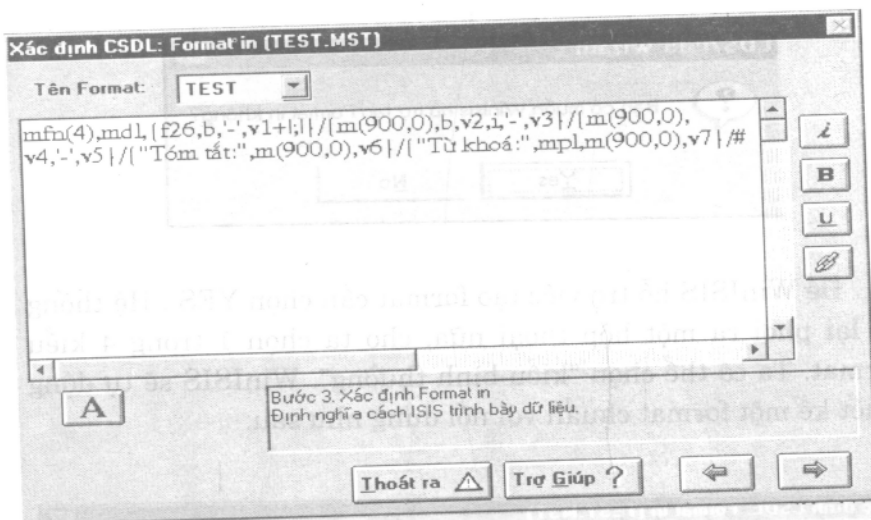


Để WinISIS hỗ trợ việc tạo format cần chọn YES . Hệ thống sẽ lại phôi ra một hộp thoại nữa, cho ta chọn 1 trong 4 kiểu format. Ta có thể chọn "kiểu bình thường". WinISIS sẽ tự động thiết kế một format chuẩn với nội dung như sau:



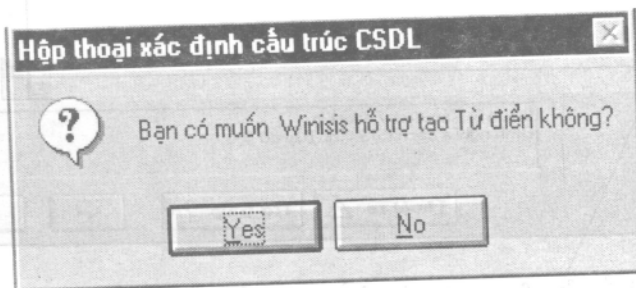
Ta có thể sửa format này bằng cách sử dụng Menu **Nhập tin**, mục **Format in/trình bày**.

Nếu chọn NO, hệ thống sẽ đưa ra cửa sổ **Format in** để ta gõ format đã chuẩn bị trước vào. Ta sẽ có một format trình bày theo ý muốn.

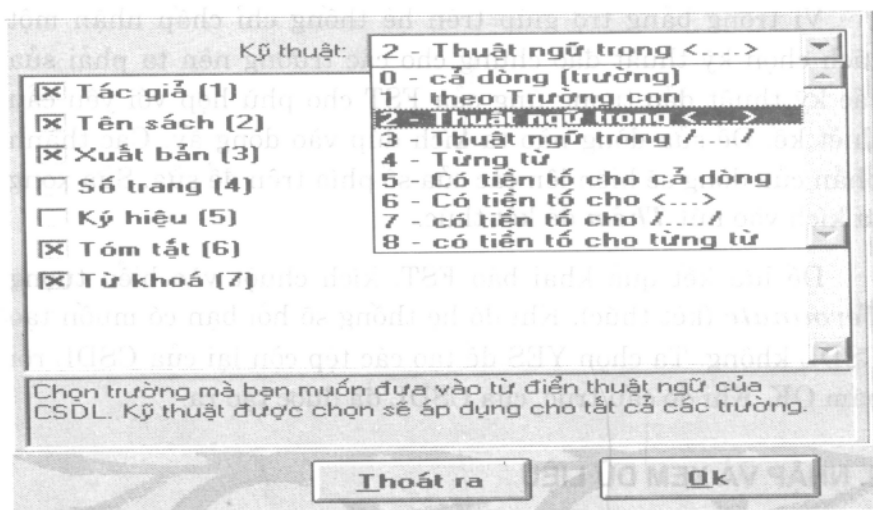


e. Khai báo bảng chọn trường (FST)

Sau khi khai báo format, kích vào nút mũi tên ở góc dưới bên phải màn hình để chuyển sang bước khai báo bảng chọn trường. Xuất hiện hộp thoại sau:

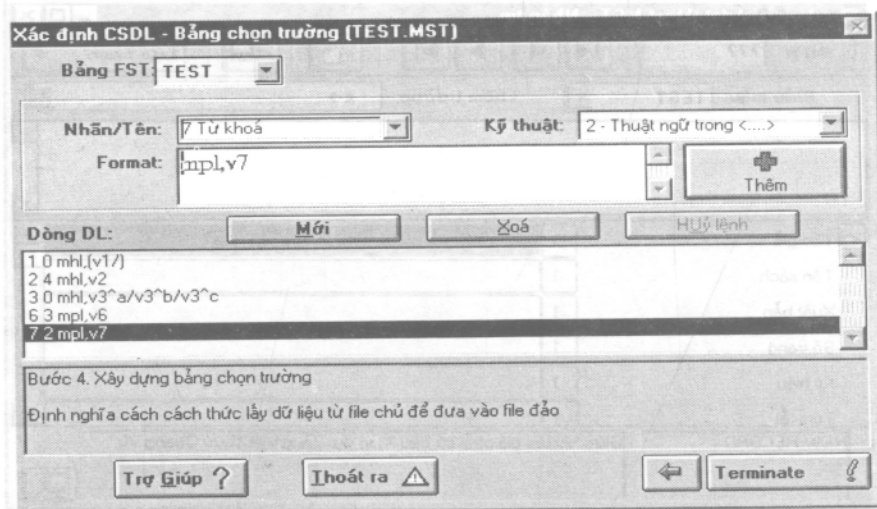


Để WinISIS hỗ trợ tạo FST, cần chọn YES. Hệ thống đưa ra màn hình khai báo dạng:



Ta có thể chọn các trường đánh chỉ số phía bên trái và kiểu đánh chỉ số phía bên phải bằng cách đánh dấu vào các mục tương ứng. Kết thúc bằng ấn nút OK.

Xuất hiện cửa sổ:



Vì trong bảng trợ giúp trên hệ thống chỉ chấp nhận một cách chọn kỹ thuật đảo chung cho các trường nên ta phải sửa các kỹ thuật đảo tương ứng của FST cho phù hợp với yêu cầu thiết kế. Để sửa dòng nào ta kích đúp vào dòng ấy. Các thành phần của dòng sẽ hiện lên các cửa sổ phía trên để sửa. Sửa xong ta kích vào nút **Thêm** để kết thúc.

Để lưu kết quả khai báo FST, kích chuột vào biểu tượng **Terminate** (kết thúc). Khi đó hệ thống sẽ hỏi bạn có muốn tạo CSDL không. Ta chọn YES để tạo các tệp còn lại của CSDL rồi bấm OK. Khi đó cấu trúc của CSDL đã được tạo ra.

2. NHẬP VÀ XEM DỮ LIỆU

2.1. Nhập, hiệu đính dữ liệu

Khi chọn lệnh **nhập dữ liệu** của **Menu Nhập tin**, cửa sổ nhập - hiệu đính sẽ hiện ra:

Tác giả	1
Tên sách	1
Xuất bản	1
Số trang	1
Ký hiệu	1
Tóm tắt	1

NEW RECORD

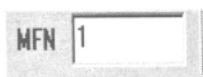
Giữa hai tác giả phải có dấu %, ví dụ: Bảng Việt%Lưu Quang Vũ

Khi bắt đầu mở, cửa sổ này đưa ra biểu ghi hiện tại. Tuy nhiên, nội dung của nó có thể sửa đổi bằng cách chọn các lệnh thích hợp trong menu **Xem dữ liệu**. Các trường của biểu ghi hiện tại được trình bày ở phần dưới của cửa sổ này theo worksheet nhập dữ liệu hiện tại.

Thanh công cụ của cửa sổ nhập dữ liệu gồm các mục sau:



Công tắc cho phép đưa ra hoặc huỷ bỏ thanh công cụ thứ hai.



Trường này chứa số MFN. Gõ vào đây số MFN mà ta muốn hiệu đính.



Hiện hình biểu ghi đầu để hiệu đính. Nếu ta đang hiệu đính kết quả tìm tin thì biểu ghi đầu tiên trong danh mục kết quả tìm được đưa ra.



Chọn biểu ghi trước đó để hiệu đính.



Chọn biểu ghi sau đó để hiệu đính.



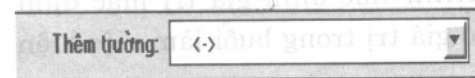
Chọn biểu ghi cuối cùng để hiệu đính.



Tạo biểu ghi mới



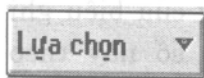
Chọn biểu mẫu nhập dữ liệu



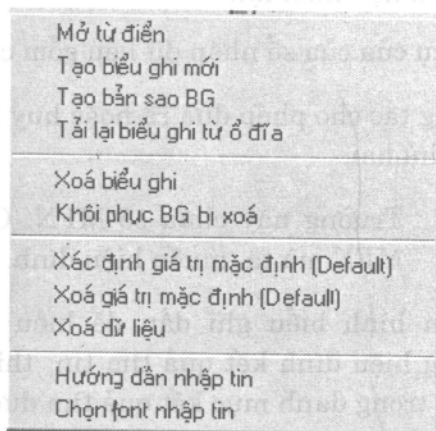
Thêm trường vào biểu ghi. Kích vào đây, danh sách các trường hiện tại trong biểu mẫu nhập sẽ được trình bày.



Lưu biểu ghi đã nhập.



Các lựa chọn bổ sung: Kích vào đây sẽ hiện ra menu phụ sau:



Mở từ điển: Mở cửa sổ từ điển.

Tạo biểu ghi mới: Worksheet hiện tại sẽ được đưa ra với các trường trống.

Tải biểu ghi từ đĩa: Huỷ bỏ mọi thay đổi và phục hồi các biểu ghi theo trạng thái ban đầu.

Xóa biểu ghi: Xóa biểu ghi hiện tại (về mặt logic).

Phục hồi biểu ghi: Phục hồi biểu ghi bị xóa (về mặt logic).

Xác định giá trị mặc định: Xác định giá trị mặc định (default). Các giá trị này chỉ có giá trị trong buổi làm việc hiện tại.

Xóa giá trị mặc định: Xóa các giá trị mặc định (default). Chỉ xóa các giá trị trong buổi làm việc hiện tại.

Xoá dữ liệu: Xoá mọi nội dung của các trường trong Worksheet hiện tại. Ta có thể sử dụng chức năng này để thay thế một biểu ghi hiện có bằng một biểu ghi mới có cùng một MFN.

Hướng dẫn nhập tin: Đưa ra thông tin trợ giúp để nhập dữ liệu cho chính xác.

Chọn font nhập tin: Cho phép thay đổi font hoặc cỡ chữ được sử dụng trong hiệu đính trường.

Thanh trạng thái ở đáy của sổ chứa hai trường:

Trạng thái biểu ghi: Đưa ra thông báo về trạng thái biểu ghi hiện tại. Chỉ có thông báo trong trường hợp biểu ghi bị xoá.

Thông báo hướng dẫn: Đưa ra thông báo hướng dẫn liên quan đến trường đang được hiệu đính, nếu có.

Dưới đây là các thao tác cập nhật dữ liệu

a- Hiệu đính trường

Để hiệu đính một trường ta phải kích vào vùng của trường đó. Nội dung của nó sẽ xuất hiện trên hộp hiệu đính trường ở phía trên. Để hiệu đính ta có thể sử dụng mọi phím chức năng của Windows, ngoài ra còn sử dụng các phím sau:

Enter: Cập nhật trường và sang trường tiếp theo.

Shift+Enter: Cập nhật trường và chuyển về trường trước đó.

F2: Xoá trường.

ESC: Huỷ bỏ mọi sửa đổi.

Nguyên tắc nhập một số trường:

Trường lập: Nếu là trường lập thì mỗi phần tử dữ liệu phải cách nhau bởi dấu %; không có ký tự trống trước và sau dấu %.

Ví dụ: Với trường tác giả có hai tác giả:

Bằng Việt%Lưu Quang Vũ

Trường con: Nếu trường có trường con, thì mỗi giá trị của trường con phải được đặt trước bởi dấu phân cách trường con. Dấu phân cách trường con là một mã gồm hai ký tự: dấu ^ và một chữ thường tiếp theo, ví dụ: ^a, ^b,...

Ví dụ: Với trường xuất bản có ba trường con:

^aHà Nội^bVăn học^c1975

Dấu hạn chế thuật ngữ tìm: Dấu hạn chế thuật ngữ tìm được dùng để nhận dạng các từ khoá được dùng làm điểm truy nhập thông tin trong biểu ghi. Đối với trường dùng kỹ thuật 3 để đảo, ví dụ trường “Tóm tắt” thì từ khoá được đặt giữa hai dấu /.../. Đối với trường dùng kỹ thuật 2 để đảo, ví dụ trường “Từ khoá” thì từ khoá được đặt giữa hai dấu <...>.

b- Bổ sung thông tin vào một trường

Chọn trường mà ta muốn bổ sung thông tin từ danh mục **Thêm trường**. Chú ý rằng ta không thể thêm giá trị thứ hai vào một trường nếu trường đó không phải là trường lập. Nếu một vài giá trị của trường lập đã có thì giá trị mới sẽ được bổ sung thêm vào.

c- Cập nhật file đảo

Nếu trong file SYSPAR.PAR tham số 130 nhận giá trị là 0 thì WinISIS không tự động cập nhật file đảo mỗi khi ta cập nhật biểu ghi. Nếu biểu ghi chưa được cập nhật thì:

- Biểu ghi mới nhập vào sẽ không có hiệu lực khi tìm kiếm.
- Biểu ghi đã sửa có thể tìm kiếm nhưng theo điểm truy nhập cũ.
- Biểu ghi đã xóa vẫn còn được định vị bởi điểm truy nhập của nó, mặc dầu biểu ghi không còn tồn tại.

Do đó sau mỗi lần cập nhật biểu ghi (thêm, xóa, sửa) ta phải cập nhật file đảo. Để cập nhật file đảo ta phải chọn lệnh **File đảo** trong **menu CSDL**. Khi đó hộp thoại **Bảo trì file đảo** sẽ hiện ra:

Bảo trì File đảo

Lựa chọn

☒ Cập nhật ☐ Tạo files liên kết

☐ Tạo File đảo hoàn chỉnh ☐ Sắp xếp files liên kết

☐ Khởi tạo lại File đảo ☐ Tải files liên kết

Từ MFN đến MFN

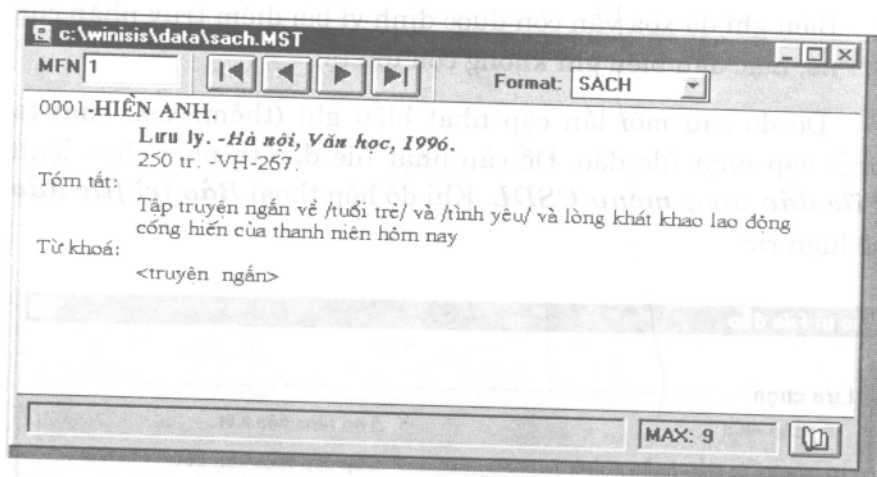
Ok Thoát ra Trợ Giúp

Trong hộp thoại trên ta có thể lựa chọn các phương án tạo file đảo: **cập nhật**, **tạo file đảo hoàn chỉnh**. Khi cần thiết ta

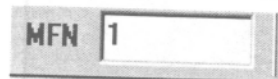
có thể phải **khởi tạo lại file đảo** và khai báo giới hạn của biểu ghi.

2.2. Xem dữ liệu

Mỗi khi mở một CSDL của sổ CSDL sẽ hiện ra. Biểu ghi hiện tại được đưa ra theo format trình bày đã chọn. Nội dung của cửa sổ này có thể thay đổi bằng cách lựa chọn các lệnh thích hợp trong menu **Xem dữ liệu**.



Thanh công cụ của cửa sổ này có các mục sau:



Trường này chứa số MFN của biểu ghi hiện tại. Kích chuột vào đây và gõ vào số MFN mà ta muốn xem.



Hiện hình biểu ghi đầu tiên của CSDL hoặc kết quả tìm tin.



Chọn biểu ghi trước đó để xem.



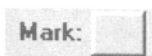
Chọn biểu ghi sau đó để xem.



Chọn biểu ghi cuối cùng để xem.



Chọn format trình bày bằng cách kích chuột vào format tương ứng.



Đánh dấu/hủy đánh dấu các biểu ghi trong kết quả tìm tin. Nhấn vào nút này ta sẽ đánh dấu biểu ghi hiện tại (chữ “V” sẽ hiện ra trên nút). Sau khi đánh dấu ta có thể tìm được biểu ghi đã đánh dấu bằng cách chọn lệnh **Chỉ biểu ghi đánh dấu** trong menu **Xem dữ liệu**.

Thanh trạng thái ở dưới cửa sổ này chứa 2 trường:

- **Trạng thái biểu ghi:** Trong trường hợp biểu ghi bị xóa, trường này sẽ đưa ra thông báo “biểu ghi bị xóa về mặt logic”. Trong trường hợp xem kết quả tìm thì ở dòng này sẽ hiện ra thông báo tựa như sau:

Câu #5: [1/8]

Điều đó có nghĩa là: ta đang xem biểu ghi thứ nhất trong 8 biểu ghi tìm được, khi thực hiện câu hỏi tìm thứ 5.

- **Số MFN tối đa:** Đưa ra số biểu ghi tối đa của tệp chủ trong CSDL.

Khi một CSDL đang được mở, ta có thể sử dụng các nút sau đây của thanh công cụ:



Mở CSDL khác.



Xoá nội dung Windows Clipboard.



Sao biểu ghi hiện tại sang Windows Clipboard.



Mở cửa sổ tìm tin.



Mở hộp thoại in để khai báo các tham số in.



Chia cửa sổ CSDL thành hai phần: một phần trình bày biểu ghi trong chế độ đồ hoạ, còn phần kia trong chế độ văn bản (ASCII format).



Trình bày format hiện hình của biểu ghi hiện tại trong chế độ văn bản, cho phép sửa format hiện hình.



Mở cửa sổ nhập dữ liệu.

In biểu ghi đang xem

Trong khi xem dữ liệu, ta có thể in biểu ghi đang xem mà không cần vào **Hộp thoại in** để khai báo các tham số in. Ta cần thực hiện các thao tác sau:

- Chọn biểu ghi trong chế độ xem (biểu ghi hiện tại trên màn hình).
- Kích chuột vào menu **Xem dữ liệu**, chọn mục **In trang hiện tại**.

Chương trình sẽ in biểu ghi đang xem ra máy in theo format hiện tại đang dùng.

3. TÌM TIN TRONG WINISIS

3.1. Nguyên tắc tìm tin theo tệp đảo

Để tìm tin trên các trường theo tệp đảo, ta cần đưa vào một biểu thức tìm, còn gọi là phương trình tìm. Đó là một biểu thức logic bao gồm các thuật ngữ tìm (từ khoá) được liên kết bởi các phép toán của ngôn ngữ tìm tin do chương trình cung cấp.

Các toán tử tìm sẽ liên kết các thuật ngữ tìm để tạo thành biểu thức tìm. Các thuật ngữ tìm được lưu trong từ điển. Ta có thể đưa vào biểu thức tìm:

- hoặc một thuật ngữ chính xác, ví dụ: Xuân Diệu, thông tin học,...
- hoặc một thuật ngữ bị chặn phải bằng **toán tử chặt cụt \$**

Ví dụ: **developpe\$** sẽ bao gồm:

developpement

developpement rural

developpement industriel

tức là tất cả các thuật ngữ bắt đầu bằng **developpe**.

Ví dụ: **197\$** sẽ bao gồm những năm 70.

3.2. Ngôn ngữ tìm tin WinISIS

Ngôn ngữ tìm tin WinISIS (The WinISIS Search Language) được xây dựng trên cơ sở của logic Bool, nó cung cấp một phương tiện thuận tiện để trình bày các phép toán logic giữa các lớp toán hạng.

Mỗi thuật ngữ tìm liên quan tới một biểu ghi, thực chất có thể xem như là đại diện của một lớp các biểu ghi có liên quan đến thuật ngữ đó. Do đó bằng cách sử dụng các phép toán logic liên kết các thuật ngữ tìm, tạo thành một biểu thức tìm, ta có thể xác định một cách chính xác lớp các biểu ghi thoả mãn yêu cầu tìm tin đặt ra.

Ngôn ngữ tìm tin WinISIS bao gồm các toán tử logic và một số toán tử khác.

Ba toán tử logic cơ bản là:

AND	ký hiệu là	*
OR	ký hiệu là	+
NOT	ký hiệu là	^

Ý nghĩa của các toán tử này là:

$A*B$ là nhóm các tài liệu được đánh chỉ số đồng thời bằng các thuật ngữ A và thuật ngữ B.

$A+B$ là nhóm các tài liệu được đánh chỉ số bằng thuật ngữ A hoặc thuật ngữ B, hoặc cả hai.

A^B là nhóm các tài liệu được đánh chỉ số chỉ bằng thuật ngữ A, nhưng không được đánh chỉ số bằng thuật ngữ B.

Các toán tử lân cận mức trường

Các toán tử này là dạng thu hẹp của toán tử AND và đặc biệt khi tìm theo ngôn ngữ tự nhiên. Chúng được sử dụng khi các giá trị của trường hoặc trường lặp được hiểu như một đơn nguyên.

Các toán tử tìm ở mức trường và lân cận là:

(G): Kết nối các thuật ngữ của cùng một trường

Ví dụ: Một tập thơ của Xuân Diệu có nhan đề là **Hồn tôi đôi cánh**, kỹ thuật đảo 4 sẽ tách trường nhan đề đó thành từng từ: **hồn, tôi, đôi, cánh**.

Khi đó biểu thức tìm:

Tôi (G) cánh

sẽ cho ta tìm được biểu ghi trên và tất cả các biểu ghi mà có hai từ **tôi** và **cánh** trong cùng một trường.

(F): Kết nối các thuật ngữ của cùng một trường hay các tình huống của một trường lặp.

Ví dụ: Tập thơ **Hương cây Bếp lửa** của hai tác giả Bằng Việt và Lưu Quang Vũ. Trường tác giả là trường lặp.

Khi đó biểu thức tìm:

Bằng Việt (F) Lưu Quang Vũ

sẽ cho ta tìm được đúng tập thơ của hai tác giả nói trên.

∴ Toán tử dấu chấm cũng như toán tử (F), nhưng thêm điều kiện là không có quá n từ giữa hai thuật ngữ, trong đó n là số lượng dấu chấm trừ đi 1.

Ví dụ: Các biểu thức tìm:

Hồn . tôi (hai từ **hồn** và **tôi** kề sát nhau)

Hồn . . đôi (có nhiều nhất một từ giữa **hồn** và **đôi**)

Hồn . . . cánh (có nhiều nhất hai từ giữa **hồn** và **cánh**)

đều cho ta tìm được tập thơ **Hồn tôi đôi cánh** của Xuân Diệu.

\$: Cũng như toán tử (F), nhưng thêm điều kiện là có đúng n từ giữa hai thuật ngữ, trong đó n là số lượng dấu \$ trừ đi 1.

Ví dụ: Các biểu thức tìm:

Hồn \$ tôi (hai từ **hồn** và **tôi** kề sát nhau)

Hồn \$ \$ đôi (có đúng một từ giữa **hồn** và **đôi**)

Hồn \$ \$ \$ cánh (có đúng hai từ giữa **hồn** và **cánh**)

đều cho ta tìm được tập thơ **Hồn tôn đôi cánh** của Xuân Diệu.

Chú ý: Trước và sau dấu chấm hoặc dấu \$ phải có một khoảng trống.

Cú pháp của biểu thức tìm

Ta có thể tạo ra một biểu thức tìm phức hợp bằng cách kết hợp hai hoặc nhiều thuật ngữ tìm với nhau bằng các toán tử. Khi trong biểu thức tìm có nhiều loại toán tử khác nhau thì WinISIS xử lý theo thứ tự mức ưu tiên giữa các toán tử đó. Trật tự ưu tiên của các phép toán là như sau:

(cao nhất)	\$ và .
	(F)
	(G
	* và ^
	+
(thấp nhất)	

Ta có thể dùng dấu ngoặc để thay đổi thứ tự thực hiện các phép toán.

Ví dụ: Với câu hỏi “Tìm các tập thơ của Xuân Diệu hoặc Tế Hanh và tiểu thuyết của Nguyễn Đình Thi” thì biểu thức tìm là:

Thơ*(Xuân Diệu+Tế Hanh)+Tiểu thuyết*Nguyễn Đình Thi

Trình tự thực hiện các phép toán của chương trình sẽ như sau:

P = 5	Xuân Diệu
P = 3	Tế Hanh
T = 8	#2: Xuân Diệu + Tế Hanh
P = 22	Thơ
T = 8	#3: Thơ * #2
P = 7	Tiểu thuyết
P = 6	Nguyễn Đình Thi
T = 3	#4:Tiểu thuyết*Nguyễn Đình Thi
T = 11	#5: #3 + #4
T = 11	#1: #5

trong đó: #1 là số hiệu của biểu thức tìm đã cho, còn #2, #3, #4, #5 là số hiệu được gán cho các biểu thức tìm trung gian.

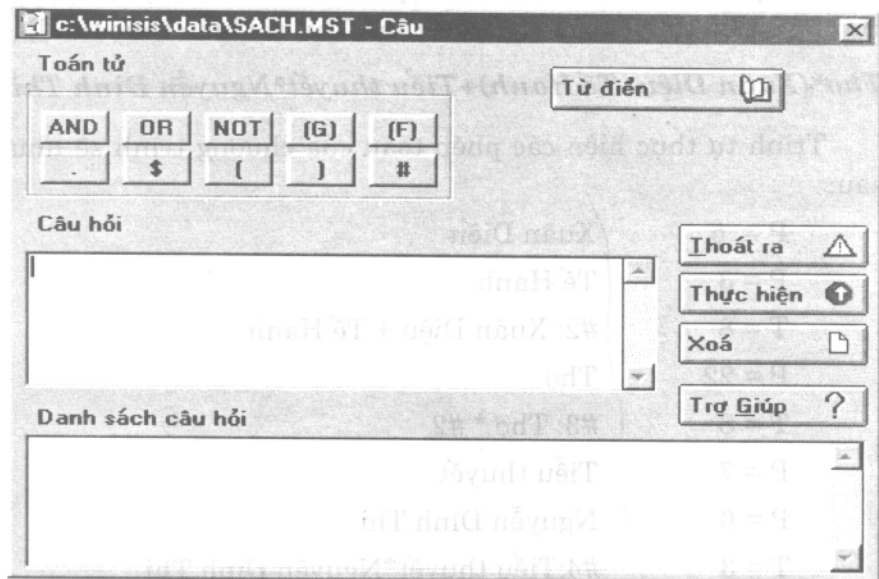
Kết quả là có 11 biểu ghi thoả mãn yêu cầu tìm tin nói trên.

3.3. Phương tiện tìm tin trong WinISIS

WinISIS cho phép tìm tin theo hai chế độ: Tìm tin trình độ cao và tìm tin có trợ giúp.

a- Tìm tin trình độ cao (Expert Search)

Kích chuột vào mục ***Tìm tin trình độ cao*** trong ***Menu tìm tin*** sẽ xuất hiện cửa sổ:



Toán tử: Nhấn chuột vào một trong các nút này ta sẽ chèn toán tử tương ứng vào hộp **câu hỏi**.

Câu hỏi: Hộp này chứa biểu thức tìm hiện tại. Người sử dụng có kinh nghiệm có thể gõ trực tiếp biểu thức tìm vào trong hộp. Số lượng ký tự tối đa trong biểu thức tìm là 1000.

Từ điển: Mở cửa sổ từ điển.

Thực hiện: Thực hiện yêu cầu tìm.

Xoá: Xoá biểu thức tìm hiện tại.

Trợ giúp: Đưa ra thông báo trợ giúp.

Danh sách câu hỏi: Liệt kê danh sách các biểu thức tìm đã được thực hiện, bao gồm số hiệu biểu thức tìm, tên CSDL, kết quả tìm và nội dung của biểu thức tìm.

Kích đúp chuột vào một trong những dòng của danh sách này, WinISIS sẽ đưa ra kết quả tương ứng trong cửa sổ CSDL.

Nếu nháy chuột đơn vào một dòng của danh sách thì số hiệu của biểu thức tìm sẽ được đưa vào hộp câu hỏi. Ta có thể thực hiện tìm lại theo câu hỏi đó, với điều kiện đó là câu hỏi của CSDL đang mở.

b- Tìm tin có trợ giúp (Guided Search)

Kích chuột vào mục ***Tìm tin có trợ giúp*** trong ***Menu tìm tin*** sẽ xuất hiện cửa sổ ***Tìm tin có trợ giúp***.

Thành phần tìm tin: Bốn hộp được đưa ra để ta gõ vào thuật ngữ tìm. Bạn có thể kích chuột vào nút ***Tủ điển*** để hiện ra cửa sổ từ điển, rồi từ đó bạn có thể chọn một thuật ngữ tìm thích hợp bằng cách kích đúp vào từ đó hoặc rê chuột để đưa thuật ngữ vào hộp tương ứng.

Trường tìm tin: Bằng cách chọn trường thích hợp trong danh sách các trường tìm tin, ta sẽ lấy ra được thuật ngữ tìm thuộc trường tương ứng.

Toán tử: Các thuật ngữ tìm có thể liên kết với nhau bằng một trong ba toán tử AND, OR, NOT bằng cách chọn toán tử thích hợp từ danh sách các toán tử.

Câu hỏi trước: Khi một hay nhiều biểu thức tìm trước đó đã được thực hiện, bạn có thể chọn một trong số đó từ danh sách các câu hỏi trước, rồi nối với các yếu tố tìm mới.

Các nút bấm: Để thực hiện một lệnh tìm ta bấm nút **Thực hiện**. Dòng văn bản của biểu thức tìm do WinISIS tạo ra sẽ hiện ra trong hộp **Câu hỏi**. Khi việc tìm kiếm đã hoàn thành, số hiệu của biểu thức tìm sẽ hiện ra trong hộp **Số câu hỏi** và tổng số các biểu ghi tìm được sẽ hiện ra trong hộp **Số biểu tìm được**.

Để hiện ra kết quả tìm, bạn bấm nút **Xem kết quả tìm**. Bấm nút **Xoá**, sẽ xoá các yếu tố tìm. Bấm nút **Thoát ra**, sẽ trở về cửa sổ CSDL, bỏ qua việc tìm kiếm.

3.4. Tìm tự do

Kỹ thuật tìm tự do của WinISIS cho phép xác định yêu cầu tìm theo các trường không được đánh chỉ số hoặc xác định những điều kiện mà ta không thể mô tả thông qua biểu thức tìm, như trường hợp so sánh các giá trị bằng số của các trường.

Tìm tự do thực hiện theo phương thức tìm tin trình độ cao.

Để phân biệt tìm tự do với chế độ tìm thông thường ta cần đặt dấu hỏi đứng trước biểu thức tìm như sau:

? Biểu thức tìm

hay

? #n Biểu thức tìm

trong đó:

? Chỉ ra rằng cần tìm tự do

#n Hạn chế việc tìm tự do trong kết quả của lệnh tìm trước đó, được thực hiện bởi câu hỏi có số hiệu là #n. Nếu không chỉ ra giới hạn này thì việc tìm tự do được thực hiện trong toàn bộ CSDL.

Biểu thức tìm: là biểu thức logic đã mô tả trong phần ngôn ngữ tìm của WinISIS.

Ví dụ:

? v1='Xuân Diệu' and val(v3^c)>1970

sẽ tìm các tập thơ của tác giả Xuân Diệu (trường 1) với năm xuất bản (V3^c) là sau năm 1970.

Ví dụ: Để tìm các tập thơ xuất bản sau năm 1970 trong một CSDL, ta dùng kỹ thuật tìm tự do để hạn chế thời gian trong câu hỏi “Thơ” đã tìm trước đó. Nếu câu hỏi đó có số hiệu là 4 thì biểu thức tìm sẽ là:

? #4 val(v3^c)>1970

Một ví dụ khác, biểu thức tìm:

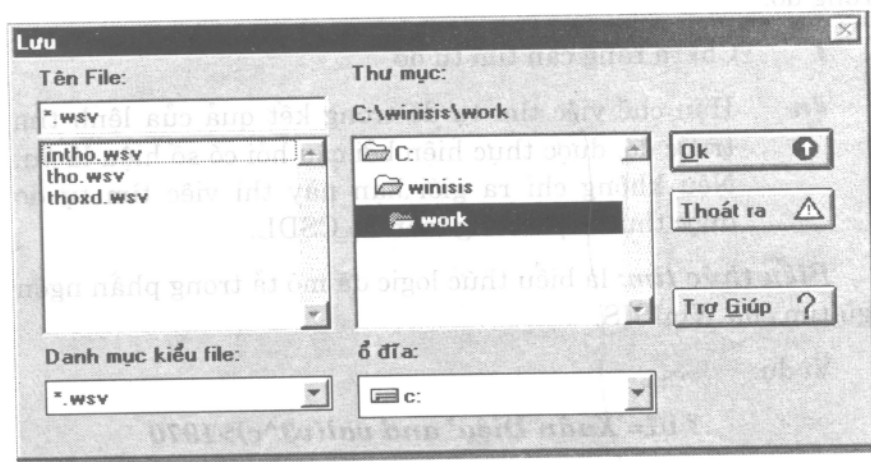
? val(v3^c)>1970

sẽ cho tất cả các tài liệu có trong CSDL xuất bản sau năm 1970.

3.5. Lưu kết quả tìm

Chọn mục **Lưu kết quả tìm** trong menu **Tìm tin**, một danh sách các câu hỏi sẽ hiện ra, cho phép ta lưu kết quả tìm

của một câu hỏi được lựa chọn trong danh sách đó. Chương trình sẽ yêu cầu ta đưa vào tên tệp lưu bằng phương tiện của một hộp thoại mở:



Bạn gõ vào mục **Tên file** tên tệp lưu. Tệp này có phần mở rộng là .wsv.

Khi cần ta có thể gọi tệp này để xem lại kết quả đã tìm. Muốn vậy ta chọn mục **Gọi lại kết quả tìm** trong menu **Tìm tin**. Sử dụng phương tiện in, ta có thể in kết quả đang xem ra giấy hoặc in ra một tệp văn bản.

4. PHƯƠNG TIỆN IN TRONG WINISIS

4.1. Nguyên tắc chung

Ta có thể in ra kết quả của một yêu cầu tìm tin hoặc in một tập các biểu ghi được chọn theo số hiệu của chúng. Trong trường hợp thứ nhất là in các biểu đang xem, còn trong trường hợp thứ

hai, ta cần chỉ ra một dãy số hiệu của các biểu ghi cần in. Ta có thể in các kết quả theo một trật tự sắp xếp tùy ý. Các kết quả có thể in ra máy in hay in ra một tệp lưu trên đĩa. Việc trình bày kết quả in được xác định bởi format in. Format này được lập từ trước hay được xác định khi đưa vào các tham số in.

Mỗi khi sử dụng phương tiện sắp xếp in, hệ thống sẽ yêu cầu cung cấp các tham số trình bày và sắp xếp cần thiết để in. Các tham số này được đưa vào nhờ các trang của **Hộp thoại in**.

Để thực hiện in dữ liệu bằng chương trình WinISIS ta chọn mục **In ấn** trong **Menu CSDL**. Trên màn hình sẽ xuất hiện **Hộp thoại in**.

Hộp thoại in gồm 5 trang màn hình:

Trang chung : Cho phép chọn in cái gì, in sang đâu (tệp hay máy in) và format in.

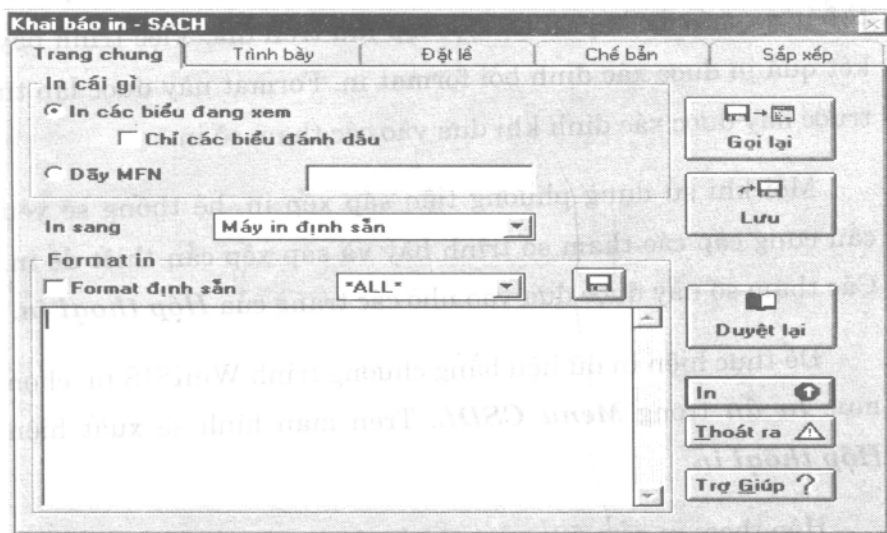
Trình bày: Xác định nhan đề, tiêu đề và phụ chú cuối trang của bản in.

Đặt lề: Xác định vùng tin của trang.

Chế bản: Xác định cách trình bày trang, đánh số trang và số lượng cột in.

Sắp xếp: Xác định tiêu chuẩn sắp xếp.

4.2. Hộp thoại in (Trang chung)



a- In cái gì

In các biểu đang xem: Cho phép in tập hợp các biểu ghi mà ta đã xem trước khi mở hộp thoại in. Nếu ta đã xem và đánh dấu các kết quả tìm kiếm, ta có thể chọn mục **Biểu ghi đánh dấu** để in các biểu ghi đã được đánh dấu.

Dãy MFN: Bạn hãy gõ vào đây giới hạn các biểu ghi cần in, theo quy định sau:

- | | |
|-----|---|
| n | In chính biểu ghi thứ n |
| -n | In biểu ghi từ thứ nhất đến thứ n |
| n- | In từ biểu ghi thứ n đến hết |
| n-m | In từ biểu ghi thứ n đến biểu ghi thứ m |

Ta có thể nối các phạm vi in bằng cách sử dụng dấu phẩy.
ví dụ:

-10,15,20-30,50-

sẽ in 10 biểu đầu, biểu ghi thứ 15, biểu ghi từ 20 đến 30 và các biểu ghi từ 50 trở đi.

In sang: Sau khi chọn in cái gì, bạn phải chọn thiết bị đầu ra, WinISIS cho các khả năng sau:

Printer: In sang máy in ngầm định hiện tại.

Postscript File: Đầu ra sẽ được lưu trữ trong một file có tên output.ps trong thư mục \WinISIS\Work\.

ASCII File(ANSI): Đầu ra sẽ được lưu trữ trong một file văn bản theo format ASCII sử dụng tập ký tự của Windows.

ASCII File(OEM): Đầu ra sẽ được lưu trữ trong một file văn bản theo format ASCII sử dụng tập ký tự của MSDOS.

Khi in ra file, trước khi chạy máy in WinISIS sẽ yêu cầu bạn đặt tên cho file đầu ra bằng phương tiện của một **Hộp thoại mở**.

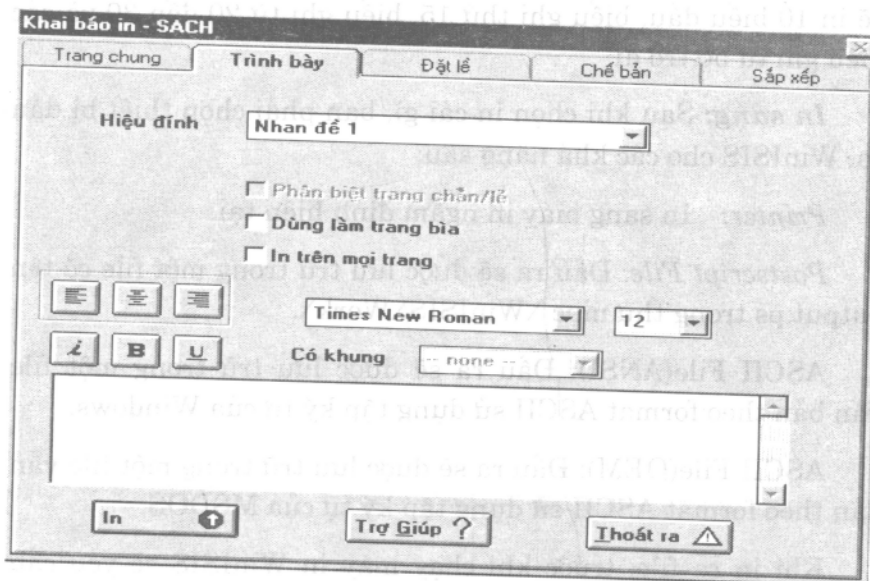
Format in: Nếu đánh dấu vào mục **Format định sẵn**, bạn có thể chọn tên một format in thích hợp trong hộp danh sách các format của CSDL. Version này của chương trình CDS/ISIS cho phép bạn sửa format in hoặc viết một format in mới trước khi in. Sau đó nếu bạn muốn lưu format đã sửa thì bạn nhấn vào nút **Lưu**.

Nút gọi lại: Nhấn nút này sẽ tìm lại worksheet in được lưu trước đó.

Nút lưu: Sẽ cất giữ tham số thiết lập worksheet in hiện tại.

Nút duyệt: Nhấn nút này sẽ cho xem trước trang đầu của bản in sẽ được in ra.

4.3. Hộp thoại in (Trang trình bày)



Sử dụng trang này khi ta muốn tài liệu in ra có nhan đề. Ở dòng trên cùng của hộp danh sách các nhan đề sẽ đưa ra các loại nhan đề mà bạn chọn để soạn thảo, có thể là: nhan đề 1, nhan đề 2, nhan đề 3, Header, Footer.

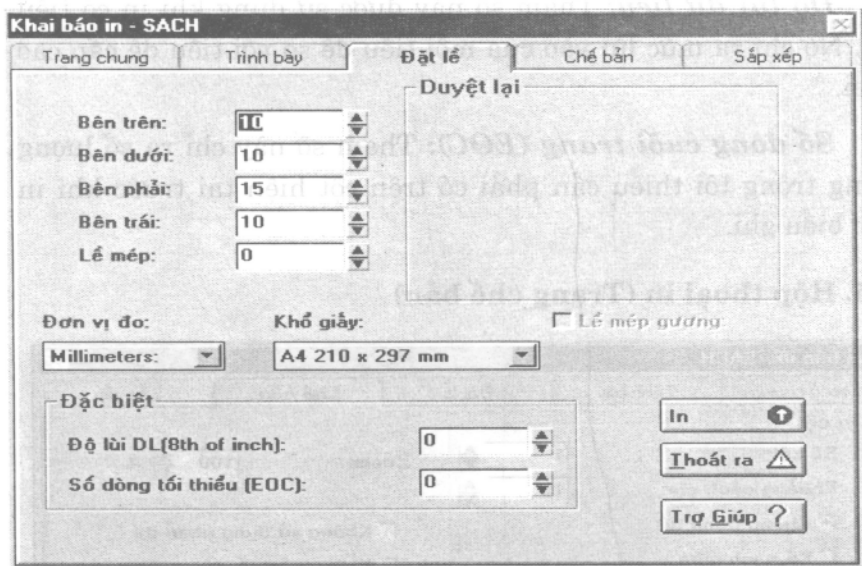
Nội dung thực sự của nhan đề được gõ vào phần ô để trống của cửa sổ. Đối với mỗi nhan đề bổ sung vào bản in bạn cần chọn:

Dùng làm trang bìa: In nhan đề trên trang bìa.

In trên mọi trang: In nhan đề trên mọi trang. Điều này chỉ cho phép với 3 nhan đề.

Các chức năng soạn thảo: Font chữ, kiểu chữ, cỡ chữ, căn dòng cho dòng nhan đề.

4.4. Hộp thoại in (Trang đặt lề)



Trong trang này ta có thể xác định lề cho trang in và các tham số liên quan.

Lề: Cũng giống các trang soạn thảo văn bản có:

Đỉnh: Khoảng trống từ đỉnh trang đến dòng in đầu tiên

Đáy: Khoảng trống từ đáy trang đến dòng in cuối cùng

Bên trái: Lề trái

Bên phải: Lề phải

Gutter: Khoảng trống bổ sung cho các lề biên của tài liệu

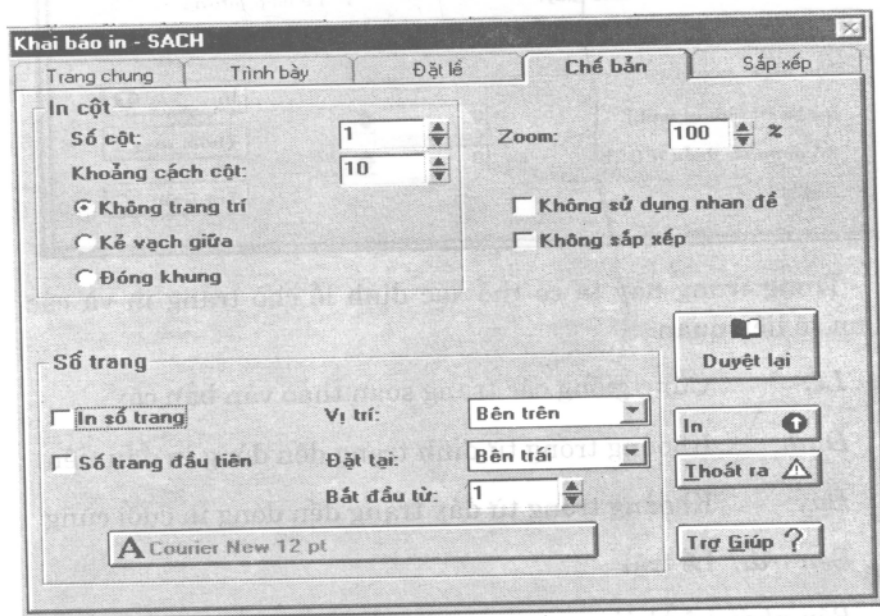
Đơn vị đo: Xác định đơn vị đo được dùng cho các loại lề, có thể là Millimet, Centimet, Inch, Twips.

Khổ giấy: Xác định khổ giấy.

Độ lùi dữ liệu: Tham số này được sử dụng khi in có tiêu đề. Nó chỉ ra mức lùi vào của mỗi tiêu đề so với tiêu đề cấp cao hơn.

Số dòng cuối trang (EOC): Tham số này chỉ ra số lượng dòng trống tối thiểu cần phải có trên cột hiện tại trước khi in các biểu ghi.

4.5. Hộp thoại in (Trang chế bản)



Trong trang này ta có thể lựa chọn các tham số liên quan đến việc chế bản trang in:

Số cột: Số lượng cột trên một trang.

Khoảng cách cột: Khoảng trống giữa các cột.

Không trang trí: Không có các trang trí bổ sung (như đường gạch đứng, đường in, khung của các cột) được in ra trên cột in.

Kẻ vạch giữa: Đường thẳng đứng giữa các cột.

Đóng khung: WinISIS sẽ tạo ra khung viền xung quanh mỗi cột.

Số trang: Chọn hộp này cho phép in ra số trang.

Số trang đầu tiên: Chọn mục này sẽ cho in số trang của trang đầu. Trang đầu không phải là trang bìa (nếu ta sử dụng nhan đề cho trang bìa).

Vị trí: Xác định vị trí số trang, trên đầu trang hay cuối trang.

Đặt tại: Vị trí số trang: bên trái, bên phải hoặc giữa.

Bắt đầu từ: Số của trang đầu tiên.

Zoom: Phóng to thu nhỏ từ 50% đến 120%. Sẽ tác động đến format in. Vì thế có thể bố trí biểu ghi trên một trang mà không cần thay đổi format in.

Không sử dụng nhan đề: Nhắc chương trình bỏ qua trang nhan đề.

Không sắp xếp: Nhắc chương trình bỏ qua việc sắp xếp.

4.6. Hộp thoại in (Trang sắp xếp)

Trang này chứa tất cả các tham số liên quan đến sắp xếp và các tiêu đề. Ta có thể sắp xếp theo bất cứ trường nào. Trong thực tế việc sắp xếp không liên quan tới tệp đảo.

Ví dụ để sắp xếp theo tác giả (mức sắp 1) rồi theo tên sách (mức sắp 2) ta phải chọn hai khoá sắp xếp. Nếu ta chỉ muốn làm nổi bật khoá thứ nhất thì số tiêu đề sẽ là 1 và kết quả in ra có thể như sau:

Tác giả 1

Tên sách 1

Tên sách 2

Tác giả 2

Tên sách 1

Tên sách 2

Tên sách 3

Tác giả 3

Tên sách 1

Tên sách 2

v.v...

Trước khi điền vào trang sắp xếp ta phải hình dung trước yêu cầu trình bày đầu ra của dữ liệu.

Mỗi mục của trang sẽ được giải thích và minh hoạ bởi ví dụ sau:

Khai báo in - SÁCH

Trang chung Trình bày Đặt lề Chế bản Sắp xếp

Số tiêu đề: File từ vô nghĩa:

Format tiêu đề:

Độ dài khoá 1: Chỉ số xử lý Tiêu đề:

☒ FST:

Độ dài khoá 2: Chỉ số xử lý Tiêu đề:

☒ FST:

Độ dài khoá 3: Chỉ số xử lý Tiêu đề:

☒ FST:

Độ dài khoá 4: Chỉ số xử lý Tiêu đề:

☒ FST:

Đầu ra được sắp xếp của cơ sở dữ liệu DHVH, với hai mức tiêu đề (Tác giả rồi Tên sách), với kết quả in ra là các yếu tố xuất bản, số trang và từ khoá (Format hiện hình là mhl, v3, '-', v4/"Từ khoá:",v6/#)

Số tiêu đề: Đưa vào số lượng tiêu đề sẽ làm nổi bật khi in. Số này không thể lớn hơn số khoá sắp xếp. Nói chung số lượng tiêu đề nên bằng số lượng khoá sắp xếp trừ đi 1 và khoá sắp xếp cuối nên tương ứng với phần tử dữ liệu đầu tiên mà ta in ra từ mỗi biểu ghi.

Ví dụ: Trong mục này ta điền vào số 2 (in 2 khoá sắp xếp đầu, ở đây là tác giả và tên sách).

File từ vô nghĩa: Đưa vào tên tệp các từ vô nghĩa. WinISIS sẽ bỏ qua các từ này khi sắp xếp.

Format tiêu đề: Là format riêng dùng để in các tiêu đề. Nếu để trường này trống, một format mặc định sẽ làm việc.

Ví dụ: mhu, v1/tab(900),v2/

mhu	Các tiêu đề là chữ in hoa
v1	Dòng đầu tiên của tiêu đề
/	Xuống một dòng
tab(900), v2	Dòng thứ hai của tiêu đề in cách lề trái khoảng 1,5 cm.

Độ dài khoá: là số các ký tự trên đó sẽ thực hiện việc sắp xếp, phụ thuộc vào độ dài của trường mà từ đó khoá được lấy ra. Cần chọn kích thước đủ dài để tạo dãy sắp xếp đúng.

Chỉ số xử lý tiêu đề:

0 : để chỉ rằng chỉ phần tử thứ nhất của trường được sắp xếp.

1 : để chỉ rằng mỗi phần tử của trường đều được sắp xếp.

FST của khoá sắp xếp: Ta đưa vào đây một bảng chọn trường (FST) chỉ dùng để sắp xếp. Nói chung ta không sử dụng lại FST được dùng để xây dựng tệp đảo. Một FST đơn giản sẽ gồm 3 phần sau, mỗi phần cách nhau một khoảng trống:

- nhãn trường (1 đối với khoá thứ nhất, 2 đối với khoá thứ hai)
- kỹ thuật đảo
- format tách dữ liệu

Ví dụ 1: Trong ví dụ đang xét, FST của hai khoá sắp xếp là:

1 0 v1+|;

1 0 v2

Ví dụ 2: 1 2 v7 + 1 3 v8

Cách viết này cho phép sắp xếp đồng thời hai trường mà thuật ngữ được tách ra với kỹ thuật khác nhau.

Cuối cùng để in thêm các yếu tố xuất bản, số trang, từ khoá bên dưới hai tiêu đề, ta phải soạn thảo một format trình bày vào mục **Format in** ở **Trang tổng quát** của **Hộp thoại in** với dòng lệnh sau:

mhl, v3,'-', v4/"Từ khoá:",v6/#

Kết quả Mục lục sắp xếp theo **tác giả** và **tên sách** cùng các thông tin về xuất bản, số trang, từ khoá của CSDL DHVH sẽ được in ra như sau:

Bùi Bình Thi

Mùa mưa đến sớm.

Hà Nội, Văn học, 1973.- 181 tr.

Từ khoá: Truyện.

Đường về cánh đồng Chum.

Hà Nội, Thanh niên, 1975.- 374 tr.

Từ khoá: Truyện.

Bùi Hiền

Một cuộc đời.

Hà Nội, Phụ nữ, 1977.- 155 tr.

Từ khoá: Truyện ký.

Trong gió cát.

Hà Nội, Văn học, 1965.- 158 tr.

Từ khoá: Truyện ngắn.

Chu Văn
Bão biển.
Hà Nội, Văn học, 1978,- 516 tr.
Từ khoá: Tiểu thuyết.
v.v...

5. TRAO ĐỔI DỮ LIỆU

5.1. Nguyên tắc trao đổi dữ liệu

Việc trao đổi dữ liệu đem lại nhiều lợi ích. Nhờ đó ta có thể bổ sung vốn tư liệu của mình và xuất dữ liệu cho người khác sử dụng. Để làm được điều đó cần phải sử dụng phần mềm và format tương thích. Đối với các CSDL tài liệu, có một format trao đổi chuẩn là format ISO-2709.

Mặc dù có ít phần mềm quản trị CSDL chấp nhận format này, việc trao đổi dữ liệu vẫn có thể thực hiện dễ dàng giữa các dữ liệu được quản lý bởi các phần mềm CDS/ISIS và một số phần mềm khác chạy trên máy lớn và máy mini. Còn đối với nhiều phần mềm khác phải tạo giao diện riêng (ví dụ đã có các giao diện với TEXT0 và DBASE).

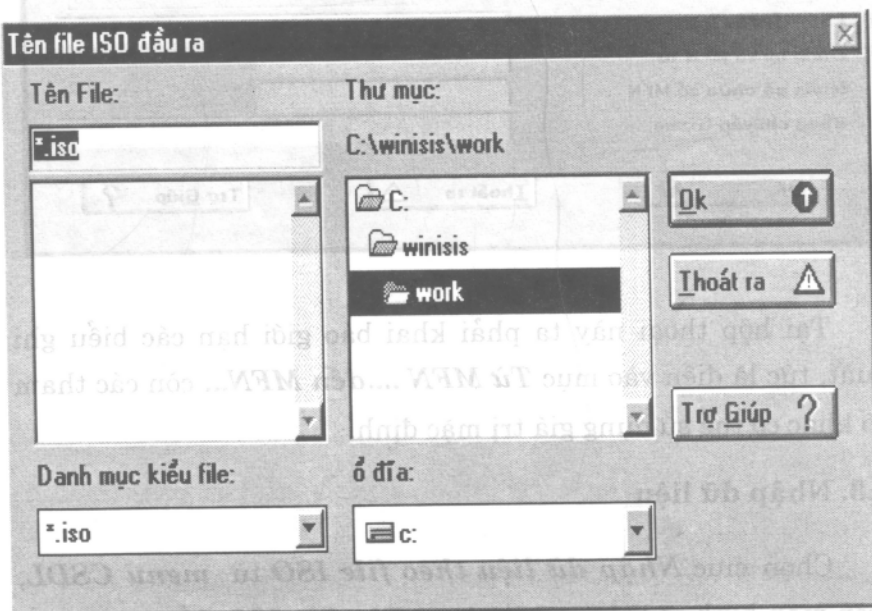
Có một phương tiện để trao đổi dữ liệu giữa các đối tượng cùng sử dụng chương trình CSD-ISIS là chép các tệp tạo thành CSDL. Tuy nhiên cách làm này không tiết kiệm bộ nhớ vì phải chép ít nhất 12 tệp.

5.2. Xuất dữ liệu

Thao tác này đòi hỏi tạo ra một đầu ra của CSDL với format ISO.2709. Khi đó CSDL này có thể sát nhập vào một

CSDL khác hoặc trên cùng một hệ đĩa cứng, hoặc trên một máy tính khác nhờ một đĩa mềm.

Chọn mục **Xuất dữ liệu theo file ISO** từ menu CSDL, trên màn hình sẽ hiện ra hộp thoại **Tên file ISO đầu ra** như sau:



Trong mục **Tên file ISO đầu ra** ta gõ vào tên của file xuất ra của CSDL đang mở, ví dụ: **Xuat1.iso**, rồi chọn **OK**. Khi đó hiện ra hộp thoại **Xuất ra theo ISO 2709**:

Xuất ra theo ISO-2709

Tên file ISO đầu ra

Dấu phân cách trường
Dấu phân cách BG

Dãy

Từ MFN	1
đến MFN	9999999

Xuất:

☒ Dãy số MFNs
☐ Kết quả tìm tin
☐ Biểu đánh dấu

FST đổi cấu trúc
Đánh lại số MFN từ
Nhãn có chứa số MFN
Bảng chuyển Gizmo

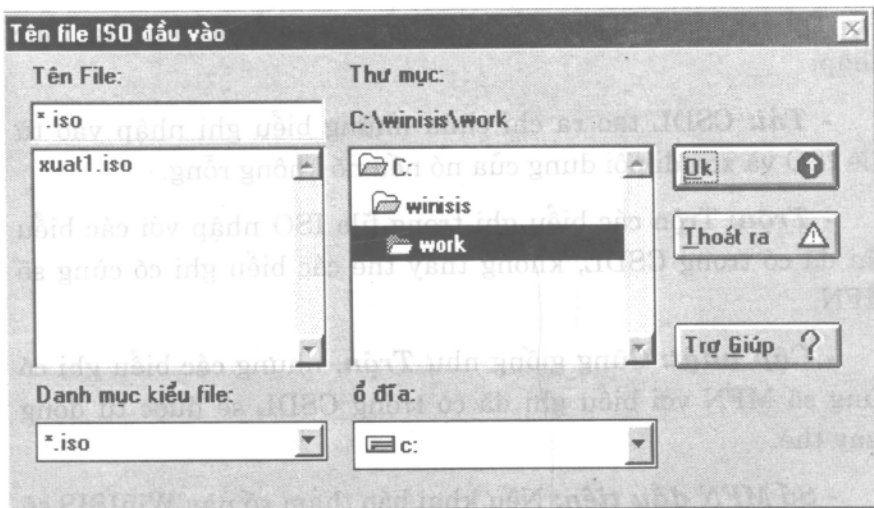
OK

Tại hộp thoại này ta phải khai báo giới hạn các biểu ghi xuất, tức là điền vào mục **Từ MFNđến MFN...** còn các tham số khác có thể sử dụng giá trị mặc định.

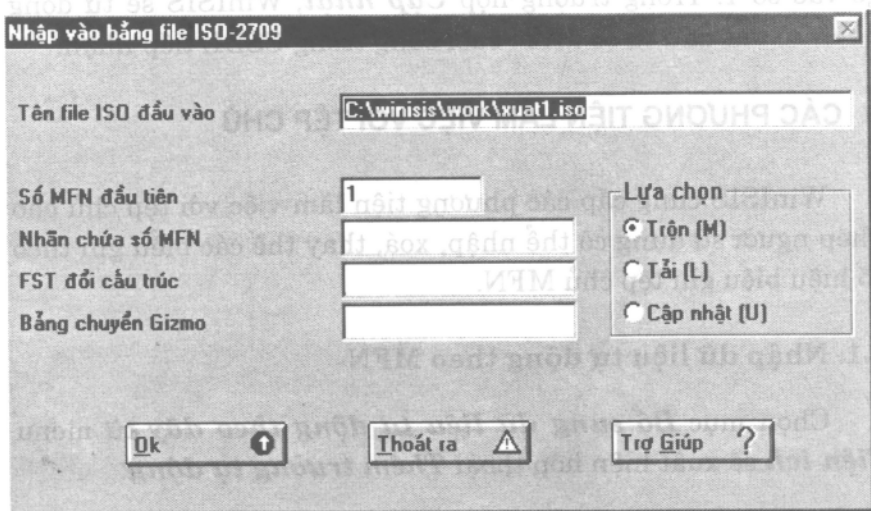
5.3. Nhập dữ liệu

Chọn mục **Nhập dữ liệu theo file ISO** từ **menu CSDL**, trên màn hình sẽ hiện ra hộp thoại **Tên file ISO đầu vào**.

Ở hộp này ta khai báo tên file ISO cần nhập. Các file ISO thường đặt trong thư mục **\WinISIS\Work**. Ta cũng có thể chọn file ISO từ một thư mục khác. Sau khi chọn xong, kết thúc bằng việc nhấn chuột vào mục **OK**.



Sau đó xuất hiện hộp thoại:



Tại hộp này ta chỉ cần chọn một trong các phương thức nhập:

- **Tải:** CSDL tạo ra chỉ chứa những biểu ghi nhập vào từ file ISO và xoá đi nội dung của nó nếu nó không rỗng.

- **Trộn:** Trộn các biểu ghi trong file ISO nhập với các biểu ghi đã có trong CSDL, không thay thế các biểu ghi có cùng số MFN.

- **Cập nhật:** Cũng giống như **Trộn**, nhưng các biểu ghi có cùng số MFN với biểu ghi đã có trong CSDL sẽ được tự động thay thế.

- **Số MFN đầu tiên:** Nếu khai báo tham số này WinISIS sẽ đánh số lại các biểu ghi nhập vào, bắt đầu từ số MFN mà bạn đã xác định. Thông thường nếu chọn phương thức **Tải**, bạn phải gõ vào số 1. Trong trường hợp **Cập nhật**, WinISIS sẽ tự động đánh số bắt đầu từ số MFN cuối cùng trong CSDL tiếp nhận.

6. CÁC PHƯƠNG TIỆN LÀM VIỆC VỚI TẬP CHỦ

WinISIS cung cấp các phương tiện làm việc với tập chủ cho phép người sử dụng có thể nhập, xoá, thay thế các biểu ghi theo số hiệu biểu ghi tập chủ MFN.

6.1. Nhập dữ liệu tự động theo MFN

Chọn mục **Bổ sung dữ liệu tự động theo dãy** từ menu **Tiện ích** sẽ xuất hiện hộp thoại **Thêm trường tự động**:

Thêm trường tự động

Dãy MFN (từ/đến)

Nhãn trường

☐ Add if not present

Nội dung trường

Thoát ra

Ok

Dãy MFN (từ/đến): Đưa vào đây giới hạn trên và dưới của số MFN của các biểu ghi cần bổ sung thêm dữ liệu. Ví dụ: 20/30.

Nhãn trường: Đưa vào mục này nhãn trường cần bổ sung dữ liệu. Ví dụ: 6.

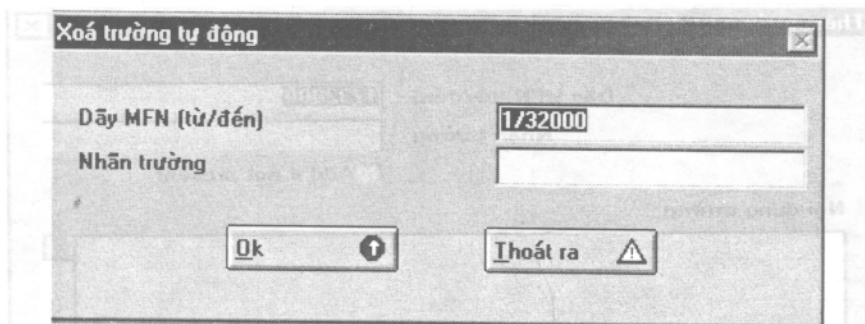
Nội dung trường: Đưa vào nội dung cần phải bổ sung cho trường.

Nút OK: Nhấn nút này để bắt đầu công việc bổ sung.

Nút thoát ra: Huỷ bỏ lệnh bổ sung dữ liệu.

6.2. Xóa dữ liệu tự động theo dãy MFN

Chọn mục **Xóa dữ liệu tự động theo dãy** từ menu **Tiện ích** sẽ xuất hiện hộp thoại **Xóa trường tự động**:



Dãy MFN (từ/đến): Đưa vào đây giới hạn trên và dưới của số MFN của các biểu ghi cần xoá dữ liệu. Ví dụ: 20/30.

Nhãn trường: Đưa vào mục này nhãn trường cần xoá dữ liệu. Ví dụ: 6.

Nút OK: Nhấn nút này để bắt đầu công việc xoá.

Nút thoát ra: Huỷ bỏ lệnh xoá dữ liệu.

6.3. Thay thế dữ liệu tự động theo dãy MFN

Chọn mục **Thay thế dữ liệu tự động** từ menu **Tiện ích** sẽ xuất hiện hộp thoại **Thay thế đoạn văn bản**.

Trong hộp thoại này có các mục sau:

Giới hạn MFN: Đưa vào đây giới hạn trên và dưới của số MFN của các biểu ghi cần xoá dữ liệu. Ví dụ: 20/30.

Đoạn văn cần tìm: Nhập vào đây dãy ký tự cần phải thay thế

Thay thế đoạn văn bản

Dãy

☒ Giới hạn MFN:

☐ Câu hỏi số: (☐ Biểu đánh dấu)

Đoạn văn cần tìm:

Văn bản mới:

Lựa chọn

☐ Phân biệt chữ hoa/thường

☐ Cả từ hoàn chỉnh

Scope

Nhãn:

Trường con:

Thoát ra **Trợ Giúp** **Ok**

Văn bản mới: Đưa vào đây đoạn văn sẽ thay thế dãy ký tự trên. Nếu bỏ trống hộp này, dòng ký tự trong hộp **Đoạn văn cần tìm** sẽ bị xóa.

Lựa chọn: Chọn một trong các phương án trong hộp thoại, hoặc phân biệt chữ hoa chữ thường hoặc cả từ hoàn chỉnh.

Scope (Phạm vi): Nếu bạn muốn thay thế văn bản chỉ trong một số trường thì điền vào hộp **Trường** các nhãn trường thích hợp, chúng cách nhau bằng một dấu phẩy (ví dụ: **6,7**). Cũng tương tự như vậy với hộp **trường con** (ví dụ: **abc**).

Nút OK: Nhấn nút này để bắt đầu công việc thay thế.

Nút thoát ra: Huỷ bỏ lệnh thay thế dữ liệu.

Chương 9

THIẾT KẾ MỘT HỆ TIN HỌC ỨNG DỤNG TRONG CÔNG TÁC THƯ VIỆN

1. MỞ ĐẦU

Hệ tin học ứng dụng trong công tác thư viện là một thực tế phức tạp. Vì vậy việc xây dựng một hệ tin học ứng dụng trong công tác thư viện phải tuân theo một quy trình chặt chẽ và là sự hợp tác của nhiều cán bộ chuyên môn, mà thành phần tối thiểu cần thiết có là: cán bộ phân tích hệ thống, cán bộ chuyên môn trong lĩnh vực thư viện và cán bộ tin học.

Nhiệm vụ tổng quát của việc thiết kế một hệ thống tin học ứng dụng trong công tác thư viện có thể được xác định như sau:

- Tập hợp các nhu cầu thông tin và các nhu cầu ứng dụng tin học trong hoạt động thư viện.
- Tích nhập các nhu cầu đó trong một hệ tin học ứng dụng.

Mục tiêu phải đạt tới của quá trình thiết kế một hệ thống Thư viện tin học hoá là:

- Cải tiến quản lý vốn tư liệu của Thư viện. Xây dựng các CSDL thư mục thoả mãn nhu cầu lưu trữ và tìm kiếm thông tin.

- Cải tiến quy trình làm việc của Thư viện. Sử dụng máy tính điện tử trong các khâu kỹ thuật và quản lý Thư viện.
- Hợp lý hoá dây chuyền, cải tiến tổ chức thư viện trong điều kiện tin học hoá.

Để hoàn thành các nhiệm vụ và đạt được các mục tiêu trên đây, quá trình thiết kế và xây dựng một hệ tin học ứng dụng trong thư viện phải qua các giai đoạn sau đây:

- Phân tích chức năng hệ thống.
- Thiết kế mô hình hệ thống.
- Lựa chọn phần mềm và phương tiện kỹ thuật.
- Tính toán các chi phí và xác định các nguồn kinh phí.

2. PHÂN TÍCH CHỨC NĂNG HỆ THỐNG

Nhiệm vụ của thư viện là thu thập bảo quản tài liệu và thoả mãn nhu cầu về tài liệu của người đọc.

Nhu cầu của người đọc được kích thích bởi các thông tin chuyển tới họ bằng nhiều nguồn khác nhau, có thể từ:

- Các nhà xuất bản (các mục lục, các thông báo về tài liệu mới)
- Các trích dẫn thư mục (trong các ấn phẩm thư mục, các CSDL thư mục)
- Các dòng thông tin không chính thức (hội thảo, hội nghị, triển lãm...)

Các nhu cầu đó có thể và phải được sự hỗ trợ của các cơ quan thông tin tư liệu, trong đó có thư viện. Có thể nói thư viện

là cầu nối giữa nguồn cung cấp tài liệu (nhà xuất bản, hiệu sách, tổ chức chuyên môn) với người sử dụng cuối cùng.

Theo cách hiểu truyền thống quá trình xử lý Thư viện bao gồm các khâu chính sau:

- Bổ sung: xây dựng vốn tài liệu.
- Xử lý tài liệu (sách, tạp chí): biên mục, lưu trữ và bảo quản.
- Lưu thông: phục vụ bạn đọc tại chỗ, mượn về, tra cứu.

Ta hãy phân tích các chức năng trên đây dưới giác độ của tin học.

Bổ sung

Chức năng:

- Kiểm soát tài liệu mới nhập về theo các nguồn khác nhau: đặt mua, trao đổi, tặng biếu.

- Thống kê báo cáo kịp thời các số liệu và tài liệu đặt và kinh phí sử dụng.

Các thông tin cần xử lý và sản phẩm đầu ra:

- Quản lý đơn đặt theo từng cơ sở.
- Danh mục các tài liệu đang đặt và mới về.
- Danh mục các tài liệu nhận được qua các nguồn, các tổ chức.
- Các báo cáo về số lượng và kinh phí mua tài liệu.

Xử lý tài liệu

Chức năng:

- Xử lý hình thức, nội dung tài liệu: biên mục, định chủ đề, tóm tắt, chú giải...

- Tổ chức bộ máy tra cứu về sách và tạp chí, phục vụ cho việc lưu trữ và tìm kiếm thông tin.

Các thông tin cần xử lý và sản phẩm đầu ra:

- Các yếu tố mô tả thư mục, các từ khoá chủ đề, các ký hiệu phân loại, ngôn ngữ.
- Các bộ phiếu in để bổ sung và cập nhật các tử mục lục truyền thống (phân loại, chữ cái, công vụ, vị trí, tra cứu).
- Các ấn phẩm thư mục.
- Thông báo tài liệu mới.
- Các câu trả lời theo yêu cầu tìm tin hỏi đáp.

Phục vụ bạn đọc

Chức năng:

- Cấp phát thẻ đọc, sổ mượn về.
- Quản lý hồ sơ bạn đọc.
- Theo dõi tình hình phục vụ bạn đọc: quá trình phục vụ và hiệu quả.
- Thống kê báo cáo về tình hình phục vụ bạn đọc.

Các thông tin cần xử lý và sản phẩm đầu ra:

- Tên, địa chỉ, chức vụ, trình độ văn hoá, cơ quan công tác và các dữ liệu khác liên quan đến bạn đọc.
- Các báo cáo định kỳ về: thành phần bạn đọc, diện đọc, số lượt đọc, số lượt mượn...

Từ phân tích chức năng hệ thống ở trên ta thấy rằng các đối tượng quản lý chính trong thư viện là:

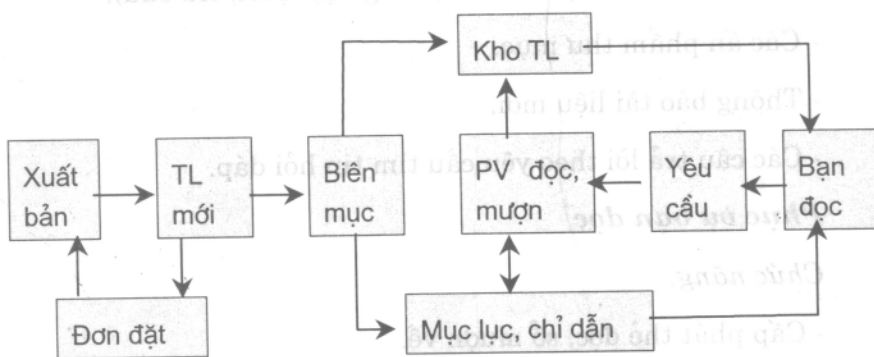
a- Sách

b- Tạp chí và các ấn phẩm định kỳ

c- Bạn đọc

d- Đơn đặt

Sơ đồ lưu thông tài liệu của quá trình phục vụ thư viện có thể mô tả bằng sơ đồ sau:



Hình 22. Quá trình lưu thông tài liệu trong thư viện

Yếu điểm của hệ thống thư viện chưa tin học hoá là toàn bộ các khâu trong dây chuyền xử lý thư viện còn thực hiện thủ công. Khâu bổ sung khó theo dõi quá trình thực hiện đơn đặt. Khâu xử lý sách, tạp chí vẫn bị chi phối bởi quy trình làm phiếu thủ công, tài liệu chậm được đưa vào sử dụng. Yếu điểm cơ bản của các bộ phiếu thủ công là không cho phép người sử dụng tra cứu đầy đủ về những thông tin có trong tài liệu, thời gian tra cứu lại lâu. Khâu phục vụ bạn đọc do tổ chức thủ công nên rất khó theo dõi việc mượn quá hạn, thủ thư không biết được sách có còn trong kho hay không. Khả năng phân tích thống kê rất hạn chế.

Tất cả những hạn chế nêu trên nói lên rằng việc áp dụng máy tính điện tử vào các khâu xử lý thư viện là cần thiết và đòi hỏi có những giải pháp và bước đi hợp lý cho quá trình này.

3. THIẾT KẾ MỘT HỆ TIN HỌC ỨNG DỤNG TRONG THƯ VIỆN

Từ việc phân tích chức năng hệ thống và lược đồ lưu thông tài liệu ở trên ta thấy rằng hệ thống tin học ứng dụng trong thư viện ở giai đoạn đầu bao gồm hai mảng sau đây:

a- Các hệ chương trình quản trị các CSDL. Đó là các CSDL thư mục về sách và CSDL về tạp chí. Các CSDL này thực hiện chức năng lưu trữ và tìm kiếm thông tin tự động hóa. Sản phẩm đầu ra sẽ là:

- Các loại thư mục chuyên đề, các bản tra cứu chỉ dẫn.
- Các bộ phiếu in trên máy để bổ sung và cập nhật các tử mục lục truyền thống.
- Các câu trả lời cho các yêu cầu tìm tin.

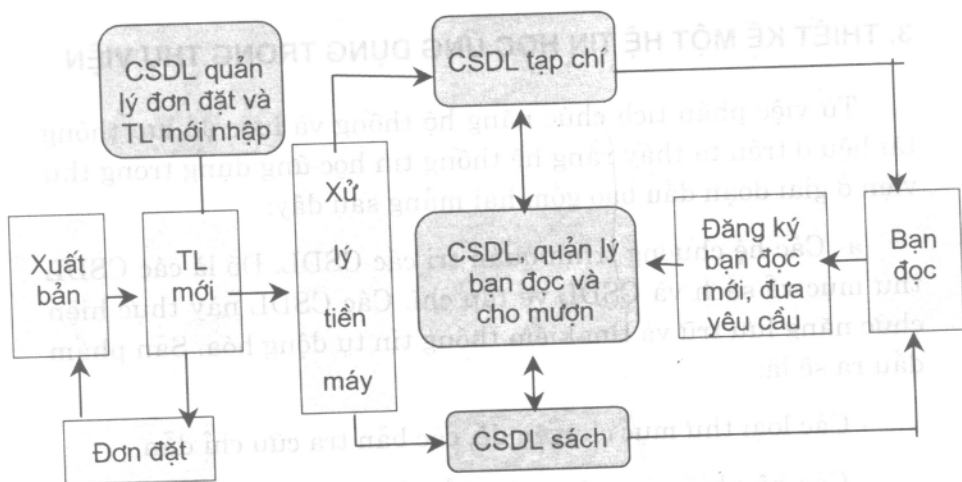
b- Các chương trình ứng dụng thực hiện các chức năng quản lý đặc thù. Cụ thể là:

- Một hệ chương trình quản lý đơn đặt và kiểm soát tài liệu nhập.
- Một hệ chương trình quản lý bạn đọc và quá trình phục vụ bạn đọc.

Sơ đồ tổng quát của hệ thống có thể mô tả ở hình 22.

Một trong những yêu cầu đặt ra đối với một hệ thống Thư viện tin học hoá là nó phải hoạt động trong môi trường bảo đảm cho nhiều người sử dụng cùng một lúc (multiuser). Điều này không những cho phép người tại cùng thời điểm có thể truy

nhập tới cùng một CSDL mà còn sử dụng tối ưu các thiết bị ngoại vi như đĩa cứng, máy in...



Hình 23. Hệ thống các CSDL cơ bản trong một thư viện

Vì vậy, bước tiếp theo là phải xây dựng một mạng cục bộ các máy tính để điều hành và khai thác toàn bộ hoạt động của hệ thống.

Lựa chọn phần mềm

Các CSDL về sách và tạp chí là những CSDL văn bản, quản lý các dữ liệu mang tính chất tư liệu như các dữ liệu thư mục về sách và tạp chí. Vì vậy hệ quản trị CSDL văn bản CDS/ISIS là phần mềm thích hợp để xây dựng các CSDL này. Đây là phần mềm đang được sử dụng rộng rãi trong các trung tâm Thông tin tư liệu và Thư viện của ta hiện nay.

Tuy nhiên những phương tiện chuẩn của CDS/ISIS không thể đáp ứng được các yêu cầu mang tính đặc trưng khác của hệ quản lý đơn đặt và tài liệu nhập, cũng như hệ quản lý bạn đọc

như các yêu cầu về tính toán, thống kê, phân tích và xử lý số liệu... Vì vậy giải pháp phần mềm ở đây có thể là:

a- Sử dụng CDS/ISIS nhưng phải thiết kế các chương trình bổ sung bằng ngôn ngữ lập trình CSD/ISIS Pascal để thoả mãn các yêu cầu quản lý khác. Tuy nhiên điều này đòi hỏi cần phải có cán bộ tin học có khả năng lập trình tốt.

b- Sử dụng hệ phần mềm quản trị CSDL Foxpro. Đây là phần mềm quản trị CSDL đa năng, hiện đại và đang được ưa chuộng ở Việt Nam. Với hoạt động thông tin - thư viện nó tỏ ra phù hợp vì có những ưu điểm nổi bật sau đây:

- CSDL trong Foxpro được xây dựng theo mô hình quan hệ nên khai thác được nhiều ưu điểm của mô hình này.
- Có khả năng giao tiếp với các phần mềm phổ biến khác.
- Giao diện với người dùng dễ hiểu, thẩm mỹ.
- Khả năng lưu trữ thông tin lớn.
- Khả năng tính toán mạnh. Đặc biệt có các hàm chuyên xử lý thống kê.
- Có ngôn ngữ lập trình mạnh, đa năng và dễ sử dụng.

Vì vậy chúng ta thiên về lựa chọn giải pháp thứ hai.

Khi triển khai xây dựng mạng cục bộ, ta có thể chọn hệ điều hành mạng Novell. Mạng Novell Netware là mạng có nhiều ưu thế về mặt phổ biến, khả năng phát triển, dễ cài đặt, dễ sử dụng cho phép người sử dụng lựa chọn cấu hình hệ thống sao cho phù hợp với nhu cầu của mình.

Việc thiết kế cấu trúc của các CSDL nói trên phụ thuộc vào yêu cầu của từng đơn vị thư viện. Ở đây chúng ta chỉ phác thảo

những chức năng chủ yếu định hướng cho việc thiết kế cụ thể sau này.

Mô tả các chức năng chủ yếu của các CSDL

1. CSDL sách

CSDL sách có chức năng chủ yếu sau:

a- *Xử lý tiền máy toàn bộ sách nhập vào từ bộ phận bổ sung.* Công đoạn này bao gồm toàn bộ khâu xử lý hình thức và xử lý nội dung tài liệu: biên mục, phân loại, đánh chỉ số, làm tóm tắt. Do đặc thù của hệ xử lý bằng máy tính nên việc mô tả dữ liệu cần có những thay đổi nhất định và việc tiêu chuẩn hoá mô tả về nội dung và hình thức đòi hỏi ở mức độ cao hơn nhiều so với xử lý truyền thống. Các mô tả được trình bày dưới dạng một format mô tả, chúng gọi là phiếu nhập dữ liệu (worksheet).

b- *Nhập dữ liệu vào máy, sửa đổi dữ liệu* qua worksheet nhập dữ liệu.

c- *Tìm và tra cứu thông tin.* Sử dụng khả năng tìm tin mềm dẻo và linh hoạt của CDS/ISIS, chức năng tìm tin có thể mở rộng cho phép tìm theo nhiều dấu hiệu khác nhau như: nhan đề tài liệu, tác giả, ký hiệu phân loại, ký hiệu kho, chỉ số ISBN, các yếu tố xuất bản, ngôn ngữ xuất bản, từ khoá...

d- *In các bộ phiếu mục lục.* Khả năng tạo ra các sản phẩm đầu ra như mong muốn của CDS/ISIS cho phép tạo ra các bộ phiếu chỉ dẫn thư mục khác nhau như bộ phiếu chính, bộ phiếu chỉ chỗ, bộ phiếu tên sách ngoài bìa, bộ phiếu chỉ bản dịch...

e- *In thông báo sách mới.* Hệ thống có khả năng tạo ra một tệp văn bản cho phép trình bày một ấn phẩm thông báo sách mới theo một format cho trước.

2. CSDL tạp chí

CSDL tạp chí có chức năng chủ yếu sau:

a- *Xử lý tiền máy toàn bộ tạp chí nhập vào từ bộ phận bổ sung.* Chức năng xử lý tiền máy trong CSDL tạp chí cũng gần giống như công tác biên mục của xử lý sách, đó là mô tả ấn phẩm để tổ chức mục lục thư viện. Kết quả của khâu xử lý này đối với một đơn vị ấn phẩm là tờ phiếu nhập tin với những mô tả đầy đủ các trường của nó.

b- *Nhập dữ liệu vào máy.* Do đặc điểm của tạp chí nên việc cập nhật tạp chí có hai yếu tố đặc thù cần chú ý là: tên tạp chí và số tạp chí. Với tên tạp chí có: tên chính, tên song song, tên khác, tên cũ, tên mới.

c- *Tìm và tra cứu thông tin.* Có thể tìm theo: tên tạp chí, ký hiệu kho, chỉ số ISSN, ngôn ngữ xuất bản, nước xuất bản, từ khoá.

d- *In các bộ phiếu mục lục.* Hệ thống cho phép tạo ra các bộ phiếu cơ bản cho tạp chí như bộ phiếu theo vần chữ cái và bộ phiếu phân loại.

e- *In thông báo tạp chí mới.* Hệ thống có thể tạo ra một tệp văn bản và cho phép trình bày ấn phẩm thông báo tạp chí mới theo một format định trước.

3. CSDL quản lý đơn đặt và tài liệu mới bổ sung

Do nhu cầu tính toán và thống kê, CSDL này được thiết kế trên cơ sở hệ quản trị CSDL Foxpro với một chương trình ứng dụng nhằm thực hiện các chức năng sau:

a- *Quản lý đơn đặt tài liệu.* Theo dõi các thông tin về đơn đặt tài liệu như: mã số đơn đặt, ngày đặt, số đầu tên tài liệu

đặt, số đầu tên tài liệu nhận, cơ quan cung cấp tài liệu, phương thức tiếp nhận (mua, tặng, trao đổi), số tài liệu chưa nhận.

b- *Quản lý tài liệu mới bổ sung.* Cập nhật tài liệu mới bổ sung, theo dõi thông tin về tài liệu mới bổ sung như: tên tài liệu, tác giả, ngôn ngữ, ISSN, ISBN, cơ quan xuất bản/biên tập, nhà xuất bản, nơi xuất bản, đơn giá, số lượng, tổng giá tiền, ngày nhận tài liệu.

Từ phân tích các chức năng trên ta thấy rằng CSDL quản lý đơn đặt và tài liệu nhập cần được xây dựng từ hai quan hệ cơ bản là: quan hệ ĐƠN ĐẶT và quan hệ TÀI LIỆU NHẬP với các thuộc tính là các yếu tố dữ liệu đã nêu ở trên. Hai quan hệ này tạo thành hai tệp dữ liệu DBF của CSDL. Khoá kết nối hai tệp dữ liệu này là "mã số đơn đặt". Một chương trình viết bằng ngôn ngữ lập trình Foxpro, với nhiều ưu điểm vốn có của nó sẽ cho phép thực hiện các chức năng quản lý mà bộ phận bổ sung đặt ra.

4. CSDL quản lý bạn đọc và cho mượn

CSDL quản lý bạn đọc và cho mượn thực hiện các chức năng sau đây:

a- *Đăng ký bạn đọc mới.* CSDL quản lý việc đăng ký bạn đọc mới và quản lý bạn đọc với các yếu tố dữ liệu sau: họ và tên, giới tính, dân tộc, năm sinh, trình độ văn hoá, chức vụ công tác, học hàm, học vị, chuyên ngành đào tạo, trình độ ngoại ngữ, đơn vị công tác, địa chỉ cơ quan, nơi ở, mã số thẻ đọc, ngày cấp thẻ, mã số thẻ mượn.

b- *Theo dõi thông tin về bạn đọc.* Thống kê số lượng bạn đọc theo chuyên môn, trình độ, ngành nghề...

c- *Quản lý việc mượn sách của bạn đọc tại chỗ*. Quản lý và theo dõi tên sách và số lượng sách cho bạn đọc mượn ở phòng đọc trong ngày.

d- *Quản lý việc mượn sách của bạn đọc mượn về*. Quản lý và theo dõi tình hình mượn sách của bạn đọc mượn về với các dữ liệu sau: mã số thẻ mượn, tên và số lượng sách mượn, ngày mượn, ngày trả.

Từ những điều phân tích ở trên ta thấy rằng CSDL quản lý bạn đọc và cho mượn về cơ bản được xây dựng từ 3 quan hệ. Đó là quan hệ **BẠN ĐỌC**, quan hệ **MƯỢN TẠI CHỖ**, quan hệ **MƯỢN VỀ** với các thuộc tính là các yếu tố dữ liệu được mô tả như trên. Tương ứng với 3 quan hệ nói trên là 3 tệp dữ liệu DBF của CSDL. Khoá kết nối các tệp dữ liệu này là "mã số thẻ đọc" và "mã số thẻ mượn".

Một chương trình viết bằng ngôn ngữ lập trình Foxpro sẽ cho phép thực hiện các chức năng quản lý của CSDL quản lý bạn đọc và cho mượn theo yêu cầu đặt ra.

Trên đây là mô hình triển khai ứng dụng tin học trong công tác thư viện ở giai đoạn đầu, nó phù hợp với điều kiện và khả năng của các thư viện của ta hiện nay, nhất là đối với thư viện cấp tỉnh và tương đương. Để triển khai các phần mềm tích hợp quản trị thư viện, cần có sự đầu tư không nhỏ và sự phối hợp đồng bộ ở tầm quốc gia.

PHỤ CHƯƠNG

A. XÂY DỰNG VÀ KHAI THÁC CSDL BẰNG CHƯƠNG TRÌNH CDS/ISIS CHẠY TRÊN DOS

1. XÂY DỰNG CẤU TRÚC CSDL

CDS/ISIS 3.08 do UNESCO phát hành năm 1997 là một trong những version gần đây nhất của chương trình CDS/ISIS chạy trên môi trường DOS. Hiện nay hầu hết các thư viện ở nước ta vẫn còn đang sử dụng version này để xây dựng và quản lý các cơ sở dữ liệu tư liệu. Vì vậy chúng tôi giới thiệu ở đây cách sử dụng chương trình này để bạn đọc tham khảo.

Cũng giống như CDS/ISIS for Windows, các tệp xác định cấu trúc của CSDL trong CDS/ISIS 3.08 cũng gồm 4 tệp: bảng xác định trường, bảng chọn trường, Worksheet nhập dữ liệu và format trình bày.

Việc xây dựng cấu trúc của CSDL cũng gồm bốn bước liên tục để xác lập các tệp trên. Tuy nhiên, do không có các phương tiện trợ giúp nên việc xây dựng Worksheet nhập dữ liệu phải tiến hành qua nhiều bước khá phức tạp. Ngoài ra cần phải nắm vững các lệnh của ngôn ngữ tạo format của chương trình SDS/ISIS chạy trên DOS để lập format trình bày.

Hầu hết các lệnh của ngôn ngữ tạo format của CDS/ISIS 3.08 cũng trùng với các lệnh của ngôn ngữ tạo format của WinISIS. Tuy nhiên CDS/ISIS 3.08 không có các lệnh liên kết siêu văn bản LINK, không có một số hàm và có sự khác biệt đáng kể trong các lệnh định dạng văn bản.

CDS/ISIS 3.08 sử dụng các lệnh sau để định dạng văn bản:

Xn: Chèn vào n khoảng trống trên dòng.

Cn: Chuyển đến vị trí thứ n của dòng.

Vtt (a,b): Dòng đầu của trường thụt vào a khoảng trống; các dòng còn lại thụt vào b khoảng trống. Nếu có một tiêu đề trước trường thì dòng đầu của trường bao giờ cũng bắt đầu ngay sau tiêu đề (trong trường hợp này giá trị a không có tác dụng).

Các lệnh này có tác dụng tương tự như các lệnh **Tab(giá trị)** và **M(indent, findent)** của WinISIS.

Trước khi tạo lập một CSDL bằng chương trình CDS/ISIS 3.08 ta cần chuẩn bị trước trên giấy một danh mục các trường, lựa chọn các trường để đánh chỉ số, dự kiến hình thức trình bày của một biểu mẫu nhập tin (Worksheet), dự kiến một format hiện hình để trình bày các biểu ghi.

Quy trình tạo lập cấu trúc của CSDL bao gồm bốn bước liên tục sau đây:

- Lập bảng xác định trường.
- Tạo lập phiếu nhập tin.
- Tạo lập format trình bày.
- Xây dựng bảng chọn trường.

Sau đây ta sẽ trình bày chi tiết quá trình xây dựng cấu trúc một CSDL bằng chương trình CDS/ISIS 3.08.

Sau khi gọi chương trình CDS/ISIS, menu chính sẽ xuất hiện trên màn hình.

```
*****  
Micro CDS-ISIS - Version 3.08  
*****  
C - Chọn cơ sở dữ liệu  
L - Chọn ngôn ngữ làm việc  
  
E - ISISENT - Cập nhật dữ liệu  
S - ISISRET - Tìm kiếm dữ liệu  
P - ISISPRT - Sắp xếp và in  
I - ISISINV - Bảo trì tệp đảo  
D - ISISDEF - Xác định CSDL  
M - ISISXCH - Trao đổi dữ liệu và bảo trì tệp chủ  
U - ISISFUL - Các phương tiện hệ thống  
A - ISISPAS - Lập trình phát triển Pascal  
  
X - Trở về DOS  
?  
*****  
Micro CDS/ISIS - Ver.3.08. (C) Unesco 1997
```

Để tạo lập cấu trúc của một CSDL, từ menu chính ta chọn chức năng D. Khi đó menu con **Xác định CSDL** sẽ hiện ra :

Phương tiện ISISDEF	Xác định CSDL	Menu FXDEF
L - Chọn ngôn ngữ làm việc		
C - Xác định cấu trúc CSDL mới		
U - Sửa đổi cấu trúc CSDL		
I - Khởi tạo lại file chủ		
X - Trở về menu trước		

Chú ý rằng, một khi chương trình tạo lập cấu trúc của CSDL hoạt động thì 4 tệp xác định CSDL được tạo thành một cách liên tiếp, không gián đoạn. Do đó, nếu có chỗ nào các trường không xác định đúng thì bắt buộc phải bỏ qua sai sót này đến tận cuối cùng. Sau đó, tất nhiên ta có thể sửa chữa các sai sót này. Tuy nhiên, tốt hơn hết là ta thực hiện chính xác ngay từ đầu. Muốn vậy cần phải chuẩn bị kỹ càng trước trên giấy nội dung của các tệp xác định cấu trúc của CSDL.

Từ menu **Xác định CSDL** ta chọn chức năng C để tạo lập một CSDL mới. CDS/ISIS sẽ đưa ra thông báo sau:

Tên CSDL?

Trả lời: gõ vào tên CSDL mới (tối đa 6 ký tự) <CR>

Nếu bạn gõ vào tên một CSDL đã tồn tại thì CDS-ISIS sẽ đưa ra thông báo:

CSDL đang tồn tại. Xoá các tệp (Y/N)?

Trả lời Y nếu muốn xác định lại hoàn toàn CSDL.

Nếu trả lời N thì sẽ trở về menu trên.

Bước 1- Lập bảng xác định trường

Giả sử trả lời cho thông báo **Tên CSDL?** ta gõ vào SACH <CR> để xác định tên CSDL mới là SACH.

Khi đó chương trình soạn thảo dòng để lập FDT của CDS-IACH sẽ hoạt động. Khi chương trình này hoạt động trên màn hình xuất hiện một bảng như sau:

Bảng xác định trường (FDT)					CSDL:SACH
? Nhấn	Tên trường	Dài	Kiểu	Lập	Dấu pc/mẫu
.....					
Vào nhân trường hay <CR> để ra					
EDIT: Thay thế					

Khi chương trình soạn thảo dòng của CDS-ISIS hoạt động, một thông báo trợ giúp xuất hiện cuối màn hình để chỉ rõ nội dung của cột phải gõ vào. Để chuyển từ cột này sang cột khác, ấn phím <CR>.

Ta đưa vào các trường những thông tin cần thiết và chuyển từ trường này sang trường tiếp theo bằng phím <CR>.

Phím <TAB> sẽ chuyển vị trí con trỏ về trường trước đó.

Để sửa đổi hay kết thúc việc lập bảng xác định trường, ấn <CR> khi con trỏ ở cột 1 rỗng (nhấn). Khi đó một menu sẽ xuất hiện ở cuối màn hình:

| <CR> - Trang sau | Sửa đổi | A - Huỷ bỏ | X - Ra và lưu |

Trả lời:

<CR>: Sang trang màn hình sau.

A: Trở về menu chính, không lưu.

X: Lưu các dữ liệu đã nhập và chuyển sang màn hình xác định biểu mẫu nhập dữ liệu (biểu mẫu).

M: Sửa đổi các thông tin đã nhập.

Nếu trả lời M thì chương trình soạn thảo dòng sẽ chuyển con trỏ đến vị trí lệnh của dòng đầu tiên trong bảng và khi đó một menu mới xuất hiện:

A - Chèn (sau) | B - Chèn (trước) | C - Sửa đổi | D - Xoá dòng

P - Trang trước | N - Trang sau | T - Đầu trang | E - Cuối trang
| X - Thoát ra | <CR> - Dòng sau

Trả lời:

A: Chèn một dòng mới vào sau dòng này.

B: Chèn một dòng mới vào trước dòng này.

C: Sửa đổi nội dung của dòng hiện thời (dòng có con trỏ).

D: Xoá dòng.

N: Trang sau.

P: Trang trước.

E: Xem phần cuối của bảng.

T: Xem phần đầu của bảng.

X: Thoát ra.

<CR>: Chuyển con trở xuống dòng sau.

Dưới đây là bảng xác định trường của CSDL SACH:

Bảng xác định trường (FDT)					CSDL SACH
Nhãn	Tên trường	Dài	Kiểu	Lập	Dấu pc/mẫu
1	Tác giả	60	X	R	
2	Nhan đề	120	X		
3	Xuất bản	60	X		abc
4	Mô tả vật lý	30	X		abc
5	Tùng thư	120	X		
6	Ký hiệu phân loại	30	X		
7	Ký hiệu kho	30	X		ab
8	Tóm tắt	300	X		
9	Từ khoá	120	X		

trong đó:

Trường 1 (tác giả) là trường lập (R).

Trường 3 (xuất bản) có 3 trường con: nhà xuất bản (a), nơi xuất bản (b), năm xuất bản (c).

Trường 4 (mô tả vật lý) có 3 trường con: số trang (a), minh hoạ (b), khổ (c).

Trường 7 (ký hiệu kho) có 2 trường con: ký hiệu kho (a), năm nhập tài liệu (b).

Bước 2 - Lập biểu mẫu nhập dữ liệu

Biểu mẫu nhập dữ liệu, còn gọi là worksheet là một hình mẫu trình bày dữ liệu trên màn hình. Để tạo lập biểu mẫu ta phải trả lời các thông báo của chương trình soạn thảo biểu mẫu, đặt ở cuối màn hình.

Với mỗi một trường, chương trình yêu cầu phải cho biết: nhãn trường, tên trường hay tiêu đề của trường, vị trí của các tiêu đề này trên màn hình và nội dung của nó, kiểu hiển thị (bình thường, đảo video, gạch chân,...), độ dài của trường và cuối cùng ta có thể đưa vào thông báo hướng dẫn hoặc giá trị định trước cho trường này.

Khi chương trình soạn thảo biểu mẫu hoạt động, một màn hình gần như rộng xuất hiện:

Vào nhãn trường: SACH/1
--

Dòng cuối cùng của màn hình là tên của biểu mẫu mà ta sẽ tạo cùng với số trang của biểu mẫu. Tên của biểu mẫu trùng với tên của CSDL.

Trong vùng thông báo (dòng 22 - 23) xuất hiện thông báo đầu tiên:

Thông báo 1: *Đưa vào nhân trường*

Trả lời:

?: Nếu bạn không nhớ danh sách các trường đã được xác định ở trên. Khi đó, một trường đã hiện ra với nhân và tên của nó. Bạn có thể gõ:

X: Để chọn trường này

<CR>: Để hiển thị trường sau

<TAB>: Để hiển thị trường trước

T: Nếu bạn muốn hiển thị chỉ một tiêu đề (không là một trường)

Gõ vào trực tiếp nhân trường, nếu bạn biết rõ, rồi ấn <CR>.

Sau khi chọn trường, thông báo sẽ thay đổi và trở thành:

Thông báo 2: *Đưa vào vị trí của tên trường (dòng/cột)*

Vị trí tiếp theo là: 1/1

CDS/ISIS luôn đưa ra một vị trí định trước. Đó là vị trí sau trường mới được đưa vào.

Trả lời:

- Hoặc bằng tọa độ L/C trên màn hình, trong đó L là số dòng và C là số cột, rồi ấn <CR>. Ví dụ: 3/1<CR>.

- Hoặc <CR> để chấp nhận vị trí định trước.

Thông báo 3: *Đưa vào vị trí cho giá trị trường (dòng/cột)*

Vị trí tiếp theo là: x/y

Trả lời (tương tự như trên):

- Hoặc C/L để chỉ ra vị trí hiển thị của giá trị trường.
- Hoặc <CR> để chấp nhận vị trí định trước.

Thông báo 4: **Xác định thuộc tính của trường**
(0 - Bình thường 1 - Đảo 2 - Đậm
3 - Gạch dưới 4 - Nhấp nháy 5 - Ẩn)

Thuộc tính màn hình cho nội dung của trường có thể là:

- 0: bình thường.
- 1: đảo video.
- 2: đậm, độ sáng cao hơn.
- 3: đậm, độ sáng cao hơn.
- 4: nhấp nháy.
- 5: ẩn.

Thông báo 5: **Đưa vào độ dài của trường (xxx)**

Trả lời:

- Hoặc một số để xác định số ký tự của trường.
- Hoặc L theo sau là số dòng (1 dòng bằng 80 ký tự).
- Hoặc S theo sau là số dòng của số.
- Hoặc <CR> để chấp nhận độ dài định trước đặt trong () của FDT.

Ví dụ:

ĐỘ DÀI	Ý NGHĨA
250	250 ký tự
L4	4 dòng, kể từ dòng được xác định trong vị trí L/C của giá trị trường
S3	cửa sổ với 3 dòng

Thông báo 6: *Đưa vào thông báo hướng dẫn cho trường này*

Trả lời:

- Hoặc một vài dòng hướng dẫn cho việc nhập dữ liệu cho trường, rồi ấn <CR>. Lúc nhập dữ liệu ta có thể yêu cầu cho hiện ra thông báo này bằng cách ấn phím <F1>.

- Hoặc <CR> nếu không muốn đưa vào thông báo hướng dẫn nào.

Thông báo 7: *Đưa vào giá trị định trước*

Trả lời:

- Hoặc một nội dung cố định của trường, rồi ấn <CR>;

- Hoặc ấn <CR> nếu không đưa vào giá trị định trước nào.

Bước này sẽ kết thúc việc xác định một trường trong biểu mẫu. CDS-ISIS lại cho hiện thông báo đầu tiên để ta xác định trường tiếp theo, nếu có.

Khi biểu mẫu đã hoàn thành hay muốn xác định trang sau, ta trả lời <CR> cho thông báo này. Ở cuối màn hình sẽ hiện ra một menu con như sau:

| <CR> - Trang sau | B - Trang trước | R - Khôi phục |

| E - Sửa đổi | S - Ra và lưu | C - Huỷ bỏ |

Trả lời:

<CR>: để tạo một trang mới;

B: xem lại hoặc sửa đổi trang trước, nếu có;

S: lưu biểu mẫu vừa tạo ra và chuyển sang bước tiếp theo;

C: kết thúc xử lý nhưng không lưu biểu mẫu;

R hoặc E: để sửa đổi biểu mẫu. Khi đó một menu con sẽ hiện ra như sau:

| A- Thêm trường | R- Làm phẳng phải | S- Dịch chuyển | D- Xoá | C- Vào giữa |

| <CR>- Trường sau | <TAB>- Trường trước | M- Sửa đổi | X- Thoát ra |

Nguyên tắc: Để đưa con trỏ về trường cần sửa đổi dùng phím <CR> và <TAB>. Trường này có thể:

M: sửa đổi; khi đó CDS-ISIS sẽ trở về chế độ tạo biểu mẫu như trên.

R: làm phẳng lề phải (nếu chỉ có một trường trên dòng).

S: dịch chuyển; ở đây tất cả các trường bắt đầu từ vị trí con trỏ được dịch chuyển lên hoặc xuống n dòng bằng cách trả lời thông báo:

Thông báo: ***Đưa vào số dòng cần chuyển dịch***
(n = xuống dưới, - n = lên trên)

D: xoá trường.

C: đưa vào giữa dòng (nếu chỉ có một trường trên dòng).

X: kết thúc việc sửa đổi và trở về menu trên:

| <CR> - Trang sau | B - Trang trước | R - Khôi phục |
| E - Sửa đổi | S - Ra và lưu | C - Hủy bỏ |

Bây giờ biểu đã lập xong. Gõ S để lưu và chuyển sang bước tạo format hiện hình.

Bước 3- Tạo format hiện hình

CDS-ISIS sử dụng ngôn ngữ tạo format để trình bày dữ liệu trên màn hình. Ta chỉ cần viết một dãy các lệnh của ngôn ngữ này để CDS-ISIS thực hiện (từ trái sang phải).

Màn hình dưới đây hiện ra; chương trình soạn thảo trường hoạt động.

Tên CSDL: SACH	Tên format: SACH
EDIT: Thay thế	

Ta chỉ cần gõ vào văn bản của format. Muốn vậy ta phải chuẩn bị trước và phải tuân theo các quy tắc đã trình bày ở phần trên. Sau khi gõ vào các dòng của format, bạn ấn <CR> để lưu. Tên mặc định của format là tên của CSDL.

Ví dụ: Dưới đây là format của CSDL SACH:

Tên CSDL:SACH	Tên format :SACH
<pre> mfn(4),mdl, '-', v1+ ; / mhl,v2(10,10), '--', v3^a(10,10), '-', v3^b(10,10),',', v3^c(10,10),', / v4^a(10,10), if p(v4^b) then '-',v4^b(10,10) fi, '-', v4^c(10,10), '-', if p(v5) then '-,' (, v5(10,10), ') fi/ "Tóm tắt:",v8(10,10)/ "Từ khoá: ",mpl, v9(10,10)/#,"c10,"Ký hiệu PL:",v6,c40, "Ký hiệu kho:", v7^a,/,v7^b /## </pre>	

Chương trình chuyển sang bước lập bảng chọn trường.

Bước 4- Lập bảng chọn trường

Trên màn hình hiện ra bảng sau đây và chương trình soạn thảo dòng bắt đầu hoạt động.

Tên CSDL: SACH		FST cho tệp đảo	Tên FST: SACH
? ID	IT	Format tách dữ liệu	
---	---	-----	
Đưa vào/soạn thảo dấu nhận biết trường			
EDIT: Thay thế			

Ta cần xác định trường nào là trường được chọn để tạo tệp đảo và kỹ thuật đảo. Lần lượt theo các cột, ta xem phải xác định các yếu tố đó như thế nào.

ID: đưa vào nhãn của trường dùng để tạo tệp đảo (theo bảng xác định trường);

IT: đưa vào kỹ thuật đảo.

CDS-ISIS có thể đảo:

- Theo dòng văn bản (nói chung là cả một trường): kỹ thuật 0
- Theo dòng hoặc trường con: kỹ thuật 1
- Theo thuật ngữ đặt giữa dấu < >: kỹ thuật 2
- Theo thuật ngữ đặt giữa dấu / \: kỹ thuật 3.

Chú ý: Các kỹ thuật 1, 2, 3 phải kết hợp với chế độ MPL (hoặc MPU) trong format tách dữ liệu.

- Theo từng từ của văn bản: Kỹ thuật 4

Từ CDS/ISIS version 3.08 bạn có thể sử dụng 4 kỹ thuật đảo mới cho phép xác định tiền tố cho mỗi thuật ngữ được tách ra từ các kỹ thuật đảo 1, 2, 3, 4. Cũng giống như trong WinISIS các kỹ thuật này được đánh số tương ứng là 5, 6, 7, 8.

Format tách dữ liệu:

Trước khi đưa các thuật ngữ vào tệp đảo, CDS/ISIS "đọc" nội dung của trường theo format đã chỉ ra trong vùng này. Từ đó tách ra các thuật ngữ theo kỹ thuật đảo đã chọn. Format tách dữ liệu được viết như format trình bày dữ liệu.

Bảng chọn trường FST của CSDL SACH giả sử được lập như sau:

Tên CSDL: SACH		FST cho tệp đảo	Tên FST: SACH
? ID	IT	Format tách dữ liệu	
1	0	mhl,(V1/)	
2	4	mhl,V2	
3	0	mhl.V3^a/V3^b/V3^c	
8	3	mpl,v8	
9	2	mpl,v9	

Đến đây cấu trúc của CSDL đã được xác định. Ta cần kiểm tra xem cấu trúc đó được xây dựng chính xác theo yêu cầu đặt ra chưa. Muốn vậy ta phải trở về menu chính, chọn menu cập nhật dữ liệu và nhập vào 3 hoặc 4 biểu ghi. Sau đó từ menu tìm, cho hiện hình các dữ liệu và từ điển các thuật ngữ đã được chọn để đưa vào tệp đảo. Việc sửa đổi cấu trúc của CSDL liên quan tới một trong bốn tệp đã xác định ở trên sẽ được trình bày dưới đây.

2. SỬA ĐỔI CẤU TRÚC CSDL

Ta có thể sửa đổi các tệp xác định cấu trúc CSDL ở bất kỳ thời điểm nào. Sau khi chọn tùy chọn D của menu chính, chọn

M của menu VSDEF. Menu dưới đây sẽ hiện ra trên màn hình. Việc sửa đổi sẽ được thực hiện khi ta chọn những chức năng thích hợp.

Ta có thể tạo lập các biểu mẫu nhập dữ liệu và các format hiện hình phụ trợ bằng cách sử dụng tương ứng các tùy chọn B và H của menu này.

Tùy chọn K cho ta danh sách các tệp đã được xác định cho một CSDL. Có thể dùng tùy chọn này một cách thuận tiện để sửa, xóa hay sao một tệp kể trên thay cho việc sử dụng các tùy chọn tương ứng của menu VSDBU.

Phương tiện ISISEF	Các phương tiện xác định CSDL	Menu VSDBU
a-	Cập nhật bảng xác định trường	
b-	Tạo cập nhật biểu mẫu	
c-	Sao biểu mẫu	
d-	Xoá biểu mẫu	
e-	Tạo/ cập nhật FST	
f-	Sao bảng chọn trường	
g-	Xoá bảng chọn trường	
h-	Tạo/cập nhật Format hiện hình	
i-	Sao format hiện hình	
j-	Xoá format hiện hình	
k-	Xem/sửa các tệp xác định cấu trúc CSDL	
x-	Trở về menu trước	
Tên CSDL: SACH		Biểu mẫu: SACH
Số biểu ghi: 12		Format: SACH

Tạo lập tệp format

Menu trên cũng cho phép tạo lập những tệp chứa format hiện hình khác nhau. Ta có thể sử dụng các format hiện hình này, xuất phát từ tùy chọn F trong menu các phương tiện tìm tin.

Để tạo lập một tệp format, ta có thể tiến hành bằng hai cách sau:

- Tùy chọn I: để sao chép tệp format đã có. Chương trình yêu cầu đưa vào tên các tệp cần chép và tên tệp muốn tạo ra. Màn hình hiện ra các lệnh của format. Ta không sửa đổi gì trên format này mà chỉ ấn <CR> để ra.

- Tùy chọn H: để tạo một format mới. Ta đưa vào tên format mới, một màn hình hiện ra. Ta gõ vào các lệnh theo ngôn ngữ format rồi ấn <CR> để ra.

Ví dụ, với CSDL SACH ta tạo thêm format THUMUC với các dòng lệnh sau:

```
mhl,v2(5,5), ' ' v1(5,5)+|;|, '- ', v3^a(5,5), ':', v3^b(5,5), ',', v3^c(5,5),  
'-', v4^a(5,5) , if p(v4^b) then ':', v4^b(5,5) fi, ',', v4^c(5,5), ',', if p(v5)  
then '- ', '(', v5(5,5), ')' fi/c12, "Ký hiệu kho: " v7^a,',v7^b(10,10), c12,  
"Ký hiệu PL: " v6/c12,"Từ khoá: ",mhl,v9(24,24)/#
```

Format này sẽ được sử dụng khi cần in thông báo thư mục của CSDL SACH, trong đó không có trường tóm tắt.

3. CẬP NHẬT DỮ LIỆU

Các phương tiện nhập dữ liệu ISISENT của CDS-ISIS cung cấp mọi chức năng liên quan đến việc cập nhật dữ liệu như: tạo

biểu ghi mới, xoá hay sửa các biểu ghi đã có. Chú ý rằng các chức năng này chỉ áp dụng đối với CSDL đã được chọn.

Để cập nhật dữ liệu, từ menu chính ta chọn E. Menu sau đây sẽ hiện ra trên màn hình.

Phương tiện ISISENT Các phương tiện cập nhật dữ liệu Menu VXE1	
L- Chọn ngôn ngữ làm việc	
W- Chọn biểu mẫu khác	
N- Tạo biểu ghi mới	
E- Hiệu đính biểu ghi	
R- Hiệu đính kết quả tìm cuối cùng	
P- Gọi lại biểu ghi cuối cùng đã sửa	
D- Xác định các giá trị định trước	
C- Xoá giá trị định trước	
X- Trở về menu trước	
?_	
Tên CSDL: SACH	Biểu mẫu: SACH
Số biểu ghi: 12	Format: SACH
Micro CDS/ISIS -(C) Unesco 1997	

Tạo biểu ghi mới

Để tạo biểu ghi mới cho CSDL, từ menu VXE1 ta chọn chức năng N. Khi đó biểu mẫu nhập dữ liệu đã lập ở trên sẽ hiện ra như hình dưới đây:

PHIẾU NHẬP TIN

Tác giả:	
Tên sách:	
Xuất bản:	
Mô tả vật lý:	
Tùng thư:	
Ký hiệu PL:	
Ký hiệu kho:	
Tóm tắt:	
	
	
Từ khoá:	
	
Hiệu đính: chèn	Tiếp trang sau
	MFN=1

Ta nhập dữ liệu theo từng trường vào biểu mẫu trên. Để nhập dữ liệu nhanh chóng, ta cần nắm vững cách sử dụng các phím sau:

<CR>: Chuyển sang trường sau

<TAB>: Chuyển về trường trước

<F1>: Hiện thông báo trợ giúp

<F6>: Xóa tất cả các ký tự của trường từ vị trí con trỏ.

Các thao tác trên khối cho phép chép nội dung của một trường vào một trường khác. Để làm được điều đó ta sử dụng 3 phím chức năng sau đây:

<F3>: Đánh dấu đầu khối

<F4>: Đánh dấu cuối khối bị xóa khỏi màn hình. Để chèn khối vào cùng một vị trí khác, ta ấn phím:

<F5>: Chèn khối vào vị trí con trỏ. Ta có thể chép lại khối đó bao nhiêu lần cũng được vào chính biểu ghi đó hay vào một biểu ghi khác. Nội dung của khối được lưu lại đến lần xác định sau của khối.

Khi trang đầu tiên của biểu ghi được nhập, một menu con sau đây sẽ hiện ra ở cuối màn hình:

|<CR>-Trang sau|B- Trang trước | M- Sửa đổi | N- Biểu ghi mới |
|X- Ra và lưu | D- Xoá | C- Huỷ bỏ | T- Kết thúc |

Trả lời:

<CR>: Sang trang sau của biểu thức mẫu: nếu ta đang ở trang cuối thì <CR> tương đương với X.

B: Về trang trước của biểu mẫu.

M: Sửa đổi nội dung vừa cập nhật khi đó con trỏ được đặt ở trường thứ nhất.

N: Lưu biểu ghi vào tệp chủ và sau đó tạo biểu ghi mới; khi đó biểu ghi đã cập nhật được ghi lại và một biểu mẫu rỗng sẽ hiện ra. Số MFN được tăng thêm 1.

C: Biểu ghi không thay đổi (mọi sửa đổi đều bị bỏ qua).

X: Lưu biểu ghi vào tệp chủ và trở về menu.

D: Xoá nội dung biểu ghi.

T: Xoá biểu ghi cuối cùng vừa cập nhật và trở về menu.

Ta sẽ thấy rằng khi đổi biểu ghi, một số trong các tùy chọn này có thay đổi ý nghĩa một chút.

Nguyên tắc viết một vài trường

Trường lập: Nếu là trường lập thì một phần tử dữ liệu phải cách nhau bởi dấu %, không có ký tự trống trước và sau dấu %. Ví dụ với trường tác giả có hai tác giả:

Bảng Việt%Lưu Quang Vũ

Trường con: Nếu trường có trường con thì mỗi trường con phải được đặt trước bởi dấu ^ theo sau là một chữ như đã xác định trong bảng chọn trường, không có ký tự trống trước và sau dấu phân cách.

Ví dụ: ^aHà Nội^bVăn học^c1987

Trường kiểu khuôn mẫu: Nếu một trường như thế đã được xác định, thì chương trình sẽ đưa ra khuôn mẫu của trường. Ta phải nhập dữ liệu theo khuôn mẫu đó. Khuôn mẫu này đã được xác định trong bảng chọn trường, đó là một dãy các ký tự liên tiếp xác định vị trí và kiểu ký tự của dữ liệu, theo quy ước như sau:

X: Vị trí có thể chứa ký tự là số hoặc chữ

A: Vị trí phải chứa ký tự là chữ

9: Vị trí phải chứa ký tự là số

Ký tự khác: Vị trí phải chứa chính ký tự đã chỉ ra.

Ví dụ: Khuôn X-99A thì dữ liệu vào phải là 3-58X hoặc C-72R...

Hiệu đính biểu ghi

Muốn hiệu đính một biểu ghi hay một số biểu ghi, ta chọn E trong menu trên. Một thông báo hiện ra ở cuối màn hình.

MFN hay dãy MFN cần sửa đổi (n1 n2)?

Ta gõ vào số của biểu ghi cần sửa đổi hoặc một dãy các biểu ghi cần sửa đổi bằng cách gõ vào số của biểu ghi đầu, khoảng trống, số của biểu ghi cuối dãy.

Một biểu mẫu hiện ra với nội dung của biểu ghi đã được chọn. Số của biểu ghi được chỉ ra ở góc dưới bên phải màn hình. Một menu con tương tự như menu con đã giới thiệu ở trên hiện ra. Ta sẽ phải thực hiện các thao tác sửa đổi theo sự chỉ dẫn của menu đó.

Xóa biểu ghi

Tuỳ chọn D của menu con trên cho phép ta xoá nội dung của biểu ghi đã được chọn. Chú ý rằng biểu ghi chỉ bị xoá về mặt logic chứ không phải về mặt vật lý. Ở đây biểu ghi chỉ bị đánh dấu để sau này sẽ bị xoá về mặt vật lý và ta không thể tra cứu biểu ghi này được nữa. Các biểu ghi này sẽ bị xoá về mặt vật lý khi tệp chủ được tổ chức lại. Trước thời điểm đó ta còn khả năng phục hồi lại chúng bằng cách gọi lại số của biểu ghi đó để hiệu đính. Khi đó thông báo sau đây sẽ hiện ra:

R- Phục hồi | N- Tạo mới | <CR>- bỏ qua

Trả lời:

<CR>: Biểu ghi vẫn bị đánh dấu xoá.

N: Thay biểu ghi bị xoá bằng biểu ghi mới; mang số của biểu ghi đó.

R: Phục hồi lại biểu ghi đã bị đánh dấu xoá.

Cập nhật tệp đảo

Sau mỗi lần tạo ra các biểu ghi mới, ta cần cập nhật ngay tệp đảo vì khi thao tác này chưa được thực hiện thì ta chưa thể tìm được các biểu ghi này bằng các dấu hiệu tìm tin được lưu trong tệp đảo.

Các phương tiện bảo trì tệp đảo ISISINV cho phép ta thực hiện các chức năng khác nhau trên tệp đảo đối với một CSDL đang được chọn như: tạo tệp đảo (tùy chọn F), cập nhật tệp đảo (tùy chọn U) v.v... Tuy nhiên sau mỗi lần tạo ra biểu ghi mới hoặc có sự sửa đổi biểu ghi cũ thì CDS - ISIS luôn đưa ra thông báo:

Có muốn cập nhật tệp đảo bây giờ không (Y/N) ?

Nếu trả lời Y, CDS/ISIS sẽ cập nhật tệp đảo.

4. TÌM VÀ XEM DỮ LIỆU

Việc tìm và xem dữ liệu được thực hiện từ tùy chọn S của menu chính. Khi đó menu ***các phương tiện tìm tin*** sẽ hiện ra:

Phương tiện ISISRET	Các PT tìm tin	Menu FXGEN
	L- Chọn ngôn ngữ làm việc B- Xem tệp chủ T- Xem từ điển S- Tạo biểu thức tìm D- Xem kết quả tìm F- Thay đổi format hiện hình R- Xem lại các biểu thức tìm G- Tìm lại theo biểu thức trước	

P- Lưu kết quả tìm
X- Trở về menu trước
?

Tên CSDL: SACH

Biểu mẫu: SACH

Số biểu ghi: 12

Format: SACH

Micro CDS/ISIS – (C) Unesco 1997

4.1. Xem tệp chủ

Ta có thể xem một đoạn tệp chủ bắt đầu từ biểu ghi với số MFN xác định. Từ menu các phương tiện tìm tin ta chọn B. Thông báo sau đây sẽ hiện ra:

Bắt đầu xem từ MFN nào?

Trả lời:

- Đưa vào số MFN của biểu ghi đầu tiên cần xem rồi ấn <CR>.

- <CR> để xem từ đầu tệp chủ.

Ví dụ: Với format hiện hình đã lập, biểu ghi số 12 của CSDL SACH sẽ được trình bày dưới dạng sau:

0012- Trần Quốc Vương.

Văn hoá Việt Nam-Tìm tòi và suy nghĩ.-H.:Văn hoá Dân tộc,
2000.

984 tr. ; 21 cm

Tóm tắt: Thông qua các /lễ hội dân gian/, /phong tục tập quán/, /danh nhân văn hoá/ tác giả trình bày những tìm tòi và suy nghĩ về /bản sắc văn hoá Việt Nam/.

Từ khoá: <văn hoá><văn hoá dân gian><Việt Nam>

Ký hiệu PL: P3(1)73

Ký hiệu kho: VL402/2001

Chú ý rằng biểu ghi trên không có từng thư và minh hoạ. Nếu biểu ghi có từng thư và minh hoạ nhưng không có tóm tắt thì nội dung của nó sẽ được trình bày như ví dụ sau:

0008- Phạm Văn Lâm; Ngô Văn Thảo.

Cơ học lý thuyết.-H.: Khoa học Kỹ thuật, 1995.

245 tr.-(minh hoạ;19 cm.-(Tủ sách khoa học kỹ thuật)

Từ khoá: <cơ học><cơ lý thuyết>

Ký hiệu PL: E101.3

Ký hiệu kho:Vb433/1998

Các biểu ghi được đưa ra tuần tự theo format trình bày đã chọn (xem tùy chọn F dưới đây). Để ngừng xem và trở về menu, ấn X.

4.2. Tìm tin bằng cách viết trực tiếp biểu thức tìm

Từ menu các phương tiện tìm tin, chọn S. Thông báo sau đây sẽ hiện ra ở góc trên màn hình:

Đưa vào biểu thức tìm?

Ta gõ vào biểu thức tìm bằng chương trình soạn thảo trường, rồi ấn <CR> để thực hiện tìm theo biểu thức đó.

Ví dụ: Để tìm thơ của Xuân Diệu hoặc Nguyễn Đình Thi, ta gõ:

Thơ*(Xuân Diệu +Nguyễn Đình Thi) <CR>

Để tìm các tập sách xuất bản sau năm 1995, ta gõ vào biểu thức tìm tự do:

?val(v3^c)>1995 <CR>

Khi thực hiện CDS/ISIS sẽ cho biết số lượng các biểu ghi ứng với từng thuật ngữ trong biểu thức tìm (ký hiệu là P=), số lượng các biểu ghi ứng với các biểu thức con của biểu thức tìm (ký hiệu là T=).

CDS/ISIS sẽ gán cho mỗi biểu thức tìm một số hiệu duy nhất. Khi được sử dụng số hiệu của biểu thức tìm được được viết sau dấu #.

Để hiển thị kết quả tìm, ta chọn D. Khi đó trên màn hình sẽ hiện ra các biểu ghi tìm được với format hiện hành đã chọn.

Muốn in kết quả tìm trước hết ta phải lưu kết quả tìm bằng cách chọn P từ menu các phương tiện tìm tin. Thông báo sau đây sẽ hiện ra:

Hãy đưa vào tên tệp lưu

Trả lời: Gõ vào tên dài nhiều nhất 6 ký tự. Nếu tên này đã tồn tại, CSD/ISIS sẽ thông báo:

Tệp lưu đang tồn tại. OK để xoá (Y/N)?

Trả lời:

Y: để dùng lại tên chính tệp lưu đó (Xoá tệp cũ)

N: để bỏ qua yêu cầu và cho hiện lại menu.

Thay đổi format hiển hình

Kết quả tìm được hiển thị theo format đã lập của CSDL (gọi là format mặc định, nó mang tên của CSDL). Tuy nhiên đó không phải là cách trình bày dữ liệu duy nhất. Trong thực tế ta có thể sử dụng nhiều format đã được lưu trên đĩa và có thể sửa đổi format đang sử dụng.

Nếu muốn sửa format đang làm việc, hãy trở về menu và chọn F. Khi đó format hiện hình sẽ hiện ra. Ta có thể sửa format đó và có thể kiểm tra ngay format đã sửa bằng cách xem tệp chủ với tùy chọn B. Mỗi khi dùng tùy chọn này để sửa format, các thay đổi chỉ là tạm thời, cho đến khi ta ra khỏi chương trình và chọn format khác. Để lưu thay đổi ta phải ấn phím <F8>. Tuy nhiên phải thận trọng khi sử dụng phương tiện này vì nếu lưu format có lỗi, CDS/ISIS sẽ đưa ra thông báo lỗi và người chưa thành thạo ngôn ngữ tạo format sẽ khó sửa chữa được.

Nếu trong CSDL đã có nhiều format hiện hình, ta có thể chọn một format mong muốn từ tùy chọn F bằng cách đưa vào tên của nó đặt sau dấu @. Để xoá khỏi màn hình nội dung của format trước, ấn <F6>.

Việc tạo format hiện hình mới được thực hiện từ chức năng ***sửa đổi cấu trúc CSDL*** (tùy chọn U) trong menu ***Xác định CSDL***.

Ví dụ khi sử dụng format THUMUC (gọi vào bằng lệnh @THUMUC) thì các biểu ghi 0012 và 0008 của CSDL SACH sẽ được trình bày như sau:

Văn hoá Việt Nam-Tìm tòi và suy nghĩ./ Trần Quốc Vương.-H.: Văn hoá Dân tộc, 2000.-984 tr. ; 21 cm

Ký hiệu kho: VL402/2001

Ký hiệu PL: P3(1)73

Từ khoá: Văn hoá; Văn hoá dân gian; Việt Nam

Chú ý rằng biểu ghi trên không có từng thư và minh hoạ. Nếu biểu ghi có từng thư và minh hoạ, nhưng không có tóm tắt thì nội dung của nó sẽ được trình bày như ví dụ sau:

Cơ học lý thuyết/ Phạm Văn Lâm;Ngô Văn Thảo..-H.: Khoa học Kỹ thuật, 1995.- 245 tr.:minh hoạ;19 cm.-(Tủ sách khoa học kỹ thuật)

Ký hiệu kho: Vb433/1998

Ký hiệu PL: E101.3

Từ khoá: Cơ học; Cơ lý thuyết

4.3. Tìm tin theo từ điển

Chọn T từ menu **Các phương tiện tìm tin** một màn hình trống hiện ra với thông báo ở cuối màn hình yêu cầu đưa vào khoá cần tìm. Nếu trả lời bằng <CR> thì một từ điển các từ khoá sẽ hiện ra. Ta có thể lật xem các trang của từ điển và chọn các thuật ngữ tìm theo cách sẽ được giải thích dưới đây.

Các phím lật trang

<PgDn> hay <PgUp>: cho xem trang sau hoặc trang trước của từ điển.

T: Cho phép xem phần khác của từ điển. Thông báo trên sẽ hiện ra để đưa từ khoá đầu vào.

Các phím chọn thuật ngữ

Các phím mũi tên: cho phép đưa con trỏ về thuật ngữ cần tìm.

S: để chọn thuật ngữ dưới con trỏ. Nếu chọn nhiều thuật ngữ mà không chỉ rõ các phép toán, CSD/ISIS sẽ tự động chèn vào giữa các thuật ngữ toán tử OR (+).

***,+:** Để nối một thuật ngữ với thuật ngữ đã chọn từ trước bằng toán tử đã chỉ ra.

Các phím để ra

C: Bỏ qua sự lựa chọn thuật ngữ vừa làm và trở về menu.

X: Để ra khỏi từ điển các thuật ngữ và hiển thị biểu thức tìm.

Khi biểu thức tìm đã hiện ra trên màn hình, ta có thể sửa đổi và hoàn thiện biểu thức tìm, rồi ấn <CR> để thực hiện biểu thức đó.

Tìm theo từ điển tương đương với cách tìm có trợ giúp trong WinISIS.

5. SẮP XẾP VÀ IN KẾT QUẢ TÌM TÌN

Ta có thể in ra kết quả của một yêu cầu tìm tin hoặc in một tập các biểu ghi được chọn theo số hiệu của chúng. Trong trường hợp thứ nhất, kết quả tìm phải được lưu từ trước và khi in ta phải gọi lại tên tệp lưu này. Trong trường hợp thứ hai, ta cần chỉ ra một dãy số hiệu của các biểu ghi cần in. Ta có thể in các kết quả theo một trật tự sắp xếp tùy ý. Các kết quả có thể in ra máy in hay in ra một tệp lưu trên đĩa. Việc trình bày kết quả in được xác định bởi format in. Format này được lập từ trước hay được xác định khi đưa vào các tham số in. Trong trường hợp thứ nhất, ta phải đưa vào tên format, đặt sau dấu @. Trong trường hợp đưa vào bản thân format, ta phải nắm vững ngôn ngữ tạo format của CDS/ISIS.

Mỗi khi sử dụng các phương tiện sắp xếp và in, hệ thống yêu cầu ta cung cấp các tham số trình bày và sắp xếp cần thiết để in. Các tham số này được đưa vào qua hai biểu mẫu hệ thống gọi là biểu mẫu in và biểu mẫu sắp xếp.

5.1. Biểu mẫu in hệ thống

Từ menu chính chọn P – Sắp xếp và in. Trên màn hình sẽ hiện ra menu con Sắp xếp và in. Từ menu này lại chọn P thì **Biểu mẫu in hệ thống** sẽ hiện ra.

Tên CSDL	_____	Giới hạn MFN 1/32000	Tên tệp lưu	_____	
Tiêu đề 1	_____				
Tiêu đề 2	_____				
Tiêu đề 3	_____				
Format in @	_____				
Độ dài dòng	70	Số lượng cột	1	Độ rộng cột	70
Số dòng/trang	60	Số trang đầu tiên	1	Khoảng trống tối thiểu	3
Độ lùi dữ liệu	0	Sắp xếp ?	N	Tên biểu mẫu sắp xếp	_____
				Tên tệp in	_____
EDIT: Chèn	Trang sau				

Điền biểu mẫu này giống như điền biểu mẫu nhập dữ liệu. Một số trường chứa những giá trị mặc định. Ta có thể sửa các giá trị đó.

Tên CSDL: Chương trình sẽ cho hiện ra tên CSDL đang làm việc.

Giới hạn MFN: Dùng để chỉ ra giới hạn của các biểu ghi cần in. Ví dụ muốn in từ biểu ghi số 20 đến biểu ghi số 50 ta gõ vào: 20/50. Trong trường hợp cần in kết quả tìm thì không cần sửa đổi giới hạn này.

Tên tệp lưu: Nếu muốn in kết quả tìm đã được lưu thì điền vào trường này **tên tệp lưu**. Nhớ rằng kết quả tìm có thể lưu bằng tùy chọn P trong menu các phương tiện tìm tin.

Tiêu đề 1, 2, 3: Để in một, hai hoặc ba tiêu đề trên trang đầu của báo cáo. Để in dòng tiêu đề này trên mỗi trang, hãy bắt đầu dòng này bằng một dấu * (dấu * sẽ không được in ra).

Format in: Để trình bày các trường cần in và cách trình bày các trường đó. Có thể đưa vào tên một format đã xác định từ trước (tên phải đặt sau dấu @) hoặc viết trực tiếp một format mới. Để thuận tiện ta nên xác định trước một vài format, kiểm tra lại chúng để có thể sử dụng khi cần thiết.

Độ dài dòng: Xác định độ rộng của văn bản bằng số ký tự được in ra theo dòng. CDS/ISIS đưa ra giá trị mặc định là 70 ký tự.

Số lượng cột: Để xác định số lượng cột trên trang in. Trong thực tế ta thường dùng hai cột để in các chỉ mục còn nói chung chỉ dùng một cột là đủ.

Độ dài cột: Xác định số ký tự tối đa của mỗi cột, số này phải lớn hơn hoặc bằng độ rộng của dòng (số ký tự dùng trong một dòng của cột). Hiệu số giữa độ rộng cột và độ rộng dòng cho ta khoảng trống giữa hai cột.

Số dòng/trang: Để xác định số dòng tối đa trên một trang in. Mặc định là 60 dòng.

Số trang đầu tiên: Để xác định cách đánh số trang. Có các tùy chọn sau:

- 0 (giá trị mặc định): Không đánh số trang, nhưng mỗi trang chứa số dòng như đã chỉ ra ở trên.

- 1 (hay một số khác): Số trang sẽ được in ra ở giữa và phía dưới mỗi trang và việc đánh số trang được bắt đầu từ số đã xác định.
- N: In nội dung văn bản, không đánh số trang. Tùy chọn này dùng để tạo ra các tệp văn bản nhằm xử lý bên ngoài CDS/ISIS.

Khoảng trống tối thiểu: Để xác định số dòng trống tối thiểu cần có trên cột hoặc trang hiện tại trước khi in biểu ghi. Số mặc định là 3.

Độ lùi dữ liệu: Để chỉ độ lùi vào của lề trái hay trong trường hợp sắp xếp là sự lùi vào của văn bản sau mức tiêu đề cuối cùng.

Sắp xếp: Để chỉ ra ta có muốn sắp xếp các biểu ghi trước khi in hay không. Câu trả lời ngầm định là N (không). Khi đó các biểu ghi sẽ được in theo thứ tự tăng dần của MFN. Nếu trả lời Y (có) thì ta phải điền tiếp các tham số cần thiết vào biểu mẫu sắp xếp.

Tên biểu mẫu sắp xếp: Để chỉ ra một biểu mẫu sắp xếp đã định trước. Nếu để trống, CDS/ISIS sẽ đưa ra biểu mẫu sắp xếp hệ thống. Chú ý rằng trường này chỉ có nghĩa khi bấm Y ở trường trước.

Tên tệp in: Có thể là:

- LPT1 để chuyển kết quả ra máy in (LPT1 tương ứng với máy in trong DOS)
- Tên tệp in để ghi kết quả trên một tệp, lưu trong thư mục \ISIS\WORK
- CON để chuyển kết quả ra màn hình.

Khi đã điền xong biểu mẫu in, menu con sau đây sẽ hiện ra:

| <CR> - Ra | M - Sửa đổi | C - Huỷ bỏ |

Trả lời:

<CR>: Để khẳng định CDS/ISIS sẽ bắt đầu in hoặc đưa ra biểu mẫu sắp xếp (nếu đòi hỏi).

M : Để sửa các tham số, nếu cần.

C : Huỷ việc in và quay về menu.

5.2. Biểu mẫu sắp xếp hệ thống

Nếu muốn kết quả đầu ra được sắp xếp, trả lời Y ở trường sắp xếp của biểu mẫu in. Sau khi biểu mẫu này được điền xong, biểu mẫu sắp xếp hệ thống sẽ hiện ra để ta điền vào các tham số sắp xếp.

Cũng giống như WinISIS, trong biểu mẫu sắp xếp ta phải khai báo các tham số sau:

- Số lượng tiêu đề
- Format tiêu đề
- Độ dài của khoá 1, 2, 3, 4
- FST của khoá 1, 2, 3, 4

Ví dụ:

Giả sử cần in mục lục sắp xếp các cuốn tiểu thuyết của CSDL DHVH, theo tác giả và nhan đề, cùng với kết quả in ra là các yếu tố xuất bản, số trang, ký hiệu kho và số thứ tự biểu ghi MFN.

Trước hết ta phải tìm tất cả các cuốn tiểu thuyết của CSDL, lưu kết quả tìm trên một tệp, sau đó sắp xếp và in kết quả tìm

Kết quả tương ứng được in ra như sau:

MỤC LỤC SẮP XẾP TIỂU THUYẾT THEO TÁC GIẢ VÀ NHÂN ĐỀ

Cơ sở dữ liệu DHVH

Chu Văn

Bão biển

Hà Nội, Văn học, 1978.-516 tr.

Ký hiệu kho: V23.V115/MFN=0018

Đất mặn

Hà Nội, Thanh niên, 1975.-397 tr.

Ký hiệu kho: V23.V116/MFN=0021

Huy Phương

Xi măng

Hà Nội, Văn học, 1968.-335 tr.

Ký hiệu kho: V23.H532P/MFN=0037

Hồ Phương

Những tâm cao

Hà Nội, Quân đội Nhân dân, 1973.- 423 tr.

Ký hiệu kho: V23.H450PH/MFN=0036

v.v...

6. TRAO ĐỔI DỮ LIỆU

Cũng giống như với WinISIS, nguyên tắc trao đổi dữ liệu trong CDS/ISIS 3.08 dựa trên format trao đổi ISO.2709.

6.1. Xuất dữ liệu

Từ menu chính chọn M để vào menu Trao đổi dữ liệu và bảo trì tệp chủ:

Phương tiện ISISXCH	Trao đổi dữ liệu và bảo trì tệp chủ	Menu VXCH
	L- Chọn ngôn ngữ làm việc	
	B- Sao lưu tệp chủ	
	R- Khôi phục tệp chủ	
	C- Tổ chức lại tệp chủ	
	I- Nhập tệp từ ngoài hệ thống	
	E- Xuất tệp của CDS/ISIS	
	X- Trở về menu trước	
Tên CSDL: SACH		Biểu mẫu: SACH
Số biểu ghi: 12		Format: SACH
Micro CDS/ISIS – (C) Unesco 1997		

Chọn tiếp E, biểu mẫu hệ thống dùng để xuất dữ liệu hiện ra như hình dưới đây. Ta điền trang này như điền biểu mẫu nhập tin. Có thể dùng phím <F1> để tham khảo cho mỗi trường.

Trong các tham số của tệp ISO, ta chỉ thay đổi, nếu cần, tên tệp ISO ở đầu ra (tên mặc định là MST.ISO). Các tham số khác chỉ thay đổi nếu cần truyền dữ liệu qua mạng viễn thông.

Trao đổi dữ liệu – Các tham số xuất

Tên CSDL _____

Các tham số tệp ISO

Tên tệp ISO đầu ra MST.ISO

Dấu ngắt trường #

Dấu ngắt biểu ghi #

Các tham số lựa chọn

Giới hạn MFN 1/32000 Tên tệp lưu_____ Tệp kết quả (Y/N)? N

Các tham số để format lại

Tạo lại bảng FST_____ Tệp chuyển đổi Gizmo_____

Đánh số lại từ _____

Các tham số lựa chọn dùng để xác định phần CSDL xuất ra có thể là:

- Một khoảng các biểu ghi: đưa và giới hạn MFN của các biểu ghi.
- Các biểu ghi đã được lựa chọn qua tìm tin: đưa vào tên tệp lưu.
- Các biểu ghi đã sắp xếp, được lưu trong tệp kết quả (Hit file, được tạo ra bằng chương trình ISISPT): gõ Y vào trường này.

Trong trường hợp xuất đơn giản, có thể bỏ qua các tham số format lại. Chúng chỉ được sử dụng để sửa đổi cấu trúc của CSDL hay khi chỉ xuất một vài trường của CSDL.

6.2. Nhập dữ liệu

Biểu mẫu hệ thống để nhập dữ liệu từ tệp ISO hiện ra mỗi khi chọn I trong menu trao đổi dữ liệu và bảo trì tệp chủ.

Trao đổi dữ liệu – Các tham số nhập	
Tên CSDL _____	
Các tham số tệp ISO	
Tên tệp ISO đầu vào	MST.ISO
Dấu ngắt trường	#
Dấu ngắt biểu ghi	#
Các tham số lựa chọn	
Số MFN đầu tiên 1	Nhấn vào có chưa MFN _____
Các tùy chọn nhập vào (L/M/U)	
Các tham số để format lại	
FST để format lại _____	Tệp chuyển đổi Gizmo _____

Trường đầu tiên của biểu mẫu là tên CSDL tiếp nhận. Mặc định là CSDL hiện hành. Tiếp theo là tên tệp ở format ISO mà ta nhập vào.

Các tùy chọn nhập vào cho phép ta xác định cách thức nhập. Cũng tương tự như trong WinISIS ta có các phương thức nhập sau:

- Nhập tệp ISO vào CSDL và xóa đi nội dung của nó nếu nó không rỗng, chọn L (Tải)
- Trộn các tệp ISO với CSDL, không thay thế các biểu ghi có cùng số MFN, chọn M (Trộn)
- Trộn các tệp ISO với CSDL, nhưng thay thế các biểu ghi có cùng số MFN, chọn U (Cập nhật)

Số MFN đầu tiên: Nếu khai báo tham số này CDS/ISIS sẽ đánh số lại các biểu ghi nhập vào, bắt đầu từ số MFN mà bạn đã xác định. Thông thường nếu chọn phương thức tải (L), bạn phải gõ vào số 1. Trong trường hợp trộn các biểu ghi (M hoặc U), CDS/ISIS sẽ tự động đánh số bắt đầu từ số MFN cuối cùng trong CSDL tiếp nhận.

Nhãn vào có chứa MFN: Nếu ta đã ấn định MFN cho một trường trong biểu ghi ISO đầu vào thì ta phải chỉ rõ nhãn ISO của trường có chứa MFN.

Đối với **các tham số format lại** ta có thể bỏ trống nếu không cần sửa đổi cấu trúc của CSDL nhập.

B. BÀI TẬP VỀ MÔ HÌNH DỮ LIỆU

Bài tập 1

Giả sử cần thiết kế một CSDL lưu trữ các thông tin về sự sở hữu xe gắn máy của các công dân ở một thành phố. Đặt tên CSDL là QLXEMAY. Hãy xây dựng lược đồ khái niệm của CSDL QLXEMAY và mô tả bằng mô hình quan hệ.

Cho ví dụ về một thể hiện của CSDL đó.

Hướng dẫn: CSDL được xây dựng từ thực thể bao gồm hai thực thể: NGƯỜI và XE MÁY.

Thực thể NGƯỜI có các thuộc tính: họ tên (HT), giới tính (GT), năm sinh (NS), địa chỉ (ĐC), số chứng minh (SCM).

Thực thể XE MÁY có các thuộc tính: biển số (BS), nhãn hiệu (NH), dung tích (DT), loại xe (LX).

Quan hệ giữa hai thực thể đó là quan hệ SỞ HỮU. Vì một người có thể sở hữu nhiều xe nên quan hệ đó là quan hệ 1: n. Quan hệ SỞ HỮU có ba thuộc tính là số chứng minh (SCM), biển số xe (BS) và ngày đăng ký (NĐK).

CSDL QLXEMAY được xây dựng từ ba quan hệ sau:

NGƯỜI (HT, GT, NS, ĐC, SCM)

XEMÁY (BS, NH, DT, LX)

SỞHỮU (SCM, BS, NĐK)

Dưới đây là một thể hiện của CSDL nói trên:

NGƯỜI

HT	GT	NS	ĐC	SCM
Trần Quang Minh	Nam	1972	43 Hàng Ngang	H-112876
Vũ Thuý Hiền	Nữ	1968	107 Bà Triệu	B-109872
Lê Hoài Nam	Nam	1974	24 Quán Sứ	T-234518
Trịnh Văn An	Nam	1965	18 Ngô Văn Sở	H-140054
Nguyễn Hùng	Nam	1969	4 Hàng Đào	A-247715
Ngô Hồng Vân	Nữ	1970	14 Hàng Khoai	A-431087
Trần Thu Hà	Nữ	1973	55 Nguyễn Trãi	N-510186

XEMÁY

BS	NH	DT	LX	NDK
F2-5460	Honda	50	Nữ	23-2-1995
F2-5461	Suzuki	100	Nữ	23-2-1995
F2-5462	Honda	100	Nam	24-2-1995
F2-5463	Honda	70	Nữ	24-2-1995
F2-5464	Vespa	125	Nữ	24-2-1995
F2-5465	Suzuki	100	Nữ	25-2-1995
F2-5466	Honda	70	Nữ	25-2-1995

SỞHỮU

SCM	BS	NDK
H-112876	F2-5460	23-2-1995
B-109872	F2-5461	23-2-1995
T-234518	F2-5462	24-2-1995
H-140054	F2-5463	24-2-1995
A-247715	F2-5464	24-2-1995
A-431087	F2-5465	25-2-1995
N-510186	F2-5466	25-2-1995

Bài tập 2

Dựa vào các thông tin cho trong CSDL QLXEMAY ở bài tập 1 viết biểu thức quan hệ trả lời các câu hỏi sau:

1- Cho biết biển số của các xe Honda đăng ký trước ngày 25-2-1995.

2- Cho biết biển số của các xe Honda nữ đăng ký trước ngày 25-2-1995.

3- Cho biết tên chủ sở hữu của các xe Honda đăng ký trước ngày 25-2-1995.

Tối ưu hoá các câu hỏi đó nếu thấy cần thiết và tìm quan hệ kết quả.

Bài tập 3

Giả sử một CSDL có tên là CÔNGTY lưu trữ thông tin về công việc của các nhân viên trong một công ty kinh doanh được xây dựng từ quan hệ sau:

NHÂN VIÊN (MÃNV, HỌTÊN, ĐỊACHỈ)

CÔNGVIỆC(MÃCV, TÊNCV, ĐƠNVI)

PHÂNCÔNG (MÃNV, MÃCV, NGÀYKD, NGÀYKT)

trong đó:

MÃNV = Mã nhân viên

HỌTÊN = Họ và tên nhân viên

ĐỊACHỈ = Địa chỉ của nhân viên

MÃCV = Mã công việc

TÊN CV = Tên công việc

ĐƠN VỊ = Đơn vị quản lý công việc

NGÀY KĐ = Ngày khởi đầu công việc

NGÀY KT = Ngày kết thúc công việc

Dưới đây là một thể hiện của CSDL đó:

NHÂN VIÊN

MÃ NV	HỌ TÊN	ĐỊA CHỈ
A15	Lê Huy	12 Hàng Đào
B70	Vũ Hùng	54 Hàng Bông
C34	Trần Thu Hà	28 Hàng Gai
...	...	...

CÔNG VIỆC

MÃ CV	TÊN CV	ĐƠN VỊ
F5	Bán hàng	Kinh doanh
D7	Bảo vệ	Kinh doanh
S25X	Thư ký	Tổ chức
S26Z	Thủ quỹ	Tài vụ
...	...	...

PHÂN CÔNG

MÃ NV	MÃ CV	NGÀY KĐ	NGÀY KT
A15	F5	1-9-92	30-9-95
A15	D7	1-10-93	30-9-95
B70	F5	1-10-93	30-9-95
C34	S25X	1-3-92	30-4-94
C34	S26Z	1-5-93	30-4-94
...	...	...	...

Dựa vào các thông tin cho trong CSDL CÔNGTY hãy viết biểu thức quan hệ trả lời các câu hỏi sau và tìm quan hệ kết quả:

- 1- Hãy liệt kê tên các công việc ở bộ phận kinh doanh.
- 2- Hãy cho biết tên các nhân viên làm công việc bán hàng kết thúc công việc vào ngày 30-9-1995.

Bài tập 4

Cơ sở dữ liệu quản lý công tác giảng dạy của các giảng viên trong một trường đại học được xây dựng từ 3 quan hệ sau:

GIẢNGVIÊN (MÃGV, HỌTÊN, KHOA)

MÔN HỌC (MÃMH, TÊN MÔN HỌC)

PHÂN CÔNG (MÃGV, MÃMH, LỚP, NGÀYKD, NGÀYKT)

trong đó:

MÃGV = Mã giảng viên

MÃMH = Mã môn học

NGÀYKD = Ngày khởi đầu

NGÀYKT = Ngày kết thúc

Viết các biểu thức quan hệ trả lời cho các câu hỏi sau:

- 1- Cho biết tên các môn học kết thúc trước ngày 25-3-1998.
- 2- Cho biết họ tên giảng viên và môn học bắt đầu dạy sau ngày 25-3-1998.
- 3- Cho biết họ tên và đơn vị công tác của giảng viên đang dạy cho lớp TV23A.

Bài tập 5

Cơ sở dữ liệu quản lý đơn đặt mua tài liệu của một đơn vị thông tin được xây dựng từ 3 quan hệ sau:

KHÁCH HÀNG(MÃKH, HỌTÊN, NĂM SINH, ĐỊA CHỈ, CỞ QUAN)

TÀI LIỆU(MÃTL, TÊN TL, DẠNG TL, NXB, NĂM)

ĐƠN ĐẶT(MÃKH, MÃTL, ĐƠN GIÁ, SỐ LƯỢNG, THÀNH TIỀN, NGÀY ĐẶT, NGÀY NHẬN)

trong đó:

MÃKH = Mã khách hàng

MÃTL = Mã tài liệu

DẠNG TL = Dạng tài liệu

NXB = Nhà xuất bản

NĂM = Năm xuất bản

Viết các biểu thức quan hệ trả lời cho các câu hỏi sau

1- Cho biết tên tài liệu và số lượng tài liệu phải trao cho khách hàng vào ngày 25-2-1995

2- Cho biết họ tên của các khách hàng thuộc Bộ Giáo dục đặt mua cuốn sách Nhật ký trong tù.

3- Cho biết họ tên khách hàng đặt mua tài liệu dạng "sách" sau ngày 25-3-1995.

Bài tập 6

Cơ sở dữ liệu quản lý việc học tập các chuyên đề của sinh viên được xây dựng từ 3 quan hệ sau:

SINHVÊN (MÃSV, HỌTÊN, NĂM SINH, LỚP, KHOA)
CHUYÊN ĐỀ (MÃCD, TÊNCD, GIẢNGVIÊN, PHÒNGHỌC,
NGÀY)

ĐĂNG KÝ HỌC (MÃSV, MÃCD)

trong đó:

MÃSV = Mã sinh viên

MÃCD = Mã chuyên đề

TÊNCD = Tên chuyên đề

Viết các biểu thức quan hệ trả lời cho các câu hỏi sau:

1- Cho biết họ tên các sinh viên lớp TV4 đăng ký học chuyên đề vào ngày 20-8-1998.

2- Cho biết họ tên sinh viên ở khoa Thông tin – thư viện dự chuyên đề do giáo sư Vũ Bình giảng.

3- Cho biết họ tên và lớp của những sinh viên dự chuyên đề " Foxpro ứng dụng" trình bày vào ngày 25-8-1998.

Bài tập 7

Cơ sở dữ liệu quản lý các đề tài nghiên cứu khoa học của sinh viên được xây dựng từ 3 quan hệ sau:

SINHVÊN (MÃSV, HỌTÊN, NĂM SINH, LỚP, KHOA)

ĐỀ TÀI (MÃĐT, TÊN ĐỀ TÀI, CHỦ NHIỆM, KINH PHÍ)

ĐĂNG KÝ ĐỀ TÀI (MÃSV, MÃĐT, NGÀY KD, NGÀY KT)

trong đó:

MÃSV = Mã sinh viên

MÃĐT = Mã đề tài
NGÀYKD = Ngày khởi đầu
NGÀYKT = Ngày kết thúc

Viết các biểu thức quan hệ trả lời cho các câu hỏi sau:

- 1- Cho biết họ tên và lớp của các sinh viên nhận đề tài có kinh phí lớn hơn 1.000.000 đồng.
- 2- Cho biết họ tên của các sinh viên khoa Thông tin –thư viện nhận đề tài do giáo sư Vũ Bình là chủ nhiệm.
- 3- Cho biết họ tên sinh viên và tên đề tài phải hoàn thành trước ngày 25-3-1998.

TÀI LIỆU THAM KHẢO

Tiếng Việt

- 1- *Hệ thống lưu trữ và tìm kiếm thông tin CDS/ISIS version 2.3*. Trung tâm TTTL Khoa học và Công nghệ Quốc gia, Hà Nội, 1990. – 261 tr.
- 2- Lê Tiến Vương. *Nhập môn cơ sở dữ liệu quan hệ*. – H.: Khoa học kỹ thuật, 1994.
- 3- *Hệ quản trị CSDL từ liệu CDS/ISIS for Windows*. Trung tâm TTTL Khoa học và Công nghệ Quốc gia, Hà Nội, 1999.- 95 tr.
- 4- Đoàn Phan Tân. *Tin học trong hoạt động Thông tin - Thư viện*. H.: NXB Văn hoá - Thông tin, 1997. - 242 tr.
- 5- Đoàn Phan Tân. *Về cấu trúc của cơ sở dữ liệu thư mục*. Tạp chí Thông tin và Tư liệu, số 4 - 1993, trang 4 - 7
- 6- Đoàn Phan Tân. *Ứng dụng của logic mệnh đề trong các hệ thống tìm tự động hoá*. Tập san Thư viện, số 3 - 1993, trang 2 - 4.
- 7- Đoàn Phan Tân. *Tìm hiểu mô hình cơ sở dữ liệu quan hệ*. Tạp chí Thông tin và tư liệu, số 2 - 1995, trang 15 - 21.

8- Đoàn Phan Tân. *Hệ quản trị cơ sở dữ liệu với tệp đảo*.
Tập san Thư viện, số 4 - 1995, trang 11-15.

9- Đoàn Phan Tân. *Phương pháp thiết kế khái niệm CSDL bằng mô hình thực thể quan hệ*. Tạp chí Thông tin & Tư liệu, số 3 - 2000, trang 12-15.

10- Vũ Văn Sơn. *Giáo trình Biên mục mô tả*. - H.: NXB Đại học Quốc gia Hà Nội, 2000. - 284 tr.

Tiếng Anh và Pháp

11- Claire Guinchat et Michel Menou. *Introduction aux sciences et techniques de l'information et de la documentation*. Les Presses de l'Unesco, 1981. - 402 p.

12- Maier D. *The Theory of Relational Databases*. Computer Science Press, Rockville. Md, 1983.- 603 p.

13 - Peter Zorkoczy. *Information technology - An introduction*, Pitman, London, 1985.- 163 p.

14- Jacques Chaumier. *Systèmes d'information - Marché et Technologies*. Entreprise moderne d'édition, Paris, 1986.- 152 p

15- André Dewese. *Informatique documentaire*. Masson, Paris, 1985.- 213 p.

16- *Prise en main de CDS/ISIS version 3.0*. Organisation des Nations Unies pour l'éducation, la science et la culture, Paris, 1992.- 77 p.

17- *Mini-micro CDS/ISIS Reference manual (Version 2.3)* Unesco, March 1989. - 285 p.

18- *CDS/ISIS for Windows.- Reference Manual (Version 1.0)*. Unesco, November 1997. - 128 p.

19- *Understanding MARC Bibliographic: Machine Readable Cataloguing* / published by the Cataloguing Distribution Service, Library of Congress, in collaboration with The Follett Software Co. in Betty Furrie, 2000.- 51p.

20- Mary Mortimer. *Descriptive cataloguing*. Canberra, DocMatrix Pty Ltd, 1999. – 239 p.

21- Carlo Batini, Stéfano Ceri, Shamkent B. Navathe. *Conceptional Database Design.-* Te Benjamin/Commings Publishing Company, Inc.1994. - 465 p.

22- *Manuel UNIMARC*: version francais. - Ed. Saur, 1991. - 414 p.

Chịu trách nhiệm xuất bản:

Giám đốc: NGUYỄN VĂN THỎA

Tổng biên tập: NGUYỄN THIÊN GIÁP

Biên tập và sửa bản in: LAN HƯƠNG

Trình bày bìa: ĐOÀN LINH CHI

TIN HỌC TRONG HOẠT ĐỘNG THÔNG TIN - THƯ VIỆN

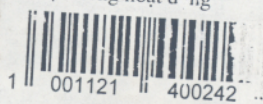
Mã số: 01.211.DH2001 - 1440.2001.

In 1000 cuốn, tại Nhà in Giao thông Vận tải.

Số xuất bản: 5/1440/CXB. Số trích ngang: 369/KH/XB.

In xong và nộp lưu chiểu Quý IV năm 2001.

tin học trong hoạt động



1 001121 400242 ..

25.000 VND

25 000 đ