

HỆ THỐNG CÁC CƠ SỞ DỮ LIỆU TẠI VIỆN THÔNG TIN KHOA HỌC XÃ HỘI

Ks. NGÔ THẾ LONG

Trưởng phòng Tin học hóa
Viện Thông tin Khoa học Xã hội

Tự động hóa hoạt động thông tin - thư viện là một tiến trình tất yếu trong việc xây dựng và phát triển từ thư viện truyền thống tới thư viện hiện đại, cho nên từ những năm đầu thành lập Viện Thông tin Khoa học xã hội (5/1975), công tác tự động hoá các hoạt động thông tin thư viện đã được Viện quan tâm đặc biệt. Từ năm 1978, thông qua Viện Thông tin Khoa học kỹ thuật trung ương (nay là Trung tâm Thông tin tư liệu Khoa học công nghệ quốc gia), Viện Thông tin Khoa học xã hội đã có đường kết nối với Hệ thống Thông tin Khoa học xã hội thuộc các nước xã hội chủ nghĩa lúc đó (MISON) để khai thác các Cơ sở dữ liệu (CSDL)

do Viện Thông tin Khoa học xã hội Liên Xô (INION) quản lý. Tuy nhiên, do nhiều hoàn cảnh khó khăn về kinh phí, thời gian tiếp cận máy tính .. nên kết quả không được bao nhiêu. Cuối năm 1989, Viện được trang bị những máy vi tính đầu tiên. Cùng với việc đào tạo cán bộ và học hỏi công nghệ, Viện đã sử dụng có hiệu quả các máy tính vào các hoạt động của Viện như chế bản điện tử, công tác tài vụ và đặc biệt là từ năm 1991, bắt đầu xây dựng các CSDL thư mục của Viện. Có thể tóm tắt về quá trình xây dựng các CSDL thư mục của Viện như sau:

1. Về phần mềm: thời gian đầu sử dụng Chương trình CDS-ISIS trong môi trường DOS từ các version 2.0 đến 2.3. Mã ký tự tiếng Việt là VNLOAD. Từ khi UNESCO phổ biến CDS-ISIS trong môi trường Windows (khoảng năm 1998), Viện đã chuyển các CSDL từ CDS-ISIS for DOS sang CDS-ISIS for Windows (từ version 1.0 nay là 1.4). Mã tiếng Việt sử dụng theo tiêu chuẩn TCVN3



Phòng tra cứu của Thư viện Khoa học Xã hội
(26 Lý Thường Kiệt, Hà Nội)

(ABC), chưa sử dụng được mã UNICODE. Việc chuyển đổi các mã từ VNLOAD sang TCVN3 bằng phần mềm đổi mã ký tự tovn2.exe. Năm 2002, Viện xây dựng mạng LAN, các CSDL của Viện đã được đưa lên mạng phục vụ bạn đọc. Năm 2003, với kinh phí hạn hẹp của một đề tài khoa học cấp bộ, Viện đã chuyển thử nghiệm một số CSDL từ CDS-ISIS sang LIBOL 2.0 của Công ty Tinh Vân. Việc chuyển đổi các CSDL này đều tuân theo chuẩn ISO 2709.

2. Về biểu mẫu nhập tin: lúc đầu các biểu mẫu nhập tin do Phòng Tự động hoá các quá trình thông tin (nay là Phòng Tin học hoá) tự thiết kế, các mã trường được đặt tùy tiện, không theo chuẩn nào. Sau Hội nghị Thông tin thư viện lần thứ 2 của Trung tâm Khoa học xã hội và Nhân văn quốc gia (11-1996), để tiến tới xây dựng một hệ thống thông tin thư mục tự động hoá thống nhất trong toàn Trung tâm, Viện Thông tin Khoa học xã hội đã soạn thảo Bản quy định chung về các yếu tố thông tin thư mục cần đưa vào CSDL, cấu trúc các trường và cách nhập dữ liệu vào các CSDL thư mục sử dụng chương trình CDS-ISIS cho các thư viện thuộc Trung tâm. Bản quy định này được dựa chủ yếu theo các mã trường do Trung tâm Thông tin Thông tin tư liệu khoa học và công nghệ quốc gia biên soạn. Từ năm 2002, cộng đồng thư viện Việt Nam đã đồng thuận về việc xây dựng một khổ mẫu thống nhất cho các thư viện trong cả nước dựa trên khổ mẫu trao đổi MARC21 do Thư viện Quốc hội Mỹ và Thư viện Quốc gia Canada biên soạn. Năm 2005, Bộ

Khoa học công nghệ đã ban hành chính thức TCVN 7539-2005 : Thông tin và tư liệu. MARC 21 cho dữ liệu thư mục). Trên tinh thần đó Phòng Tự động hoá đã biên soạn tài liệu hướng dẫn nghiệp vụ dùng cho các thư viện thuộc Viện Khoa học xã hội Việt Nam: *Nhập biểu ghi CSDL thư mục dùng chương trình CDS-ISIS với các nhãn trường của MARC21*. Để có thể sử dụng những yếu tố đã nhập và phù hợp với những đặc điểm riêng của Viện Khoa học xã hội Việt Nam, trong tài liệu này đã bổ sung thêm một số trường mang tính chất nội bộ (nằm trong vùng X9X và 9XX) như môn loại khoa học (trường 691), ký hiệu kho cũ (891) và các trường liên quan đến người nhập, xử lý tài liệu Các CSDL thư mục đã có của Viện đã được chuyển đổi từ khổ mẫu cũ sang khổ mẫu mới theo hướng dẫn trên. Các yêu cầu của MARC21 vẫn có thể được thoả mãn phần lớn khi xuất dữ liệu theo khổ mẫu trao đổi ISO-2709 qua một file FST để chuyển đổi qua các phần mềm quản trị thư viện khác, kể cả các CSDL thư mục các tài liệu bằng tiếng Nga.

3. Về các Cơ sở dữ liệu :

a. Các mã ký hiệu sách của Thư viện Khoa học xã hội :

Hiện nay, các tài liệu của Viện Thông tin được chia theo các ký hiệu sau:

- *Đối với sách thuộc kho sách cũ của Thư viện của Học viện Viễn Đông Bác cổ (École française d'Extrême-Orient, gọi tắt là EFEO) :*

+ *Tài liệu in : Được chia theo khổ sách : 8⁰ - khổ giấy in cỡ khoảng 79cm x 109cm chia 8 , 4⁰ – chia 4, PLO –*

chia 2. Đối với sách 8⁰, nếu có ký hiệu ở đầu là P có nghĩa là sách của từng phần nhỏ (P8⁰) và chữ Q là sách quốc ngữ (Q8⁰). Từ năm 1995, Thư viện KHXH bổ sung thêm một ký hiệu kho mới : OCTO tương đương với 8⁰, QTO : 4⁰

+ Tài liệu viết tay :

* *Kho Thần tích thần sắc* : ký hiệu kho cũ là Q4⁰ 18, ký hiệu kho mới : TTTS.

* *Kho tài liệu Hương ương ước* : ký hiệu kho : HU và HUN.

* *Thống kê sắc phong* : ký hiệu kho TSHN

- *Đối với sách thuộc kho sách có từ Thư viện Khoa học Trung ương đến nay* : được chia theo ngữ tài liệu (n) và khổ sách theo chiều đứng (k). Trong đó:

+ n có các ký hiệu : v (tiếng Việt), l (chữ latin), s (ký tự krilic), h (chữ Hán), n (chữ Nhật)

+ k có các ký hiệu : b (chiều đứng của cuốn sách nhỏ hơn hoặc bằng 21 cm), v (lớn hơn 21 cm và nhỏ hơn hoặc bằng 30 cm), t (lớn hơn 30 cm).

Ví dụ : Vb : sách Việt bé, Lv : sách chữ Latinh cỡ vừa.

- *Đối với tên các báo, tạp chí* : được chia theo ngữ tài liệu (n) và hình loại Báo hoặc tạp chí (T).

Trong đó :

+ n có các ký hiệu sau : v (tiếng Việt), l (chữ latin), s (ký tự Krilic), h (chữ Hán), n (chữ Nhật)

+ T có các ký hiệu : B dùng cho Báo, C dùng cho Tạp chí.

Ngoài ra, Tên Báo và Tạp chí được nhập về Thư viện sau khi tách riêng là Thư viện Khoa học xã hội (1966) sẽ thêm chữ x.

Ví dụ : BH : Báo chữ Hán, Cvx : Tạp chí tiếng Việt được nhập về khi thành lập Thư viện Khoa học xã hội.

- *Đối với các sách biếu tặng* : có các ký hiệu

+ NVT : Sách của gia đình cụ Nguyễn Văn Tố tặng Thư viện

+ KV : Sách của cụ Nguyễn Khắc Viện tặng Thư viện

+ UNb, UNv : sách do tổ chức UNESCO tặng

+ *Smy* : sách do Hội các Học giả Mỹ, Quĩ châu Á và một số cá nhân Mỹ biếu tặng.

- *Một số sách có ký hiệu riêng khác:*

+ Phòng Sách tra cứu : Sách được sắp xếp theo chủ đề nên ngoài ký hiệu kho theo đăng ký cá biệt, còn có ký hiệu theo phân loại BBK. Ví dụ : P : Lịch sử, Q : Kinh tế

+ Tài liệu do Viện Thông tin khoa học xã hội dịch : Vd

+ Luận án Tiến sĩ, Thạc sĩ : La

+ Sách của miền Nam Việt Nam trước năm 1975 (thuộc kho sách hạn chế) : Vvh, Vvn.

b. Các CSDL thư mục :

Tính tới tháng 12/2008, Viện Thông tin Khoa học xã hội đã có các CSDL thư mục:

+ *Các CSDL thư mục đã hoàn thành :*

Thần tích thần sắc (Mã CSDL : tts) : 13.211 biểu ghi.

CSDL về sự tích các vị thần được thờ cúng ở làng quê Việt Nam, cùng các nghi lễ, tục lệ thờ cúng của mỗi làng, kết quả của cuộc điều tra do Học viện Viễn Đông Bác cổ tiến hành vào khoảng những năm 30 thế kỷ XX, bằng các tiếng Việt, chữ Nôm, chữ Hán,

tiếng Pháp. Ký hiệu kho cũ : Q4⁰ 18,
Ký hiệu kho mới : TTTS

○ *Hương ước* (Mã CSDL : hgu) :
5.637 biểu ghi.

CSDL về các quy định, tục lệ của các thôn, làng Việt Nam như Hương khoán, hương lệ, hương tục, hương ước, khoán bạ, khoán lệ, khoán ước, phong tục, tục lệ, điều lệ, điều ước, ước lệ, văn khế v.v. Các văn bản này có thể bằng các ngôn ngữ tiếng Việt, chữ Nôm, chữ Hán, tiếng Pháp và chủ yếu được xây dựng vào thế kỷ XVIII vào đầu thế kỷ XIX. Ký hiệu kho : HU và HUN.

○ *Thống kê sắc phong* (Mã CSDL : tks) : 4.221 biểu ghi.

CSDL về bản kê các sắc phong, bia, thần tích, thần sắc, hương ước, tục lệ, quan tước, địa bạ v.v. của các địa phương (theo đơn vị tổng, xã) trong cả nước, bằng chữ Hán. Ký hiệu kho : TSHN

○ *Mục lục tên các Tạp chí* (Mã CSDL : serial) : 827 biểu ghi

Tên CSDL tên các Báo, Tạp chí hiện có tại Viện Thông tin KHXH (được nhập về từ khi thuộc Thư viện Khoa học Trung ương đến nay) gồm các Báo tạp chí tiếng Việt, Anh, Pháp, Nga, Trung Quốc Có các ký hiệu là : BH, BL, BLx, BS, BV, BVx, CH, CL, CLx, CS, CSx, CV, CVx.

+ *Các CSDL thư mục cập nhật tiếp tục* :

○ *CSDL Sách Thư viện KHXH* (Mã CSDL : sht) : 126.336 biểu ghi.

CSDL Thư mục các sách tiếng Việt, chữ latin (Anh, Pháp, Đức, Tây ban Nha ...), tiếng Nga (nhập dữ liệu theo phiên âm) gồm các sách in bằng các ngôn ngữ có dạng chữ latin, chữ quốc

ngữ của Thư viện Học viện Viễn Đông Bác cổ để lại (fond Châu Âu) với các ký hiệu 4⁰, 8⁰, P8⁰, Q8⁰, PLO, NVT (từ biểu ghi 1 đến biểu ghi 39.598) và các sách được nhập về từ khi là Thư viện Khoa học Trung ương cho đến nay, tài liệu do Viện Thông tin Khoa học xã hội dịch (bản đánh máy), sách do UNESCO cung cấp, luận án tiến sĩ, sách của Mỹ tặng và sách của miền Nam trước năm 1945, gồm các ký hiệu : Vb, Vv, Vt, Lb, Lv, Lt, La, Sb, Sv, UNb, UNv, Smy, Vd, Vvh, Vvn. Các sách do Trường Đại học Colorado tặng có các ký hiệu từ : Lb 21701 đến Lb 22754 và Lv 23401 đến Lv 28258.

○ *Sách tra cứu* (Mã CSDL : stc) :
2.140 biểu ghi.

CSDL các sách hiện đặt tại Phòng đọc tra cứu của thư viện (26 Lý Thường Kiệt), gồm nhiều ngôn ngữ khác nhau. Sách được sắp xếp theo chủ đề nên ngoài ký hiệu kho theo đăng ký cá biệt, còn có ký hiệu theo phân loại BBK.

○ *Bài Tạp chí* (Mã CSDL : btt) :
95.034 biểu ghi.

CSDL các bài tạp chí thuộc Viện Khoa học xã hội Việt Nam và các tạp chí tiếng Việt khác.

○ *CSDL ảnh* (Mã CSDL : anh) :
12.901 biểu ghi.

CSDL về các ảnh do EFEO để lại, và một số ảnh của Thông tấn xã Việt Nam .. (nhập về Thư viện trước năm 1975), có file ảnh (độ phân giải thấp) kèm theo.

○ *CSDL đường phố Hà Nội* (Mã CSDL : phn) : 638 biểu ghi

CSDL dùng để tra tìm tên các đường phố hiện nay ở Hà Nội, có thêm tên phố cũ.

o *Sách các thư viện thuộc Viện Khoa học xã hội Việt Nam từ năm 1998 đến nay* (Mã CSDL : stt) : 744.449 biểu ghi.

CSDL sách mới nhập toàn Viện Khoa học xã hội Việt Nam từ năm 1998 đến nay trên cơ sở tích hợp các dữ liệu của các thư viện thuộc Viện Khoa học xã hội Việt Nam gồm các sách tiếng Việt, Anh, Pháp, Nga, Trung Quốc (chữ Nga và Trung Quốc dùng nhan đề phiên âm) v.v..

Ngoài ra còn một số CSDL khác (cập nhật dở dang) như : Luận án thạc sĩ, tiến sĩ , Công báo, Sách tiếng Nga, các CSDL theo chuyên đề

Các yếu tố để tìm tin trong các CSDL thư mục này là : tác giả, nhan đề (có thể tìm theo từng từ), tên tạp chí (đối với CSDL Bài Tạp chí và Serial), địa danh như thôn, làng, tổng, xã, huyện, phủ, tỉnh, châu ...(đối với CSDL Thần tích thần sắc, Hương ước và Thống kê sắc phong), nơi xuất bản, nhà xuất bản, năm xuất bản, ký hiệu kho, ký hiệu kho mới (đối với CSDL Thần tích thần sắc, kho sách cũ của EFEO), từ khoá (CSDL Kho sách EFEO và kho sách Latin chưa xây dựng từ khóa, có thể tìm thông qua các từ trong nhan đề) ... Ngoài ra còn một số yếu tố tìm tin hỗ trợ như mã ngôn ngữ (theo ISO 369-2), mã nước (theo ISO 3166) ...

Cũng cần nói thêm, từ năm 2000, Thư viện không in thêm phiếu tra cứu cho các sách mới nhập nữa. Việc tra cứu hoàn toàn qua các CSDL trên mạng LAN của Viện (tại hai địa điểm : 26 Lý Thường Kiệt và 1 Liễu Giai, Hà Nội)

c. Các CSDL toàn văn :

Viện Thông tin KHXH cũng đã và đang tiến hành xây dựng một số

CSDL toàn văn dưới dạng HTML, PDF :

- o CSDL toàn văn Tạp chí Thông tin Khoa học xã hội.
- o CSDL toàn văn các Sưu tập chuyên đề (trên đĩa CD).
- o CSDL toàn văn Tài liệu Nghiên cứu.
- o CSDL toàn văn các Báo cáo, Đề tài khoa học.
- o CSDL số hóa một số sách của Thư viện EFEO để lại (300 tài liệu) do Pháp hỗ trợ.

Số hóa kho ảnh hiện lưu trữ tại Thư viện Khoa học xã hội

Viện đang tiến hành quản lý các CSDL toàn văn bằng phần mềm Green Stone (riêng CSDL sách của EFEO dùng phần mềm của Pháp).

Tạp chí Thông tin Khoa học xã hội của Viện tham gia Website *Tạp chí Việt Nam trực tuyến* (Vietnam Journal Online).

Ngoài ra còn một số CSDL trên đĩa CD được phục vụ cho bạn đọc như : CSDL Hồ Chí Minh toàn tập, CSDL Văn bản pháp luật Việt Nam, Tập san Đô thành hiếu cổ (Bulletin des Amis du vieux Huế), Tạp chí Văn Sử Địa v.v. Đặc biệt là bộ CSDL đồ sộ của Trung Quốc “Tứ khó Toàn Thư”, gồm 182 đĩa, tương đương 110 GByte.

Từ năm 2005, Viện cũng đã thiết lập một trang Web riêng với địa chỉ truy cập là : <http://www.issi.gov.vn>.

4. Những vấn đề còn tồn tại và hướng phát triển

- *Về phần mềm* : hiện nay Viện Thông tin khoa học xã hội và các thư viện thuộc Viện Khoa học xã hội Việt Nam sử dụng chủ yếu là phần mềm

CDS-ISIS để tạo lập các CSDL (Thư viện Viện Nghiên cứu Trung Quốc, Viện Nghiên cứu Đông Bắc Á có kết hợp sử dụng phần mềm MYLIB của Công ty CMC để xây dựng các CSDL bằng tiếng Trung Quốc, Nhật Bản, Viện Khoa học và phát triển bền vững vùng Nam Bộ dùng phần mềm Libol ...). Phần mềm CDS-ISIS do UNESCO phổ biến có nhiều điểm thuận lợi như đó là phần mềm mở, miễn phí, khả năng tìm các biểu ghi theo nội dung thông qua ngôn ngữ tìm tin năng động, khả năng trình bày biểu ghi hay một phần biểu ghi theo yêu cầu, sắp xếp biểu ghi theo thứ tự yêu cầu, trao đổi dữ liệu thông qua file trao đổi ISO-2709, phát triển các ứng dụng đặc thù nhờ phương tiện lập trình liên kết, có khả năng chuyển sang CSDL dạng XML ..., tuy nhiên CDS-ISIS cũng bộc lộ nhiều hạn chế như khả năng tích hợp kém, chưa hỗ trợ UNICODE, khó quản lý các CSDL toàn văn, chuyển tải lên mạng không thuận tiện Cho đến nay, phần mềm này chỉ phù hợp với các thư viện nhỏ. Trong khi đó một số phần mềm trong nước như LIBOL của Công ty Tinh Vân, ILIB của Công ty CMC ... phần nào khắc phục được những nhược điểm trên như quản lý được các loại ấn phẩm đa dạng với số lượng lớn, tuân thủ các chuẩn như ISO 2709, MARC 21, Z39.50 ..., được hỗ trợ UNICODE, có khả năng tích hợp với các thiết bị ngoại vi như mã vạch, thẻ từ, chạy trên nhiều hệ quản trị như Oracle, MS Sql, và đặc biệt là nó bao gồm các phân hệ tích hợp trong một CSDL chung thuận lợi cho công tác quản lý thư viện, cơ chế vận hành thống nhất. Tuy nhiên đây là phần

mềm đóng gói, kinh phí mua và bảo trì cho phần mềm không phải là nhỏ.

Ngày nay, với xu hướng phát triển là chuẩn hoá, hội nhập để tiến tới hợp tác và liên thư viện, việc xây dựng kho dữ liệu mở cho việc trao đổi thông tin – thư viện, quản lý và chia sẻ thông tin trên cơ sở các giải pháp công nghệ nguồn mở và các chuẩn mở trên Internet. Có nghĩa là Web chính là công nghệ của ngành thông tin thư viện trong hiện tại và tương lai. Nhu cầu trao đổi thông tin cũng không dừng lại ở chỗ chỉ trao đổi biểu ghi thư mục mà còn dưới dạng toàn văn, âm thanh, hình ảnh và hình ảnh động. Xu hướng hiện nay trên thế giới là ứng dụng công nghệ IP-based với ngôn ngữ XML (eXtensible Markup Language) với chuẩn Dublin Core và trao đổi dữ liệu bằng Metadata, sử dụng mạng Internet để xây dựng các kho dữ liệu mở bằng công nghệ nguồn mở và tra cứu dữ liệu trên cơ sở các cơ chế tìm kiếm thông tin (search engine). Hiện nay một trong những phần mềm mã nguồn mở thư viện dựa trên ngôn ngữ XML được UNESCO giới thiệu là phần mềm miễn phí Green Stone.

Do hạ tầng cơ sở thông tin còn kém, kinh phí chưa nhiều, nên Viện Thông tin KHXH tạm thời vẫn phải sử dụng phần mềm CDS-ISIS. Viện đang xây dựng trụ sở mới (tại 1, phố Liễu Giai, Quận Ba Đình Hà Nội, dự kiến hoàn thành vào đầu năm 2011) cùng với một dự án xây dựng hạ tầng thông tin hiện đại, chắc chắn thư viện sẽ được trang bị các thiết bị phù hợp và các phần mềm trong nước hoặc nước ngoài có đủ các phân hệ để quản lý các CSDL cũng như quản lý thư viện nói chung. Các



Ông Nguyễn Minh Hiệp, Giám đốc thư viện Đại học Khoa học Tự nhiên TP. HCM trao đổi nghiệp vụ thư viện với các cán bộ thư viện thuộc Viện Khoa học Xã hội Việt Nam (tháng 3/2005). Hàng đầu từ người thứ 4 từ trái sang phải : Pgs.Ts Vương Toàn, nguyên Phó Viện trưởng Viện Thông tin Khoa học Xã hội, Ông Nguyễn Như Kim, Chủ tịch Hội Thông tin khoa học Việt Nam, Pgs.Ts Hồ Sĩ Quý, Viện trưởng Viện Thông tin Khoa học Xã hội, Ông Nguyễn Minh Hiệp.

CSDL của thư viện được chuyển tải lên mạng Internet. Đồng thời nghiên cứu và triển khai việc xây dựng các CSDL dựa trên ngôn ngữ XML, chuẩn Z39.85-2001 và các trường dữ liệu của Dublin Core.

- *Về các Cơ sở dữ liệu* : tiếp tục cập nhật và hồi cố các CSDL, bổ sung một số trường và hiệu đính các CSDL, xây dựng các CSDL đặc thù riêng của Viện (ảnh, bản đồ, ảnh tư liệu ...), các thuật ngữ tìm tin cần được thống nhất và chuẩn hoá.

Tiến hành số hóa các tư liệu, trước mắt là các tư liệu cổ, tư liệu viết tay đặc thù của Viện như Hương ước, Thần tích thần sắc, Sắc phong, Bản đồ, ảnh tư liệu ... theo các chuẩn hiện nay để vừa lưu trữ bền vững và sớm xã hội hóa các dạng tư liệu này.

Xây dựng các CSDL tích hợp trong toàn Viện Khoa học xã hội Việt Nam, có sự phân công trong các thư viện

thuộc Viện Khoa học xã hội Việt Nam, tránh trùng lặp gây lãng phí.

Mua và trao đổi các CSDL số hoá, thực hiện một số chính sách bổ sung hợp lý tạo khả năng phục vụ kịp thời cho các nhu cầu của người dùng tin.

*

Ứng dụng công nghệ thông tin trong hoạt động thông tin - thư viện, mà cụ thể là sử dụng các phần mềm để xây dựng các CSDL thư mục và CSDL toàn văn nhằm quản trị và tìm kiếm dữ liệu là công cụ hữu hiệu để quản lý và kiểm soát hiệu quả hoạt động thông tin thư viện, nâng cao chất lượng phục vụ của thư viện và mở rộng năng lực cung cấp thông tin và dịch vụ cho người sử dụng.

Hà Nội
Tháng 11 - 2008