

MỘT SỐ KINH NGHIỆM TRONG VIỆC TẠO LẬP VÀ CHIA SẺ THÔNG TIN TRÊN MẠNG INTERNET BẰNG PHẦN MỀM MÃ NGUỒN MỞ GREENSTONE - GIẢI PHÁP LỰA CHỌN CHO CÁC THƯ VIỆN VỪA VÀ NHỎ HẠN CHẾ VỀ TÀI CHÍNH

Chu Tuyết Lan

Thư viện Trường Thương mại Anh Quốc

Trong môi trường INTERNET hiện nay, số hóa được coi là mục tiêu chiến lược trong việc phát triển hoạt động thông tin thư viện, bởi lẽ ngoài việc tăng cường khả năng tiếp cận thông tin cho độc giả, thực hiện một cách hiệu quả nhất việc phân phối thông tin, thì số hóa còn là một chiến lược hữu hiệu để bảo quản vốn tài liệu hiện có, đặc biệt là tài liệu quý hiếm. Để thực hiện tốt công tác số hóa, các cơ quan thông tin thư viện cần phải xây dựng lộ trình, lên kế hoạch phù hợp và đặc biệt chú trọng đến việc lựa chọn giải pháp công nghệ để đảm bảo khả năng sử dụng lâu dài các kết quả số hóa, trong đó cần quan tâm đến 3 chức năng cơ bản của phần mềm quản lý dữ liệu số là hiển thị, xử lý và tìm kiếm thông tin.

Nhiều năm làm việc trong một thư viện chuyên ngành, nguồn kinh phí được cấp hàng năm khá eo hẹp, nhưng lại có chức năng lưu giữ vốn tài liệu quý hiếm của dân tộc, nên chúng tôi đã triển khai công tác số hóa từ khá sớm, tranh thủ sự giúp đỡ cả về tài chính và kỹ thuật của các tổ chức quốc tế. Một trong những bài học kinh nghiệm thu nhận được trong thời gian học tập tại nước ngoài mà chúng tôi thấy cần chia sẻ đó là khi tiến hành bất cứ một hoạt động gì cũng cần xem xét những vấn đề sau: (1) Answering the needs; (2) Getting started; (3) Reaping the benefits. Có thể hiểu một cách nôm na là: trước hết bạn phải xác định thật rõ ràng tiềm năng, quy mô của thư viện nơi mình công tác, cần và có thể đáp ứng được những nhu cầu nào của độc giả. Từ đó lên phương án, xây dựng lộ trình và kế hoạch triển khai công việc sao cho phù hợp với hoàn cảnh và điều kiện thực tế của thư viện. Sau đó cần quan tâm giải quyết bài toán kinh tế: việc đầu tư triển khai công việc này sẽ đáp ứng những mục tiêu gì, phục vụ đối tượng nào và sẽ đem lại những lợi ích gì về kinh tế, văn hóa và khoa học xã hội trong bối cảnh đất nước còn nhiều khó khăn như hiện nay.

I/ KINH NGHIỆM TRONG LĨNH VỰC SƯU TẦM & BẢO QUẢN TÀI LIỆU

1. Công tác sưu tầm

Để đảm bảo độ chính xác cũng như chất lượng của nguồn tài liệu sưu tầm, Viện chúng tôi đã bước đầu ứng dụng những tiến bộ của khoa học công nghệ vào việc

điều tra điền dã và thu thập tư liệu ở các địa phương. Các nhóm sưu tầm đã sử dụng máy ảnh số và máy ghi âm, để ghi hình các tài liệu cần thu thập, (điều này tránh được các lỗi không nên có trong các sản phẩm sưu tầm thực hiện bằng phương pháp thủ công trước đây, chẳng hạn việc thiếu từ trong các vé đối: có câu đối có vé 9 từ, vé kia lại chỉ có 7; hay việc chữ tắc đánh chữ tộ, rất khó khắc phục khi xử lý, ...); việc ghi tiếng khi phỏng vấn các vị túc ả ho ở các địa phương về truyền thống làng nghề, lịch sử tư liệu và di tích cũng rất hữu ích. Hiệu quả của việc sử dụng máy ảnh kỹ thuật số và máy ghi âm để ghi lại hình ảnh tư liệu gốc và âm thanh về nguồn gốc tư liệu, sẽ giúp chúng ta có tư liệu toàn vẹn về cả âm thanh và hình ảnh để kết nối với CSDL thư mục bên cạnh phần phiên âm và dịch nghĩa của chúng, tiến tới xây dựng ngân hàng dữ liệu tổng hợp về vốn tài liệu điều tra điền dã đã được tiến hành trong nhiều năm qua. Điều này tạo điều kiện đảm bảo quyền lợi của người dùng tin ở mọi nơi, mọi lúc thông qua việc phát triển nguồn tin, các công cụ giúp kiểm soát, khai thác, tìm kiếm thông tin và các dịch vụ mà họ được thụ hưởng.

Để công tác sưu tầm tư liệu nói chung và tư liệu Hán ả ôm nói riêng ngày một hiệu quả hơn, cũng cần quan tâm đến việc khảo sát và thống kê chi tiết số tài liệu đã mua được ở các địa phương, số chưa mua được (tìm lý do để lên kế hoạch bổ sung trong tương lai) và đánh giá giá trị của nguồn tài liệu này trong thực tế. Vì thế về cơ bản đã tránh được tình trạng mua tài liệu trùng và mua những tài liệu ít giá trị, vì công bồi vá và bảo quản tiếp sau đó là rất lớn. Điều quan trọng nữa cần lưu ý là sau khi sưu tầm về, tư liệu cần được đánh giá, phân loại và lên danh mục các tài liệu thuộc dạng độc bản đặc biệt quý hiếm, để có quy chế bảo quản và khai thác hiệu quả và hợp lý nhất.

2) Công tác số hóa và bảo quản tư liệu Hán Nôm

Để công tác bảo quản đạt hiệu quả cao, ngoài việc năng động phát huy nội lực, cần chủ động tranh thủ khai thác các nguồn tài trợ bên ngoài bổ sung những trang thiết bị & các loại vật liệu có độ chuẩn cần thiết phục vụ cho công tác bảo quản và phục chế. Tranh thủ các điều kiện để cử cán bộ đi học tập và trao đổi kinh nghiệm với các nước có truyền thống về số hoá và bảo quản như: Mỹ, Trung Quốc, ả hạt và Đài Loan, ... Chú ý tổ chức các workshop mời chuyên gia nước ngoài đến để giảng dạy về kỹ thuật số hoá và bảo quản, nhằm trang bị những kiến thức cần thiết cho những cán bộ làm công tác này, giúp họ chủ động và tự tin hơn trong công việc mình làm.

Hiện 100% tài liệu Hán ả ôm bản gốc đã được sao chụp và phần lớn tài liệu quý hiếm đã được số hóa để phục vụ bạn đọc, bản gốc chỉ được phục vụ trong những trường hợp đặc biệt nhằm đối chiếu so sánh, để giảm thiểu đến mức tối đa độc giả làm tổn thương đến tài liệu gốc. Cần tranh thủ sự đầu tư của nhà nước và các tổ chức quốc tế để từng bước trang bị những thiết bị hiện đại cho công tác bảo quản và số hoá tài liệu.

- **Bảo quản bằng hệ thống điều hoà trung tâm**, kết hợp với việc dùng phần mềm Data-logger để kiểm soát nhiệt độ và độ ẩm trong kho. Đây là công nghệ mới cho phép kiểm soát nhiệt độ và độ ẩm một cách chuẩn xác nhất qua hệ thống máy vi tính. Điều quan trọng là phải tuân thủ chặt chẽ những thông số chuẩn về hệ thống nhiệt độ, và độ ẩm, tránh tình trạng hệ thống điều hoà trung tâm chỉ hoạt động vào những ngày làm việc trong tuần. Ắt ẽ như vậy thì việc dao động về nhiệt độ và độ ẩm trong kho sẽ rất lớn, nhất là vào những ngày nghỉ cuối tuần, và ngày lễ, lúc đó chỉ số chênh lệch về nhiệt độ và độ ẩm sẽ ở mức báo động. Điều này cần phải được quán triệt một cách triệt để, nếu không giá trị bảo quản của hệ thống điều hoà trung tâm sẽ đi ngược lại với những mong muốn trong thực tế. Các công trình nghiên cứu trong khu vực và thế giới về lĩnh vực bảo quản cũng đã chỉ rõ: tạo ra môi trường thích hợp và ổn định về nhiệt độ và độ ẩm là phương pháp bảo quản tối ưu nhất cho nguồn tài liệu. Bởi vì sự dao động quá lớn về nhiệt độ và độ ẩm là yếu tố nguy hại lớn nhất đối với các loại hình tài liệu. Sự dao động này sẽ thúc đẩy các phản ứng hoá học tăng nhanh, phân huỷ các loại vật liệu hữu cơ, làm giảm tuổi thọ của tài liệu một cách đáng kể. Mức độ của các phản ứng này gia tăng cùng với điều kiện nhiệt độ, độ ẩm và các hợp chất acid hiện có, sẽ tạo điều kiện lý tưởng cho nấm mốc và côn trùng phát triển, gây nguy hại cho tài liệu nói chung và tài liệu số nói riêng.
- **Bảo quản bằng Microfilm**, cho đến nay, Microfilm vẫn được xem là phương tiện chính yếu trong công tác bảo quản nội dung nguồn tài liệu kể cả của thư viện và các cơ quan lưu trữ, với điều kiện là quá trình sản xuất và chế độ kho tàng phải thực hiện theo đúng chuẩn quốc tế. Kết quả nghiên cứu đã chứng minh rằng: nếu được bảo quản trong điều kiện nhiệt độ và độ ẩm lý tưởng thì tuổi thọ của Microfilm có thể kéo dài tới 500 năm.
- **Bảo quản bằng phương pháp số hóa**, đây là phương pháp mới, hỗ trợ rất đắc lực cho việc khai thác và bảo quản tư liệu, được hai trường Đại học nổi tiếng của Mỹ là Yale và Cornell nghiên cứu đưa vào sử dụng từ tháng 11/92. Công nghệ này rất đáng quan tâm, vì ưu thế của nó là cùng một lúc có thể kết hợp được cả hai chức năng bảo quản và khai thác, thông qua mạng máy tính bạn đọc có thể truy cập trực tuyến những tài liệu mà họ có nhu cầu ở mọi lúc, mọi nơi, và đồng thời cho nhiều người sử dụng, giúp ích rất nhiều cho người nghiên cứu nói riêng và bạn đọc nói chung. Tại nhiều Hội thảo quốc tế, các chuyên gia hàng đầu của thế giới về công tác bảo quản đã khẳng định số hoá là công nghệ mới có nhiều tiến bộ, rất hữu ích cho việc tìm kiếm thông tin, đồng thời cũng là phương tiện bảo quản gián tiếp, giúp bạn đọc thuận tiện trong việc tra cứu, hạn chế đến mức tối đa việc sử dụng trực tiếp bản gốc, do vậy giúp kéo dài tuổi thọ của nguồn tài liệu gốc.

ả gay từ đầu những năm 1998, kho tư liệu Hán ả om đã được số hoá tuân thủ nguyên tắc: *ưu tiên những tài liệu đặc biệt có giá trị hoặc ở trong tình trạng rách nát mà có nhu cầu khai thác cao.*

- Một CSDL đơn lẻ đã được xây dựng với các hình ảnh số hoá nguồn tài liệu Hán ả om có kết nối thông tin tiếng Việt và phần mềm đọc duyệt tài liệu. Tất cả đều được ghi trên đĩa CD-ROM. Tuỳ theo độ dày của sách mà trên một đĩa CD-ROM có thể có 1 hoặc nhiều cuốn sách, song dung lượng không vượt quá 500 Mb dữ liệu. Các ảnh số được tạo bởi máy Scanner loại chất lượng cao, độ phân giải 300 DPI, độ sâu màu 24 bit, dạng JPG (có độ nén khá cao), kích cỡ ảnh trung bình 20 cm x 30 cm, dung lượng trung bình 500 Kb.

- CSDL tập trung được thực hiện từ năm 2004 trở lại đây, CSDL này tích hợp toàn bộ thông tin hiện có trong bộ Di sản Hán ả om Việt ả am thư mục đề yếu với các tài liệu Hán ả om đã được số hoá chứa trong các đĩa CD. Độc giả có thể trực tiếp tìm tài liệu trên máy tính mà không cần phải tra cứu thủ công như trước đây, và cũng có thể đọc những sách mình cần (những tài liệu đã số hoá) trên mạng LA ả .

Hiện chúng tôi đang dùng hệ điều hành quản lý mạng **Window Server 2003**, (hệ điều hành của Microsoft có khả năng quản lý mạng với độ bảo mật cao), và hệ quản trị CSDL **MS SQL Server 2000**. Phần mềm này có khả năng quản lý dữ liệu lớn, đặc biệt với dữ liệu ảnh số, phù hợp với CSDL Hán ả om đã số hoá, việc quản lý người dùng và thông tin khá hoàn thiện, tính bảo mật cao, dễ sử dụng, có khả năng mở rộng và kết xuất với mạng I ả TER ả ET. Để khai thác các dữ liệu Hán ả om đã số hoá, chúng tôi phối hợp cùng công ty tin học eK xây dựng chương trình Quản lý đọc duyệt thư tịch Hán ả om. Phần mềm này hoạt động trên nền hệ quản trị CSDL của hãng Microsoft SQL Server 2000, giúp bạn đọc có thể đọc trực tiếp các tài liệu Hán ả om và tìm thông tin trên máy tính cá nhân có nối mạng LA ả của viện. Cũng giống như các phần mềm khác, phần mềm này giúp người đọc tìm kiếm thông tin theo nhiều tiêu chí khác nhau, như: nhan đề sách, tên tác giả, ký hiệu sách và các từ khoá địa danh, nhân vật, niên đại và chủ đề. ả goài ra độc giả còn có thể tìm đến trang bản gốc của tài liệu thông qua việc tìm từ hoặc cụm từ bất kỳ nào từ bản dịch của một hoặc toàn bộ số sách có trong CSDL

- **Số hoá kho thác bản văn khắc Hán Nôm**, trong mối quan hệ hợp tác với các tổ chức khoa học nước ngoài, Viện đã phối hợp với Học viện Viễn Đông bác cổ (E.F.E.O.) và trường Cao đẳng thực hành Pháp thực hiện chụp ảnh kỹ thuật số các thác bản văn khắc, ngoài việc ghi lưu trên ổ cứng, còn sao lưu trên các đĩa CD-ROM chất lượng cao, nhằm bảo quản lâu dài và khai thác có hiệu quả kho tư liệu văn khắc vô giá này của Việt ả am. Do các thác bản văn khắc có kích cỡ rất đa dạng và có những thác bản thuộc diện ngoại cỡ (từ 24 đến 32 các tờ giấy dó ghép lại), nên không thể dùng máy quét scanner thông thường được. Giải pháp công nghệ được lựa chọn ở đây là dùng máy ảnh kỹ thuật số ả IKO ả Cool Pix 5.000, và một hệ thống các thiết bị phụ kèm theo như hệ thống hút gió (tự

thiết kế theo cách thủ công) để làm phẳng các thác bản trước khi chụp, một Tivi nối trực tiếp với máy ảnh số, giúp người chụp có thể tự quan sát và điều chỉnh độ nét của ảnh, mục đích tạo ra những hình ảnh có chất lượng cao nhất như có thể. Cho đến thời điểm này công trình phối hợp ba bên (giữa Viện ả nghiên cứu Hán ả ôm với E.F.E.O. và trường Cao đẳng thực hành Pháp) đã chụp xong kho thác bản văn khắc Hán ả ôm với trên 56.000 đơn vị thác bản.

Để phục vụ cho việc khai thác nguồn tin số hoá kho thác bản văn khắc, chúng tôi đã và đang xây dựng CSDL các thác bản văn khắc Hán ả ôm, cung cấp cho bạn đọc những nội dung cơ bản về các đơn vị văn khắc hiện đang lưu giữ tại Viện. Hy vọng ngoài việc tra cứu thông tin trong bộ Thư mục Văn khắc Hán ả ôm Việt ả am, bạn đọc có thể khai thác kho thác bản văn khắc Hán ả ôm này thông qua mạng Iả TERả ET hoặc mạng LAả của Viện, chứ không đọc trực tiếp trên bản gốc, đây cũng là cách bảo quản gián tiếp đối với nguồn tài liệu quý hiếm này.

➤ **Xây dựng CSDL toàn văn Tạp chí Hán Nôm trên mạng INTERNET:** Để quảng bá các thành quả nghiên cứu Hán ả ôm trên mạng Internet, phục vụ ngày một tốt hơn nhu cầu nghiên cứu và khai thác của giới nghiên cứu Hán ả ôm và những người có quan tâm đến văn hóa Việt ả am ở trong nước, khu vực và trên thế giới. Trong chương trình hợp tác giữa Viện ả nghiên cứu Hán ả ôm và Viện Harvard-Yenching (Hoa Kỳ), Viện đã được tài trợ một khoản kinh phí để xây dựng Website của Viện và CSDL toàn văn Tạp chí Hán ả ôm trực tuyến. Thư viện Viện ả nghiên cứu Hán ả ôm đã phối hợp với bộ phận trị sự của Tạp chí Hán ả ôm trong việc xây dựng CSDL toàn văn Tạp chí Hán ả ôm trực tuyến. Với CSDL này, người đọc có thể tiếp xúc trực tiếp với Tạp chí Hán ả ôm toàn văn (từ lúc bắt đầu xuất bản năm 1984 đến nay) ở mọi lúc mọi nơi, thông qua các giao diện về năm xuất bản tạp chí, số của tạp chí, tổng mục lục tên tác giả, tổng mục lục nhan đề bài viết bằng cả tiếng Việt và tiếng Anh. Đây là một trong những tạp chí nghiên cứu đầu tiên đưa được đầy đủ nội dung toàn văn trên mạng Iả TERả ET, thành công này đã được ghi nhận là *một trong những sự kiện nhất của báo chí Việt Nam năm 2006*.

3) Kinh nghiệm bảo quản, lưu trữ dữ liệu phục vụ bạn đọc: Qua kinh nghiệm thực tế, chúng tôi thấy rằng không nên dùng một giải pháp chung cho cả việc bảo quản, lưu trữ dữ liệu và dữ liệu phục vụ bạn đọc. Vì mỗi công việc đều có những yêu cầu và khả năng đáp ứng khác nhau. Do vậy cần có 2 giải pháp riêng biệt:

a. *Bảo quản, lưu trữ dữ liệu:* Dữ liệu cần được lưu trữ trên 3 phương tiện và để ở các vị trí khác nhau.

1a. Lưu trên các đĩa CD-ROM hoặc DVD chất lượng cao, cứ sau 2 đến 3 năm lại sao lưu thêm một bản khác.

2a. Lưu trên một ổ cứng riêng biệt có dung lượng lớn, để ở nơi an toàn có cập nhật dữ liệu hàng tháng.

3a. Lưu trong ổ cứng của máy chủ (dung lượng lớn, chia làm nhiều ổ riêng biệt), dữ liệu sẽ chiếm một khoảng trống nhất định trong ổ cứng của máy chủ, việc này sẽ kém an toàn khi máy chủ đã có tuổi thọ cao, nhưng lại thuận tiện cho việc cập nhật dữ liệu trên mạng.

b. *Dữ liệu phục vụ bạn đọc trên mạng*: Dữ liệu trong hệ quản trị CSDL SQL Server đã được mã hoá và truyền tải trên mạng. Quá trình mã hoá, tìm kiếm và hiển thị dữ liệu là một quá trình phức tạp và tiêu tốn thời gian. Do vậy để dữ liệu được tìm kiếm, hiển thị nhanh, dữ liệu cần được tối ưu hoá. Đối với dữ liệu Hán ả ôm số hoá có dung lượng lớn, số lượng nhiều. Giải pháp của Viện là thu nhỏ ảnh sao cho dung lượng ảnh được mã hoá và truyền tải trên mạng từ khoảng 100 – 200 kb. Dữ liệu trong CSDL SQL Server thực chất đã được lưu trữ đúng, Card điều khiển luôn tạo một khoảng trống trên tất cả các đĩa cứng để lưu dữ liệu vào khu vực đó. ả người quản trị mạng còn phải thường xuyên backup dữ liệu sau mỗi lần cập nhật hoặc sửa chữa CSDL. Việc backup từ SQL Server rất thuận tiện, có thể backup toàn bộ hoặc backup từng phần. Do vậy dữ liệu ở đây thực chất cũng được nhân 3:

1b. Dữ liệu được mã hoá trong SQL Server.

2b. Card điều khiển tự động lưu trữ dữ liệu trên nhiều vùng của ổ cứng.

3b. Dữ liệu backup được copy sang một phương tiện lưu trữ khác cất giữ bên ngoài máy chủ (lưu vào ổ cứng có dung lượng lớn).

Trong nhiều năm qua chúng tôi đã và đang làm công việc tạo tiền đề cho việc xây dựng một thư viện số các tài liệu Hán ả ôm để phục vụ giới nghiên cứu Hán ả ôm ở cả trong và ngoài nước. Hy vọng trong một tương lai không xa, sẽ phần đầu xây dựng thành công Thư viện số các tài liệu Hán ả ôm, với những nhiệm vụ cụ thể sau đây:

1) Tổ chức mục lục liên hợp bao quát đầy đủ các tài liệu Hán ả ôm theo một format chuẩn thống nhất, giới thiệu toàn diện các tài liệu Hán ả ôm ở cả trong và ngoài nước;

2) Thường xuyên cập nhật và bảo trì mạng lưới thông tin thư mục các tài liệu Hán ả ôm, đảm bảo việc tìm kiếm dễ dàng qua các yếu tố tác giả, nhan đề, từ khoá và chủ đề cho người dùng tin trên toàn thế giới;

3) Từng bước phổ biến các kết quả và công trình nghiên cứu, tạo điều kiện thuận lợi cho việc trao đổi thông tin về lĩnh vực Hán ả ôm và các ngành khoa học có liên quan;

4) Duy trì việc giới thiệu toàn văn Tạp chí Hán ả ôm trên mạng Iả TERả ET, chú ý cập nhật nội dung mới và bổ sung thêm phần tóm tắt nội dung các bài viết bằng tiếng Anh;

- 5) Cung cấp tài liệu Hán ả om nguyên ngữ và bản dịch tiếng Việt đảm bảo thoả mãn nhu cầu giảng dạy và học tập Hán ả om trên mạng Iả TERẢ ET.
- 6) Từng bước cải tiến trang Web, chú ý cả hai mặt nội dung thông tin và công nghệ, đáp ứng nhu cầu thông tin ngày một cao của giới nghiên cứu Hán ả om ở cả trong nước và nước ngoài. Các CSDL thư mục và tài liệu Hán ả om số hoá sẽ là nội dung quan trọng của trang Web để phổ biến thông tin Hán ả om tới người dùng tin ở khắp mọi nơi trên thế giới.

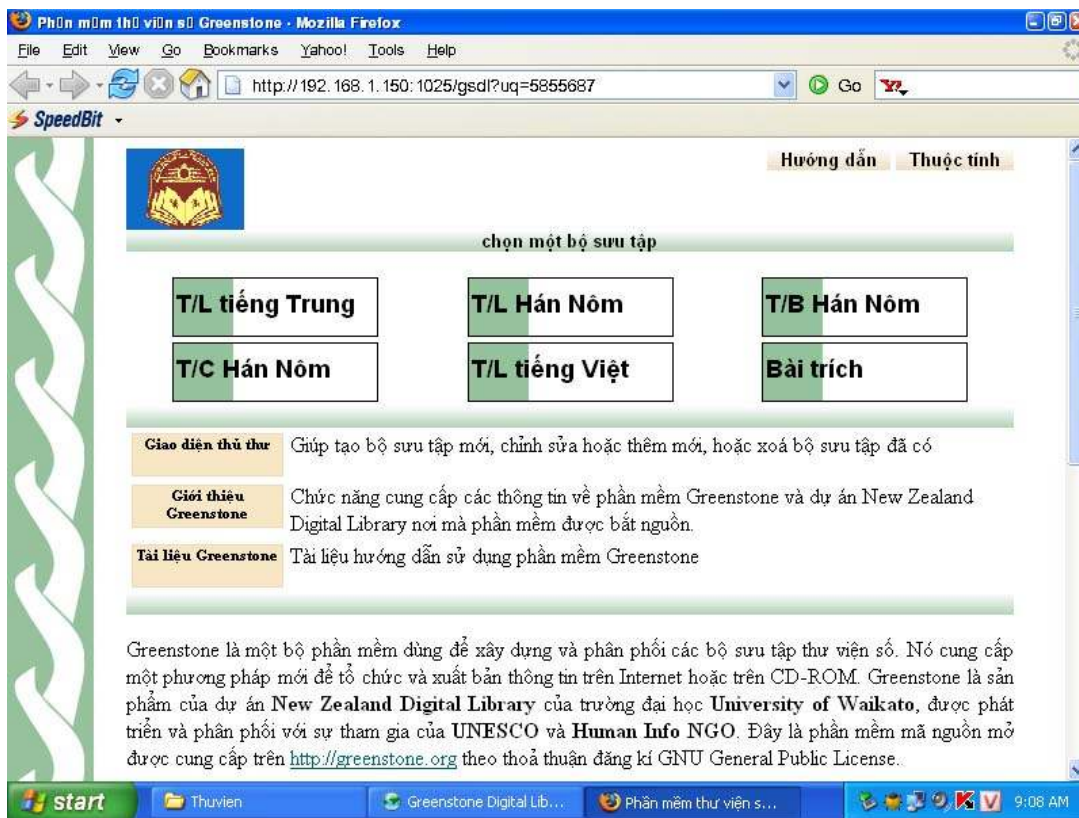
II/ GIẢI PHÁP LỰA CHỌN PHẦN MỀM VÀ XÂY DỰNG CHƯƠNG TRÌNH CHUYÊN ĐỔI DỮ LIỆU

Do kinh phí hạn hẹp, nên phần lớn các CSDL mà chúng tôi thực hiện đều được xây dựng dựa trên phần mềm ISIS với chữ Việt là font ABC, chữ Hán là font TwinBridge đã phần nào đáp ứng được yêu cầu đảm bảo thông tin và tạo các sản phẩm thông tin hiện có. Tuy nhiên để quảng bá rộng rãi vốn tài liệu nói trên cho các nhà nghiên cứu trong và ngoài nước sử dụng, thì cần phải tìm giải pháp để đưa lên mạng Iả TERẢ ET. Hiện nay phần mềm ISIS đã có một số bản dành cho CSDL trên mạng nhưng không hỗ trợ các font chữ Việt, kể cả Unicode còn các phần mềm khác đang được thương mại hóa ở Việt ả am thì còn khá nhiều lỗi và quá đắt đối với một thư viện chuyên ngành với vốn tài liệu không lớn như Thư viện Viện ả ghiên cứu Hán ả om, đây là chưa kể để chạy các phần mềm này Viện phải trang bị máy chủ đủ mạnh với khoản kinh phí không nhỏ. Với số biểu ghi như của Viện chúng tôi và giá các phần mềm đang chào hàng (30 ngàn USD) thì tính ra chi phí cho việc quản lý một biểu ghi là quá cao, không tiết kiệm, hơn nữa điều kiện tài chính của một thư viện chuyên ngành như chúng tôi không thể đáp ứng được. Để giải quyết mâu thuẫn giữa nhu cầu công việc và nguồn kinh phí được cấp, chúng tôi đã quyết định sử dụng **phần mềm mã nguồn mở Greenstone**.

Đúng như lời tự giới thiệu “Greenstone là một bộ phần mềm dùng để xây dựng và phân phối các bộ sưu tập thư viện số. Phần mềm này cung cấp một phương pháp mới để tổ chức và xuất bản thông tin trên Iả TERẢ ET hoặc trên CD-ROM. Greenstone là sản phẩm của dự án ả ew Zealand Digital Library của trường Đại học University of Waikato, được phát triển và phân phối với sự tham gia của Uả ESCO và Human Info ả GO. Đây là phần mềm mã nguồn mở được cung cấp trên <http://greenstone.org> theo thoả thuận đăng kí Gả U General Public License. Mục tiêu của phần mềm là cung cấp cho người sử dụng, đặc biệt là các trường đại học, các thư viện hoặc trong các trụ sở công cộng xây dựng các thư viện số riêng của họ. Các thư viện số làm thay đổi nhanh chóng cách thức thông tin được thu thập và phổ biến trong các thành viên của Uả ESCO và các cơ quan trong lĩnh vực giáo dục, khoa học và văn hóa trên toàn thế giới, đặc biệt là ở các quốc gia đang phát triển.”

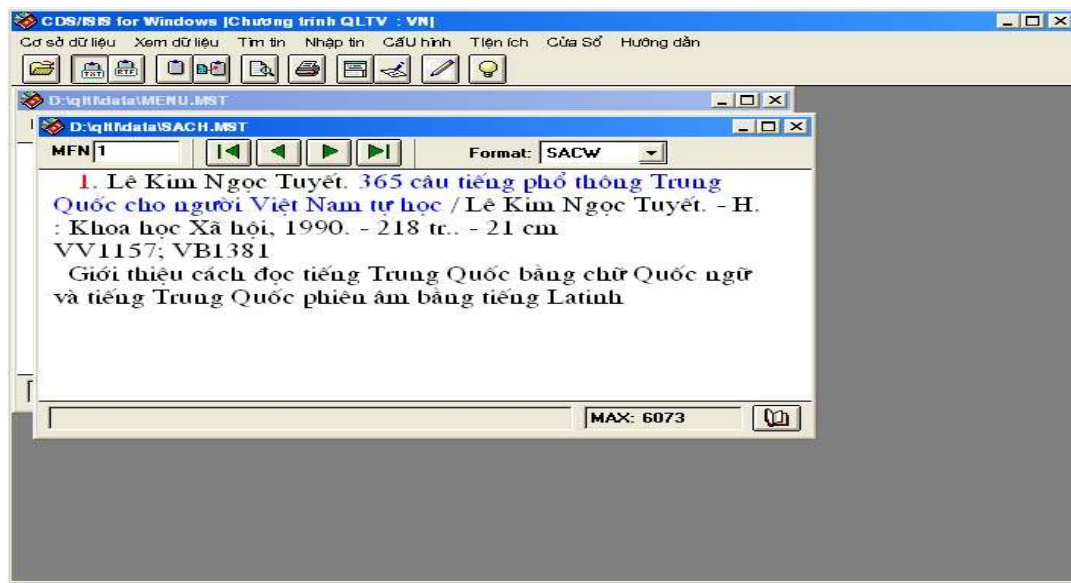
Cũng như phần mềm ISIS, phần mềm Greenstone là phần mềm mã nguồn mở nên việc sử dụng không lo vi phạm bản quyền. Tuy nhiên việc chuyển đổi biểu ghi từ CSDL ISIS sang biểu ghi CSDL Greenstone là không đơn giản vì hai CSDL này khác nhau cả về cấu trúc và font chữ: về cơ bản, CSDL ISIS dùng cấu trúc MARC và font chữ ABC cho tiếng Việt và TwinBridge cho Hán ả ôm còn CSDL Greenstone dùng cấu trúc Dublin Core và font chữ Unicode. ả hư vậy nếu muốn chuyển đổi thì trước hết phải chuyển đổi cả cấu trúc và font chữ, sau đó mới tính tới việc nhập vào biểu ghi Greenstone.

Để làm được điều này, chúng tôi đã xây dựng một chương trình chuyển đổi và đã chuyển đổi thành công. Chương trình được thiết kế theo chế độ tự động, nên có thể chuyển đổi dữ liệu 24/24 giờ mà không cần có người trực. Điều này rất hữu ích, đặc biệt là với các CSDL lớn vì có thể sử dụng giờ nghỉ để chuyển đổi dữ liệu. Quy trình chuyển đổi được tiến hành như sau: dùng ngôn ngữ format của ISIS để đổi cấu trúc biểu ghi sau đó chương trình sẽ chuyển đổi font chữ và tự động nhập biểu ghi vào CSDL Greenstone. Đối với biểu ghi tiếng Việt, công việc đơn giản hơn vì chỉ phải chuyển đổi 1 lần font chữ từ ABC sang Unicode. Còn đối với biểu ghi chữ Hán ả ôm và Trung Quốc thì phức tạp hơn nhiều, vì phải tách riêng phần chữ Việt và phần chữ Hán ả ôm hoặc chữ Trung Quốc, sau đó đổi font chữ Việt từ ABC sang Unicode, tiếp đó đổi chữ Hán ả ôm hoặc chữ Trung Quốc từ TwinBridge sang Unicode, cuối cùng ghép hai phần lại với nhau và nhập vào biểu ghi Greenstone. Hiện tại các CSDL thư mục ISIS do chúng tôi thực hiện, đã được chuyển sang CSDL Greenstone để đưa lên mạng Iả TERả ET như sau:



Dưới đây là ví dụ về biểu ghi đã được chuyển đổi:

Biểu ghi ISIS:



Xin xem chi tiết một biểu ghi::



Trong xu thế phát triển hiện nay của hoạt động thông tin – thư viện, việc đưa dữ liệu lên IẢ TERẢ ET để chia sẻ thông tin là việc làm cần thiết và hữu ích. Đa số các thư viện nhỏ hiện đang dùng phần mềm ISIS để quản trị dữ liệu, nhưng nếu muốn đưa dữ liệu lên IẢ TERẢ ET mà dùng wwisis hoặc webisis đều vướng vấn đề font chữ vì ISIS và các dị bản của nó chưa hỗ trợ Unicode, nếu mua các phần mềm tích hợp hiện bán trên thị trường thì rất đắt gây lãng phí đối với 1 thư viện có số biểu ghi không lớn. Theo kinh nghiệm của chúng tôi, thì sử dụng phần mềm Greenstone là một giải pháp tốt nhưng do sự khác nhau của hai CSDL nên không thể dùng chung được, vậy cần phải chuyển đổi. *Giải pháp xây dựng chương trình chuyển đổi biểu ghi là giải pháp tối ưu*, vì nếu không việc phải nhập lại biểu ghi sẽ tốn rất nhiều công sức và khó đảm bảo chính xác. Sử dụng Greenstone còn có lợi ích khác nữa là nếu muốn xây dựng CSDL toàn văn, ta chỉ việc tích hợp phần toàn văn vào biểu ghi của CSDL nêu trên. Dữ liệu toàn văn có thể ở dạng HTML hoặc tệp PDF.

Dưới đây là ví dụ về 1 biểu ghi có liên kết toàn văn HTML:

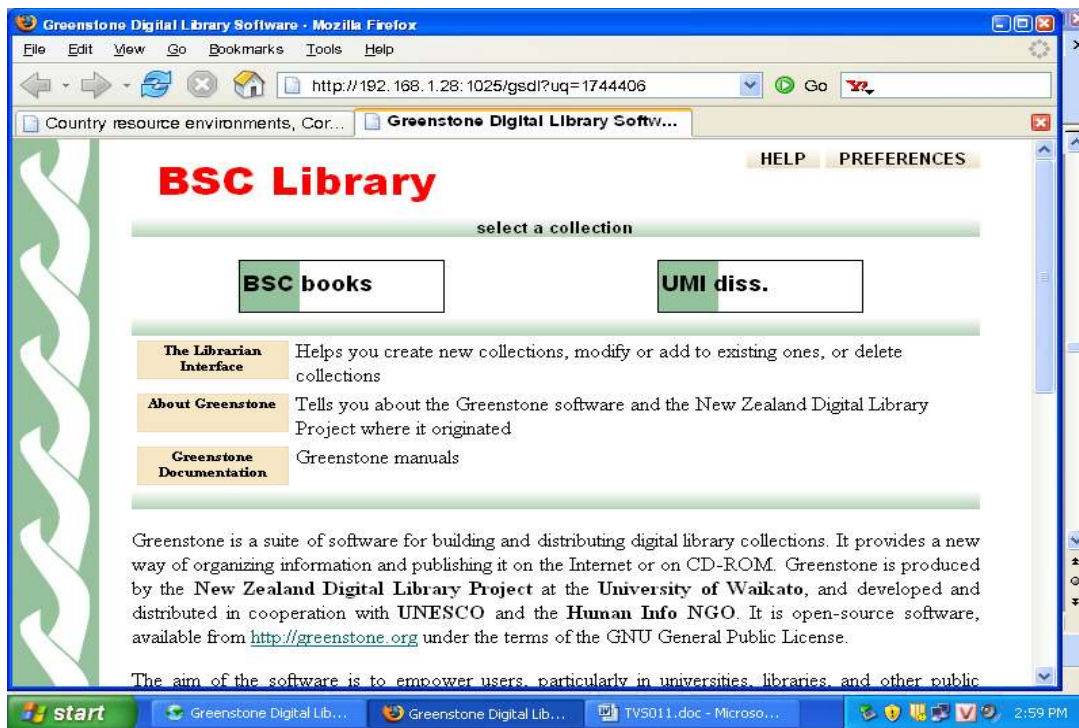


Ấu nháy vào liên kết “Toàn văn”, ta sẽ đọc được toàn văn như trang minh họa dưới đây:



Hiện chúng tôi đã chuyển sang làm việc trong một môi trường mới, công việc chủ yếu là đáp ứng nhu cầu thông tin và tài liệu tham khảo cho người dùng tin. Khác với trước đây chủ yếu tiếp xúc và xử lý tài liệu trên giấy, thì nay phần lớn thời gian chúng tôi làm việc với các CSDL trực tuyến tiềm năng như: ProQuest, ebrary và EBSCO, ... để làm tốt công việc của mình, hàng ngày chúng tôi phải download dữ liệu toàn văn trên các CSDL nói trên, tạo lập thành các CSDL theo ngành và môn học để dễ dàng phục vụ đội ngũ học viên và giảng viên của trường. Dữ liệu toàn văn download được chúng tôi cũng đưa vào GREENSTONE để dễ dàng phục vụ người dùng tin theo yêu cầu và nguyện vọng của họ.

Dưới đây là hình ảnh minh họa 2 CSDL thư mục (CSDL thư mục về sách và CSDL toàn văn các luận án, luận văn liên quan đến nội dung giảng dạy của trường) của Trường Thương mại Anh quốc (BSC British School of Commerce) được đưa vào GREENSTONE, độc giả có thể tra cứu bằng cả 2 hình thức offline và online.





Theo kinh nghiệm của chúng tôi thì việc chuyển đổi dữ liệu từ phần mềm ISIS sang phần mềm mã nguồn mở GREENSTONE và chia sẻ thông tin số hóa trên mạng INTERNET là giải pháp lựa chọn hợp lý và kinh tế cho các thư viện vừa và nhỏ có nguồn tài chính khiêm tốn. Cần chăng cần xây dựng các Tổ hợp thư viện để việc chia sẻ thông tin và sử dụng kinh phí trong việc mua sắm trang thiết bị, máy móc và phần mềm hiệu quả hơn. Giải pháp này giúp cho các đơn vị nói trên dễ dàng tạo dựng và phục vụ thông tin cho độc giả, cũng như có cơ hội quảng bá và chia sẻ tiềm năng thông tin cho những người có quan tâm đến văn hóa, đất nước và con người Việt Nam ở khắp mọi nơi trên thế giới.

Tháng 11/2011

TÀI LIỆU THAM KHẢO

Chu Tuyết Lan. Giải pháp chuyển đổi dữ liệu từ ISIS sang Greenstone để chia sẻ nguồn thông tin trên mạng Internet của Thư viện Viện nghiên cứu Hán học // Tạp chí Thư viện. – 2010. - Số 2