

Sử dụng GREENSTONE để xây dựng Bộ sưu tập Thư viện số

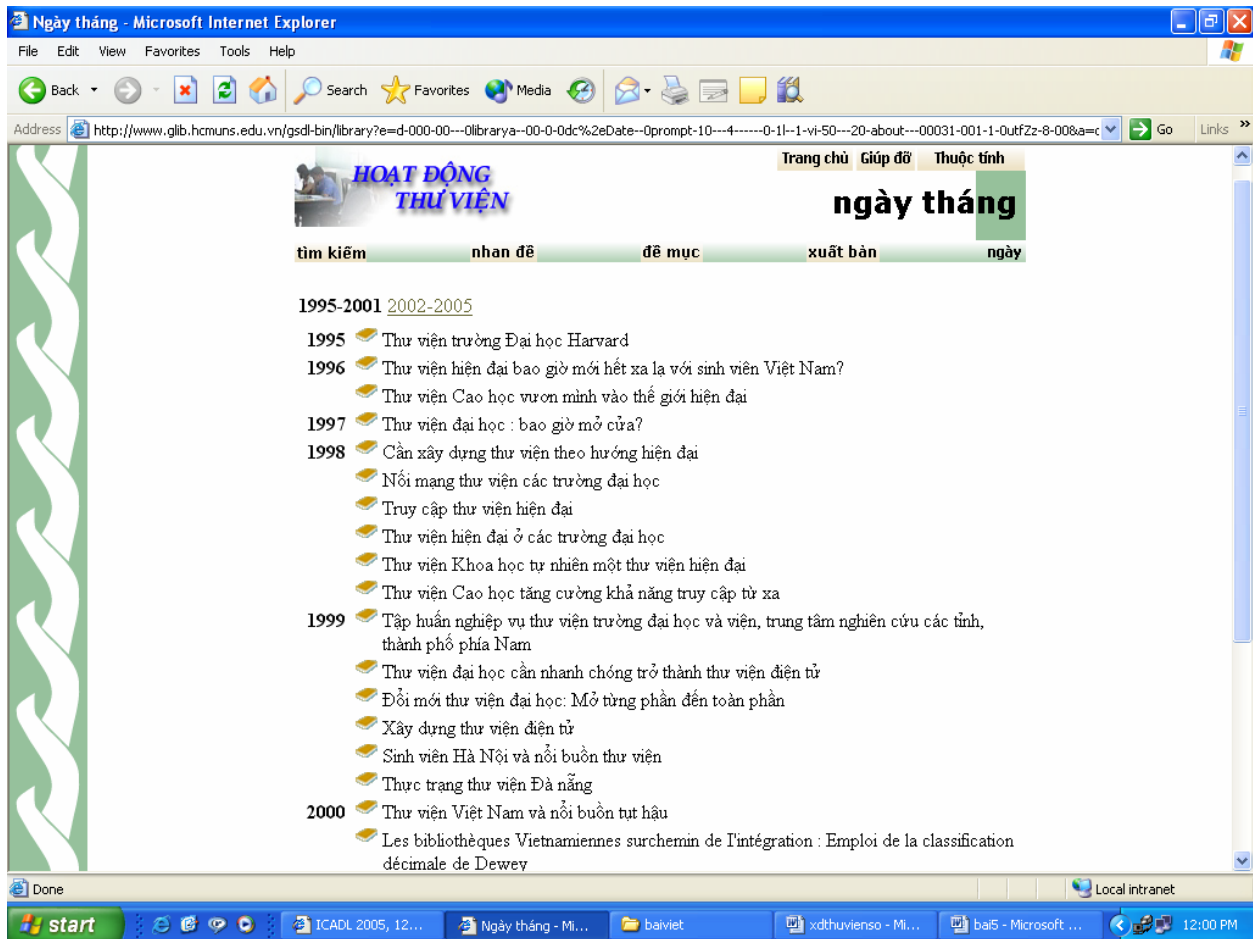
PHÒNG CÔNG TÁC KỸ THUẬT
Thư viện ĐH. Khoa học Tự nhiên
ĐHQG-HCM

Nếu chúng ta gọi *Thư viện truyền thống* là một cơ sở lưu trữ và phục vụ tài liệu in ấn thì *Thư viện điện tử* là một thư viện truyền thống có phục vụ thêm tài liệu điện tử, là tài liệu được đọc trên máy vi tính và chuyển tải trên mạng máy tính. Tài liệu điện tử bao gồm CD-ROM, cơ sở dữ liệu (CSDL) CD-ROM, CSDL trực tuyến, tạp chí điện tử, sách điện tử, vv... Đó là những tài liệu chúng ta phải bổ sung vào vốn tài liệu của thư viện với một hình thức mới: vốn tài liệu điện tử như CD-ROM chẳng hạn; hoặc chúng ta phải mua quyền sử dụng để truy cập vào những CSDL trực tuyến thương mại. Một nguồn tài liệu điện tử khác rất phong phú và đa dạng, đó chính là tài nguyên thông tin miễn phí trên mạng toàn cầu Internet. Khối lượng thông tin này được lưu thông trên Internet từ nhiều nguồn khác nhau. Có thể từ những cơ quan có thẩm quyền như chính phủ, trường đại học, viện nghiên cứu, vv...; đến những tổ chức phi chính phủ, hội đoàn, vv...; thậm chí cũng có thể từ các nhóm hay cá nhân độc lập. Bất kỳ ai cũng có thể đóng góp vào nguồn tài nguyên thông tin ngày càng trở nên đồ sộ này. Quả nhiên thông tin từ mọi nguồn ngày càng trở nên quá tải. Vấn đề đặt ra cho người quản thủ thư viện và quản lý thông tin là cần phải chọn lọc những thông tin có ý nghĩa và hữu ích, tổ chức chúng thành những *bộ sưu tập thông tin kỹ thuật số* theo từng đề tài hay môn loại. Thực hiện công việc này là một phần quan trọng trong việc hình thành *Thư viện số* và một công cụ tốt nhất hiện nay để xây dựng những bộ sưu tập như thế chính là *Phần mềm nguồn mở đa ngôn ngữ thư viện số GREENSTONE*.

Bộ sưu tập thông tin thư viện số

Theo Ian H. Witten, chuyên gia Thư viện số ĐH Waikato, New Zealand, Thư viện số là tập hợp những bộ sưu tập thông tin của các đối tượng số hoặc đã được số hóa có tổ chức và tập trung. Bằng cách sử dụng Phần mềm nguồn mở Greenstone ta có thể tạo lập một cách nhanh chóng những bộ sưu tập thông tin như thế, có tổ chức và làm tăng năng lực truy tìm và lướt tìm của người sử dụng để hình thành thư viện số.

Một bộ sưu tập thông tin bao gồm nhiều tài liệu dưới nhiều dạng thức khác nhau: văn bản, hình ảnh, âm thanh, hình ảnh động, tuy nhiên cung cấp một giao diện đồng nhất qua đó tất cả các tài liệu có thể được truy cập, mặc dù cách mà tài liệu đó hiển thị sẽ tùy thuộc vào phương tiện và dạng thức của tài liệu đó. Ví dụ muốn đọc một văn bản PDF thì phải tích hợp trình ứng dụng Adobe Acrobat hay muốn xem một đoạn video thì phải tích hợp trình ứng dụng Windows Media Player, vv... Một bộ sưu tập như thế trước khi trình bày phải qua một quá trình hình thành để tạo nên những cấu trúc hỗ trợ cho việc truy tìm và lướt tìm được dùng cho việc truy cập sưu tập, một khi xây dựng xong, bộ sưu tập có thể được xuất bản trên Internet hoặc xuất ra CD-ROM một cách hoàn toàn tự động. Một khi sưu tầm thêm tài liệu mới, ta có thể dễ dàng bổ sung thêm vào bộ sưu tập bằng cách tái xây dựng.



Hình 1: Ví dụ một bộ sưu tập Greenstone

Một thư viện nói chung bao gồm nhiều bộ sưu tập khác nhau, mỗi sưu tập tổ chức mỗi khác, tùy theo hình thức tài liệu được sưu tầm và tùy theo chủ đề, đề tài được quan tâm; tuy nhiên hoàn toàn giống nhau về phương cách xây dựng và hiển thị. Những chuyên viên thư viện số trong một thư viện với sự phối hợp và tương tác với người sử dụng sẽ tạo nên những bộ sưu tập cần thiết và hữu ích cho thư viện, phục vụ tốt công tác nghiên cứu, giảng dạy, học tập trong những cơ sở đào tạo; đồng thời đáp ứng nhu cầu khảo cứu của học giả, nhà nghiên cứu và nhiều đối tượng khác. Đây là công việc thường xuyên của một thư viện đòi hỏi chuyên viên thư viện số phải phối hợp với bộ phận Dịch vụ thông tin để nắm bắt yêu cầu thông tin của người sử dụng nhằm phục vụ tốt cho việc sưu tầm thông tin; có kỹ năng số hóa tài liệu; am hiểu chuẩn Dublin Core trong công tác biên mục tài nguyên; và nhất là thành thạo trong việc sử dụng phần mềm nguồn mở Greenstone.

Phần mềm nguồn mở đa ngôn ngữ thư viện số Greenstone

Greenstone, một bộ phần mềm dùng để xây dựng và phân phối các bộ sưu tập thư viện số, là sản phẩm của dự án **New Zealand Digital Library** của đại học Waikato, được phát triển và phân phối với sự tham gia của UNESCO và **Human Info NGO**. Đây là

phần mềm mã nguồn mở được cung cấp theo thỏa thuận của GNU General Public License.

Greenstone là phần mềm nguồn mở và tự do.

Theo Richard Stallman, chủ tịch FSF (Free Software Foundation), Phần mềm tự do và Phần mềm nguồn mở là:

- Chương trình ta có thể chạy với bất cứ mục đích nào;
- Có thể chỉnh sửa cho phù hợp với yêu cầu của mình;
- Có thể chỉnh sửa và tái phân phối bản sao, có hoặc không có thu phí trên phần chỉnh sửa của mình.

Phần mềm nguồn mở Greenstone có thể download để sử dụng miễn phí tại <http://greenstone.org>. Nếu dùng phiên bản Tiếng Việt thì phải trả phí dịch sang Tiếng Việt cho Công ty Integrated e-Solutions Việt Nam. Lệ phí này cộng với chi phí cài đặt và huấn luyện sử dụng (30 tiết) là 10.000.000 đồng. Đây là một lệ phí tương trưng so với một phần mềm tương đối lớn với nhiều tính năng hiện đại ứng dụng công nghệ mới, đã được UNESCO mua bản quyền sử dụng trở thành phần mềm nguồn mở và tự do, được phổ biến khắp nơi trên thế giới.

Greenstone là phần mềm đa ngôn ngữ.

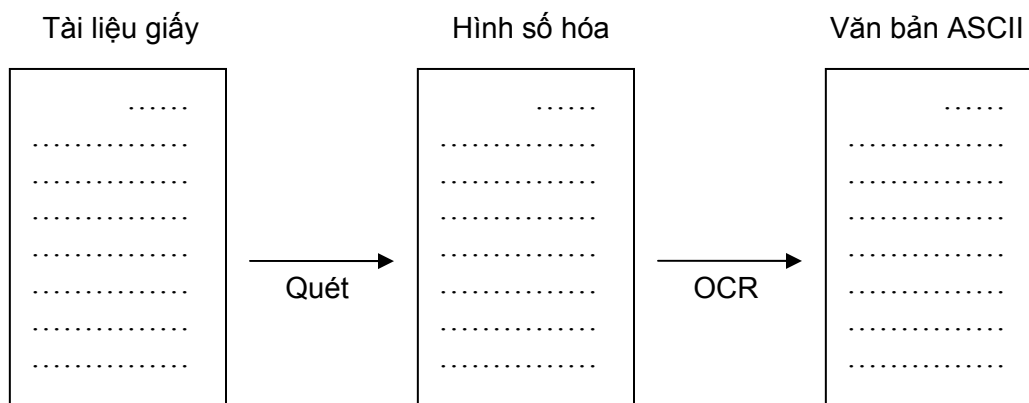
Những bộ sưu tập thông tin được tạo ra bởi Greenstone có thể chứa một số lượng lớn những tài liệu (hàng ngàn cho đến hàng triệu) với một giao diện đồng nhất. Bộ sưu tập bao gồm từ những bài báo cho đến chương trình âm nhạc. Hiện nay, Greenstone được dùng rộng rãi trong thư viện của các trường đại học trên thế giới để hình thành Thư viện số. Phần mềm Greenstone trên CD-ROM được Liên Hiệp Quốc và những cơ quan nhân đạo khác xuất bản và phân phối cho các quốc gia đang phát triển. Greenstone được dịch sang hơn 50 ngôn ngữ bao gồm:

- 4 ngôn ngữ nòng cốt: Anh, Pháp, Tây Ban Nha, và Nga. Ấn bản Greenstone Tiếng Pháp, Tây Ban Nha, và Nga được UNESCO bảo trì; ấn bản Tiếng Anh do Chương trình Dự án thư viện số New Zealand bảo trì. Khi chúng ta download Greenstone thì giao diện của bốn ngôn ngữ nòng cốt tự động được cài đặt; còn các ngôn ngữ khác cần có sự tham gia của người bảo trì hệ thống Greenstone.
- Các ấn bản của các ngôn ngữ đã hoàn chỉnh bao gồm: Arabic, Armenian, Ba Lan, Bengali, Bồ Đào Nha (Brazil), Bồ Đào Nha (Portugal), Catalan, Croatian, Czech, Đức, Farsi, Galician, Georgian, Hà Lan, Hebrew, Hindi, Hy Lạp, Indonesian, Italian, Kannada, Kazakh, Kirghiz, Latvian, Maori, Mông Cổ, Nhật, Phần Lan, Serbian, Thái Lan, Thổ Nhĩ Kỳ, Trung Hoa (phổ thông), Trung Hoa (truyền thống), Ukrainian, Việt Nam.
- Các ấn bản của các ngôn ngữ đang tiến hành phiên dịch bao gồm: Bislama, Blackfoot/Siksika, Bosnian, Esperanto (Quốc tế ngữ), Hungarian, Korean, Macedonian, Marathi, Nepali, Romanian, Shona, Tamil.

Phần cảm nang sử dụng bao gồm: “*Từ giấy đến bộ sưu tập*”, “*Hướng dẫn sử dụng*”, “*Hướng dẫn cài đặt*” được trình bày dưới các ngôn ngữ: Anh, Pháp, Tây Ban Nha, Nga, Kazakh, và Việt Nam. Các bản hướng dẫn Tiếng Việt có thể được download tại các website ở Anh, Brazil, Đài Loan, Đức, Hoa Kỳ, Ireland, Nam Phi, Nga, Nhật, Pháp, Thụy Sĩ, và Úc thông qua một địa chỉ trung gian được tìm thấy tại web site của Greenstone thế giới là <http://prdownloads.sourceforge.net/greenstone/Paper-2.39-vi.pdf>.

Công tác chuẩn bị để xây dựng Bộ sưu tập thư viện số Số hóa tài liệu

Một trong những công việc đầu tiên mà ta quan tâm khi bắt đầu xây dựng một bộ sưu tập là số hoá tài liệu. Số hoá là tiến trình chuyển tài liệu thư viện truyền thống, cụ thể là sách và văn bản sang dạng điện tử và lưu trữ trên máy tính.



Hình 2: Tiến trình số hoá

Có hai giai đoạn trong tiến trình số hoá. Giai đoạn đầu *quét hình – scanning* cho ra sản phẩm số hoá dạng hình. Giai đoạn hai cho ra một sản phẩm dạng số hoá văn bản bằng một tiến trình gọi là *nhận dạng ký tự quang học – OCR (Optical Character Recognition)*. Trong nhiều hệ thống thư viện số, tài liệu chỉ ở giai đoạn đầu, nghĩa là những gì độc giả thấy chỉ là hình ảnh. Giai đoạn hai là cần thiết đối với việc xử lý văn bản, vì tài liệu ở dạng căn bản ASCII nên ta có thể tổ chức biên mục và chỉ mục để độc giả có thể định vị bất kỳ một tổ hợp từ nào hay đối với bất kỳ kỹ thuật trích dẫn metadata tự động được định trước, chẳng hạn xác định từ khóa của tài liệu bằng cách tìm trong nội dung văn bản. Đối với văn bản Tiếng Việt, chúng ta chưa có phần mềm OCR hoàn chỉnh nên chúng ta có thể dùng dạng PDF với sự hỗ trợ của Greenstone tự động tạo nên một tập tin dạng HTML kèm theo để giải quyết việc truy tìm trong nội dung văn bản.

Số hóa cái gì?

Chuyển một thư viện thường sang dạng số là một cách làm đầy tham vọng và đắt tiền. Số hóa nội dung của một sưu tập in ấn thường là một công việc đồ sộ và chán nản. Thế nhưng có người cho rằng muốn xây dựng thư viện số thì phải số hoá toàn bộ tài liệu có trong thư viện. Đây là một quan niệm hết sức sai lầm, thực ra đây là một ảo tưởng vì

thực tế không có một thư viện nào trên thế giới có đủ nhân lực tài lực để thực hiện công việc này. Và lại cũng không cần thiết phải làm như thế, công nghệ thay đổi nhanh chóng làm cho công việc tốn kém nhiều công sức của chúng ta sẽ nhanh chóng trở nên lạc hậu. Trước khi bắt tay vào việc số hóa một sưu tập chúng ta cần phải cân nhắc thật kỹ lưỡng liệu có thật cần thiết để thực hiện không.

Một khi chúng ta đã quyết định tiến hành thì vấn đề then chốt là xác định độ ưu tiên của tài liệu để chuyển đổi. Tài liệu thư viện có thể chia làm ba loại: sưu tập đặc biệt và tài liệu một bản, chẳng hạn sách quý hiếm và bản viết tay; tài liệu được sử dụng cao, thường xuyên được yêu cầu cho giảng dạy và nghiên cứu; và tài liệu có mức độ sử dụng thấp bao gồm tài liệu nghiên cứu ít dùng thường xuyên.

Nguyên tắc số hóa tài liệu.

Có sáu nguyên tắc được xác định nhằm chọn tài liệu để số hóa hướng đến việc phát triển sưu tập thư viện số:

1. *Tính hữu dụng*: Hữu dụng là lý do cơ bản trước tất cả mọi quyết định phát triển sưu tập. Tài liệu có tần suất sử dụng cao (như giáo trình, tài liệu tham khảo mà các giáo viên thường yêu cầu tất cả sinh viên tìm đọc);
2. *Nhu cầu nội bộ*: Sưu tập nội bộ được xây dựng để phục vụ nhu cầu nội bộ và chi phí cho tài nguyên nội bộ phải được thuyết minh vì lợi ích nội bộ – chẳng hạn như đối với thư viện đại học, yêu cầu học tập, giảng dạy, và nghiên cứu là ưu tiên;
3. *Tài liệu mới*: Mặc dù sưu tập cũ mang tính lịch sử là cần thiết cho nghiên cứu, nhưng tài liệu mới vẫn ưu tiên hơn;
4. *Tài liệu liên quan đến bản gốc*: Những tài liệu mà người muốn tìm hiểu không thể tiếp cận được bản gốc (ví dụ các văn bản viết tay – "manuscript" của các nhà thơ, nhà văn, các nhà chính trị, hoặc các bản tuyên ngôn có chữ ký của các lãnh tụ như bản tuyên ngôn độc lập của Hoa Kỳ hiện có tại Thư viện Quốc hội Hoa Kỳ, vv...). Trên thực tế, còn có rất nhiều thể loại viết tay trên những chất liệu khác nhau. Việc số hoá các bản viết tay đó tạo điều kiện tiếp cận thuận lợi hơn cho các nhà nghiên cứu;
5. *Tài liệu quý hiếm*: Tài liệu quý hiếm, lâu năm, độc giả không thể trực tiếp sử dụng, dễ hư hỏng – chẳng hạn như tài liệu chữ Nôm trên giấy bồi;
6. *Chuyển đổi nhận thức*: Ngày càng có nhiều thông tin chuyển sang dạng số. Tài liệu giúp người sử dụng chuyển đổi nhận thức để làm quen việc sử dụng dạng thông tin này là ưu tiên.

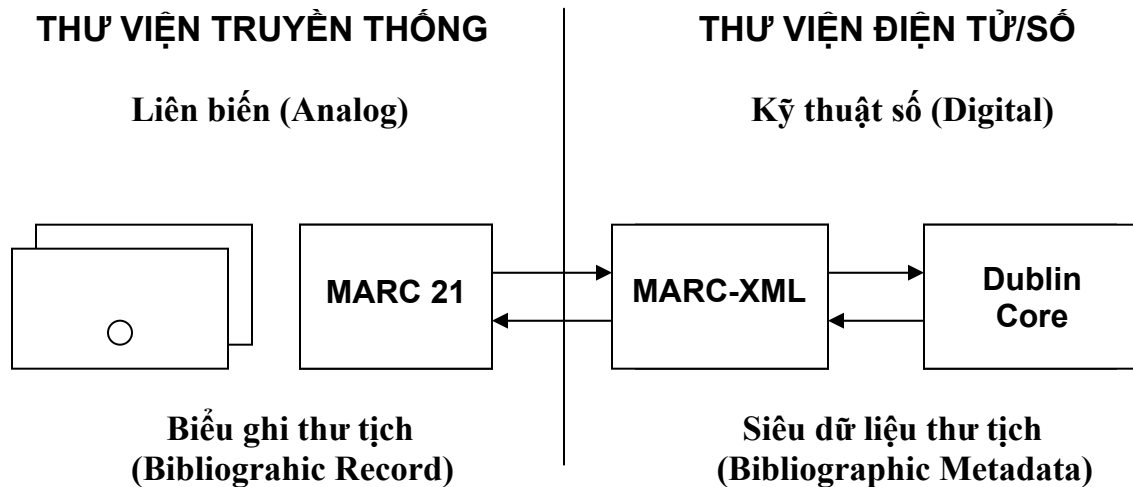
Chúng ta cần phải cân nhắc mức độ ưu tiên đối với những nguyên tắc trên trong việc chọn tài liệu để số hóa.

Siêu dữ liệu - metadata

Để xây dựng một sưu tập mới ta thường phải đối mặt với cả hai loại tài liệu: tài liệu đã ở dạng điện tử rồi và tài liệu in ấn cần phải số hoá. Nếu toàn bộ tài liệu ở dạng

điện tử thì công việc hết sức dễ dàng, ngay cả việc sưu tầm, tổ chức tập tin và chuyển đổi dạng thức; công việc này rõ hơn nhiều so với việc số hoá tài liệu.

Vấn đề là chúng ta phải xác định metadata. Có được metadata cần thiết và chuyển đổi qua dạng điện tử thường là công việc chính trong vấn đề xây dựng sưu tập. Khi số hoá một thư viện hiện hữu thì metadata có sẵn rồi, nhưng khi xây dựng sưu tập mới việc xác định metadata là phức tạp hơn.



Hình 3: Thay đổi biên mục qua môi trường Web

Trong thư viện truyền thống, người ta biên mục để tạo nên những biểu ghi thư tịch nhằm xây dựng hệ thống tra cứu qua mục lục phiếu. Biểu ghi thư tịch hay mục lục phiếu miêu tả lý lịch của tài liệu: nhan đề, tác giả, đề mục, xuất bản, vv... Khi sử dụng máy tính, biểu ghi thư tịch này được biểu thị bằng một dạng thức máy đọc được (MARC). Cách biên mục này chỉ thể hiện được dạng thư tịch tức lý lịch chứ không có toàn văn và đa phương tiện, được gọi là biên mục theo dạng liên liên (analog).

Trong môi trường số, dữ liệu được đóng gói bằng ngôn ngữ XML. Cách biên mục phải thay đổi qua môi trường Web, nghĩa là các dữ liệu thư tịch phải được đóng gói, người ta gọi là biên mục theo dạng kỹ thuật số (digital). Các biểu ghi thư tịch trở thành siêu dữ liệu thư tịch – metadata. Vậy Metadata chính là phiếu mục lục miêu tả lý lịch tài liệu được phát sinh tự động trong môi trường số.

Hai phương pháp chuẩn trình bày Metadata

Đã từng có hai phương pháp chuẩn đối nghịch nhau về trình bày siêu dữ liệu tài liệu: Dạng biên mục máy đọc được MARC và Dublin Core. Dạng MARC được phát triển công phu, kiểm soát chặt chẽ, chi ly và bao hàm đến độ khá phức tạp, được tạo nên bởi những nhà biên mục học chuyên nghiệp chủ yếu để sử dụng trong thư viện truyền thống. Chuẩn Dublin Core chủ trương đơn giản hóa để có thể áp dụng rộng rãi cho tài liệu thư viện số đối với những người không cần được huấn luyện biên mục thư viện. Hai chuẩn này không những chú ý đến giá trị đặc thù của mình mà còn lưu tâm đến những triết lý căn bản đối nghịch nhau một cách tuyệt đối.

Chuẩn MARC được Thư viện Quốc hội Hoa Kỳ phát triển vào cuối thập niên 1960 để phục vụ việc trao đổi biểu ghi mục lục giữa các thư viện. MARC được giảng dạy khá kỹ lưỡng trong những chương trình đào tạo thư viện học trên thế giới. Chúng ta khá quen thuộc với biểu ghi MARC khi tiếp xúc với mục lục trực tuyến ở thư viện đại học.

Chuẩn Dublin Core là một tập hợp những thành phần metadata được thiết kế đặc biệt cho việc sử dụng không chuyên. Được dùng chủ yếu cho việc mô tả tài liệu điện tử. Đây là kết quả của một sự hợp tác nhiều người cùng xây dựng. Dublin là tên của thành phố ở Bang Ohio, Hoa Kỳ, nơi cuộc họp đầu tiên được tổ chức vào năm 1995. Từ đó đến nay đã có 12 lần hội nghị quốc tế tổ chức tại Anh, Canada, Đức, Hoa Kỳ, Nhật Bản, Phần Lan, Trung Quốc, và Úc để hoàn thiện. Dublin Core được Tổ chức Chuẩn Quốc gia Hoa Kỳ – ANSI phê chuẩn vào năm 2001. So với dạng MARC, Dublin Core đơn giản một cách dễ chịu. Dublin Core chỉ bao gồm 15 thành phần so với hàng trăm của MARC. Như cái tên "core – nòng cốt" đã hàm ý rằng Dublin Core là một tập hợp những thành phần nòng cốt, ngoài ra còn có thể tăng thêm những thành phần phụ cho mục đích riêng. Hơn nữa, những thành phần hiện hữu có thể được cải tiến xuyên qua việc sử dụng. Tất cả những thành phần này đều có thể lập lại khi cần thiết.

Ngày nay sự đối nghịch giữa hai chuẩn không còn tồn tại mà ngày càng nhiều chuyên gia tìm những giải pháp để chuyển đổi nhau: MARC sang Dublin Core và Dublin Core sang MARC, chẳng hạn như tại website của Văn phòng Chuẩn MARC và phát triển mạng lưới của Thư viện Quốc hội Hoa Kỳ có hướng dẫn các phương thức chuyển đổi. Thậm chí còn có nhiều phần mềm chuyển đổi được phổ biến rộng rãi, chẳng hạn như những phần mềm MarcEdit của Đại học Bang Oregon, Hoa Kỳ mà ta có thể download để dùng miễn phí tại website: <http://oregonstate.edu/~reaset/marcedit/html/downloads.html>.

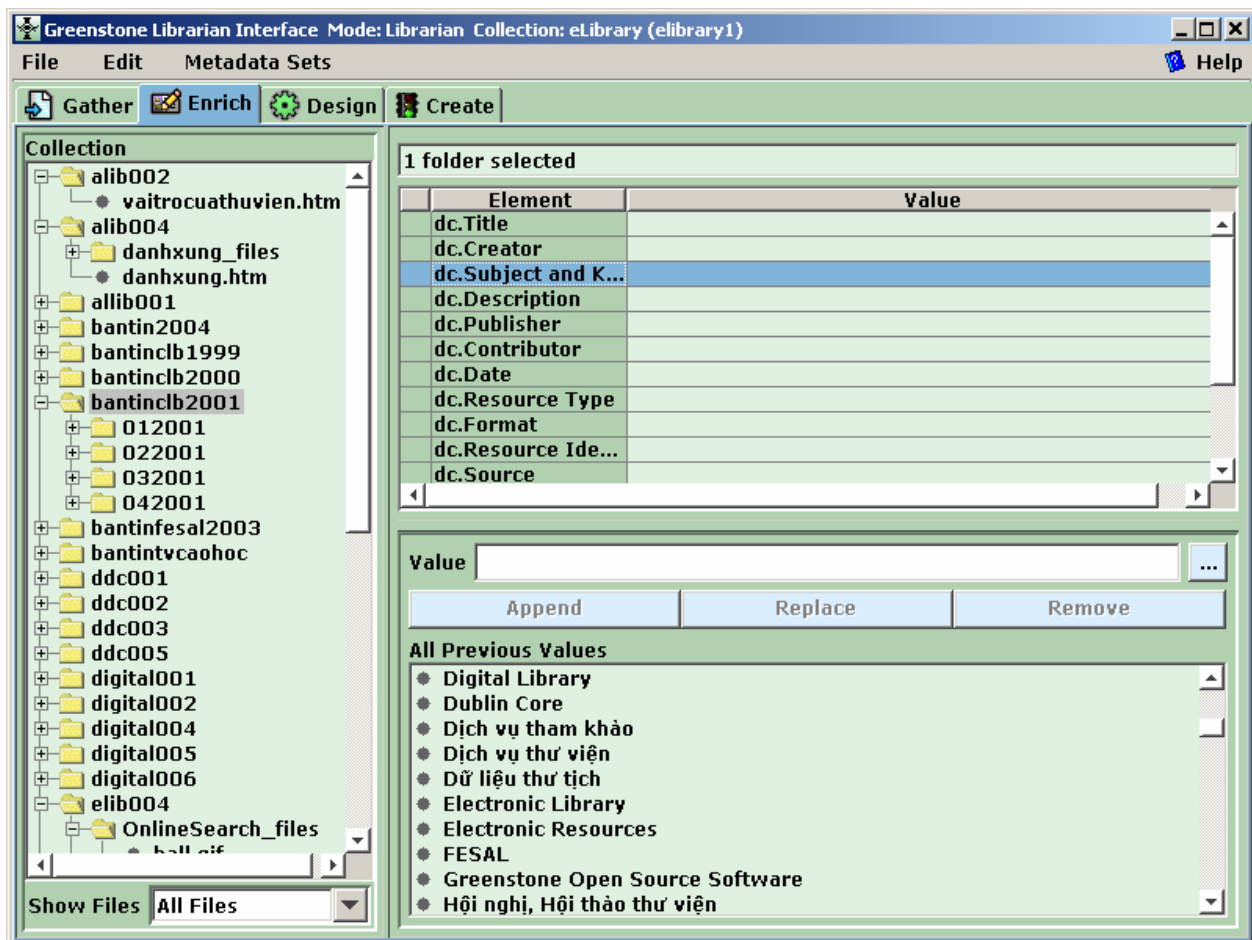
Xây dựng bộ sưu tập với Greenstone.

Phương thức chung.

Thường thì ta dùng công cụ LIBRARIAN INTERFACE để tổ chức một tài liệu, công việc này trong nghiệp vụ thông tin - thư viện gọi là biên mục (cataloging) và chỉ mục (indexing), còn trong công nghệ kỹ thuật số thì gọi là xác định metadata. Một tài liệu sau khi được tổ chức biên mục theo chuẩn Dublin Core (thủ công) và xác định metadata (tự động) sẽ trở thành một thư mục chứa năm thư mục con: *Import* (chứa tài liệu thô); *Archives* (chứa tập tin tự tạo dạng GML), *Index* (chứa các tập tin cuối cùng phục vụ người dùng kèm theo metadata), *Building* (thư mục trung gian trong quá trình xây dựng bộ sưu tập), *-etc* (thư mục bổ trợ chứa các tập tin điều khiển quá trình kiến tạo bộ sưu tập).

Giao diện INTERFACE LIBRARIAN.

Giao diện LIBRARIAN INTERFACE trình bày 15 yếu tố của Dublin Core cho ta biên mục tài liệu. Quá trình này khiến ta chọn những dẫn mục (entry) hay điểm truy cập (access point) của mỗi tài liệu để phục vụ việc truy tìm và lướt tìm sau này. Cũng bằng giao diện này, Greenstone sẽ cho ta xác định những dẫn mục và hình thức truy tìm hay lướt tìm được trình bày trên giao diện của bộ sưu tập. Chẳng hạn như: Nhan đề (Title), Tác giả (Author), Từ khoá (Keyword), Đề mục (Subject), vv... hoặc Tìm kiếm (Search)



Hình 4: Giao diện LIBRARIAN INTERFACE với 15 yếu tố của Dublin Core

thì tìm kiếm trên những điểm truy cập nào. Trong quá trình tổ chức tài liệu, việc thêm, bớt hay thay thế những yếu tố của Dublin Core là rất dễ dàng. Công việc sưu tầm và tổ chức tài liệu cứ tiếp diễn liên tục, tài liệu được lưu vào máy tính cá nhân. Greenstone hoàn toàn xử lý một cách tự động và nhanh chóng một khi chúng ta muốn xuất bản tài liệu như một bộ sưu tập lên Internet hay CD-ROM. Dĩ nhiên chúng ta vẫn có thể cập nhật tài liệu vào bộ sưu tập mỗi khi cần thiết; khi đó thì ta phải xuất bản lại bộ sưu tập.

Mỗi bộ sưu tập được xuất bản lên Internet hay CD-ROM đều có một giao diện Greenstone kèm theo. Nếu trên CD-ROM thì giao diện Greenstone có chứa sẵn một phần trình duyệt (web browser) Netscape để tải xuống (download) cho những máy cá nhân nào không sử dụng web. Như thế cho chúng ta thấy rằng mỗi khi một bộ sưu tập được xuất ra CD-ROM thì người sử dụng có thể dùng bất kỳ một máy tính với bất kỳ một hệ điều hành nào đều có thể đọc, truy tìm, lướt tìm, in ra những thông tin trên bộ sưu tập với giao diện thân thiện của Greenstone. Nếu chúng ta sử dụng và truy cập vào các bộ sưu tập của các quốc gia khác nhau khắp nơi trên thế giới thì có thể sử dụng giao diện của nhiều ngôn ngữ, trong đó có cả giao diện Tiếng Việt.

Giao diện INTERFACE LIBRARIAN cung cấp bốn giao diện tương tác phản ánh các bước thực hiện như sau:

- GATHER: Dùng để thu gom tài liệu tập trung vào bộ sưu tập;
- ENRICH: Trình bày 15 yếu tố của Dublin Core để biên mục tài liệu. Động tác biên mục được làm thủ công. Người biên mục có thể chọn trên danh sách những yếu tố có sẵn hoặc “append” một yếu tố mới vào; cũng có thể “replace” một yếu tố hay hơn và “remove” một yếu tố khác đi;
- DESIGN: Dùng để thiết kế giao diện bộ sưu tập kèm theo những chỉ thị tìm kiếm theo những dẫn mục cho ta chọn, chẳng hạn như *tác giả, nhan đề, đề mục, năm, nguồn*, vv...;
- CREATE: Dùng để xuất bộ sưu tập lên Internet hay ra đĩa CD. Đây là một thao tác tự động.

Các bước thực hiện.

Để tạo được một bộ sưu tập phải thực hiện nhiều bước, cụ thể là: khai báo thông tin bộ sưu tập, xác định dữ liệu nguồn, cấu hình và xây dựng bộ sưu tập. Trong đó, bước cấu hình đóng vai trò quan trọng nhất vì nó quyết định bộ sưu tập có hoạt động được hay không.

- Bước đầu tiên là khai báo tên của bộ sưu tập và các thông tin liên quan như địa chỉ email, thông tin mô tả sưu tập. Bước này sẽ giúp chúng ta quản lý dễ dàng khi đã xây dựng được nhiều bộ sưu tập.
- Bước thứ hai là xác định nguồn dữ liệu để đưa vào bộ sưu tập.
 - Tên thư mục trong cùng hệ thống sẽ bắt đầu bởi "file://"
 - Địa chỉ bắt đầu bằng "http://" các tập tin sẽ được lấy về từ Web.
 - Địa chỉ bắt đầu bằng "ftp://" các tập tin sẽ được lấy về qua FTP.

Bộ sưu tập sẽ gồm tất cả các tập tin, các thư mục con trong thư mục được chỉ định. Đối với “http://” thì bộ sưu tập sẽ ánh xạ đến web site được chỉ định.

Khi tạo mới một bộ sưu tập hay thêm tài liệu vào sưu tập đã sẵn có, đầu tiên tài liệu nguồn được đưa vào hệ thống – *importing*. Đồng thời tài liệu được chuyển sang dạng tương tự HTML, đó là GML (“Greenstone Markup Language”), định dạng có kèm theo metadata vào tài liệu. Tài liệu được mang mã Unicode UTF-8.

Cấu trúc của tài liệu đáp ứng việc tìm kiếm theo chỉ mục. Gồm có ba cấp của chỉ mục: tài liệu, phân khu và các đoạn, ứng với các đặc điểm khác nhau mà GML tạo cấu trúc phân cấp phù hợp để tạo chỉ mục. Các chỉ mục có thể là dạng text, metadata hay bất kỳ sự kết hợp nào. Do đó, chúng ta có thể tạo các chỉ mục tìm kiếm theo nhan đề, theo tác giả, theo nơi chỉ định trong tài liệu cũng như nội dung tài liệu.

- Bước thứ ba là cấu hình các tùy chọn, yêu cầu cần người dùng có sự am hiểu nhất định. Cấu trúc và hình thức của bộ sưu tập sẽ được định rõ trong tập tin cấu hình.

Tài liệu được chuyển sang định dạng XML một cách tự động. Các plugin thích

hợp sẽ được chọn trong tập tin cấu hình. Sau đó chỉ mục tìm kiếm và cấu trúc trình duyệt toàn tài liệu được qui định trong tập tin cấu hình đã tạo. Cuối cùng, kết quả của tiến trình xây dựng tạo ra bộ sưu tập hoạt động được.

Giai đoạn xây dựng có thể tốn nhiều thời gian. Những bộ sưu tập nhỏ mất khoảng một vài phút nhưng với những sưu tập lớn thì có thể mất cả giờ hay nhiều hơn.

Sở hữu trí tuệ và bản quyền.

Trong một thư viện truyền thống, quyền sở hữu tài liệu là quan trọng; nhưng trong lĩnh vực lưu hành tài nguyên điện tử, quyền sở hữu trí tuệ, cụ thể là quyền tác giả hay bản quyền là quan trọng hơn.

Sưu tầm thông tin và làm cho thông tin đó trở nên phổ biến hơn đối những người khác là một điều liên quan đến vấn đề xã hội, và những người xây dựng thư viện số phải am hiểu quyền sở hữu trí tuệ để hành động một cách có trách nhiệm và đúng luật xung quanh những ứng dụng cụ thể của họ.

Thư viện số có thể làm cho việc truy cập trở nên rộng rãi hơn thư viện truyền thống. Và chính điều này đã nảy sinh ra nhiều vấn đề: truy cập thông tin trong thư viện số, nói chung ít bị kiểm soát hơn truy cập sưu tập in ấn trong thư viện thường. Đưa thông tin vào thư viện số là có khả năng làm cho thông tin đó trở nên phổ biến ngay đối với một số lượng độc giả hầu như vô hạn.

Sở hữu một cuốn sách chắc chắn không phải là xác lập được quyền sở hữu đối với tài liệu đó theo nghĩa của bản quyền. Mặc dù có nhiều bản của một tài liệu nhưng chỉ có một bản quyền. Điều này không chỉ áp dụng cho bản in mà cả cho bản điện tử, dù được số hoá từ bản in hay được tạo nên dưới dạng điện tử từ đầu. Khi mua một cuốn sách, ta có thể bán lại, nhưng chắc chắn không mua quyền tái phân phối. Quyền đó tùy thuộc vào bản quyền.

Ai làm chủ một tác phẩm cụ thể? Bản quyền đầu tiên là của người sáng tác trừ phi tác phẩm được thuê sáng tác. Trong trường hợp này bản quyền thuộc về cơ quan hay tổ chức thuê theo hợp đồng; bản quyền có thể được sang nhượng hay chuyển cho một đơn vị khác thông qua một hợp đồng cụ thể, được thực hiện bằng văn bản do người chủ ký tên.

Luật bản quyền là phức tạp. Tình trạng luật pháp đối với tập tin máy tính và tài liệu cụ thể được xuất bản trên World Wide Web lại mù mờ. Muốn xây dựng thư viện số thì phải cần số hoá tài liệu. Chúng ta phải làm như thế nào để tránh vi phạm bản quyền? Trước hết chúng ta phải xem xét:

- Nếu tác phẩm được số hoá ở trong miền (domain) công cộng thì chúng ta không phải xin phép ai hết. Dĩ nhiên kết quả số hoá của chúng ta cũng không được bảo vệ bản quyền, trừ phi kết quả của ta nhiều hơn bản gốc;
- Nếu tài liệu được tặng cho cơ sở của ta để số hoá và người tặng có bản quyền, thì chúng ta tiến hành số hoá, tuy nhiên cần phải yêu cầu người tặng cung cấp cho mình quyền được số hoá – có thể bằng một mẫu giấy có ghi "*quyền sử dụng tác phẩm với bất kỳ mục đích chung của cơ sở, dưới bất kỳ phương tiện nào*".

Nếu ta muốn số hoá tài liệu mà không rơi vào hai trường hợp trên thì ta phải cân nhắc thử việc số hoá của chúng ta có phải là một việc làm có lợi ích chung mà không xâm phạm quyền lợi của người khác. Đây là một điều khó về mặt pháp lý. Cuối cùng nếu chúng ta không chắc chắn với điều cân nhắc trên thì ta phải tiến hành xin phép để được cấp phép thực hiện số hoá.

Tóm lại để tiến hành xây dựng thư viện số, ta phải lưu ý đến vấn đề bản quyền. Những người thực hiện phải cam kết hiểu biết đầy đủ về bản quyền và nhận thức sâu sắc rằng giấy phép là rất cần thiết để chuyển đổi tài liệu số hóa không thuộc trong miền công cộng.

Ứng dụng bộ sưu tập.

1. Xây dựng Kho tài nguyên học tập.

Kho tài nguyên học tập là hình thức tiêu biểu nhất của Thư viện số trong một trường đại học. Đây là kho tài nguyên điện tử, tùy theo qui mô và chức năng của Trường đại học mà một kho tài nguyên học tập có thể trở nên rất đa dạng. Thường bao gồm:

- Tài liệu đa phương tiện: Bản đồ, hình ảnh, mẫu vật, đoạn băng hình thí nghiệm, băng hình bài giảng, hội nghị khoa học, khối giáo trình (course block), vv... phục vụ như là công cụ trợ giảng, đào tạo từ xa.
- Bộ sưu tập chuyên ngành phục vụ nghiên cứu, giảng dạy, học tập.

Việc tạo lập những bộ sưu tập chuyên ngành bằng Phần mềm nguồn mở Thư viện số Greenstone để đóng góp cho Kho Tài nguyên học tập là công việc do đội ngũ giảng viên, nghiên cứu, nói chung là những người sử dụng cùng với đội ngũ nhân viên thư viện xây dựng, theo đúng tinh thần **"Thư viện số là sự tương tác giữa người sử dụng với thư viện để phục vụ chính người sử dụng"**. Công việc này sẽ trở thành một sinh hoạt bình thường trong một trường đại học trong giai đoạn hiện nay; tuy nhiên bước khởi đầu cần phải có sự đầu tư và động viên của lãnh đạo nhà trường.

Phàm là một cán bộ giảng dạy đại học hay nghiên cứu thì ai cũng tự mình sưu tầm một số tài liệu về chuyên ngành của mình, đó là sách tham khảo, photocopy những chương sách, bài báo, mẫu vật, hình ảnh minh họa, bản thiết kế, băng hình thí nghiệm, tập tin trên máy tính hay CD-ROM, vv... Ngày nay chúng ta đang bước vào kỷ nguyên tri thức với công nghệ kỹ thuật số có thể giúp cho công việc sưu tầm tài liệu của người cán bộ giảng dạy đại học dễ dàng và phong phú hơn. **Với sự hỗ trợ của thư viện, từng cán bộ giảng dạy và nghiên cứu sử dụng phần mềm nguồn mở Greenstone để xây dựng bộ sưu tập cho chính mình rồi đóng góp vào thư viện để làm phong phú Kho tài nguyên học tập.**

2. Tái đóng gói thông tin phục vụ Dịch vụ tham khảo.

Dịch vụ tham khảo cung cấp thông tin cho người sử dụng từ mọi nguồn có trong và ngoài thư viện. Sử dụng phần mềm Greenstone để tổ chức thông tin thành những bộ sưu tập chuyên ngành theo những đề tài được yêu cầu. Những bộ sưu tập này được tái đóng gói lên CD-ROM để cung cấp cho người sử dụng. Với giao diện đồ họa thân thiện của Greenstone, người sử dụng có thể truy tìm và lướt tìm thông tin theo từ khóa, tác giả, nhan đề, đề mục và những điểm truy cập khác trên chính CD-ROM của mình.

3. Lưu trữ và quản lý công văn.

Trong việc lưu trữ và quản lý công văn, mỗi văn bản được xem như một tài liệu. Sử dụng chuẩn Dublin Core linh hoạt để biên mục từng tài liệu và quản lý theo cấp, đề tài, ngày, cơ quan chủ quản, người ban hành và ký công văn, vv... Người sử dụng dễ dàng truy tìm và lướt tìm Bộ sưu tập công văn cho dù lớn bao nhiêu.

4. Phục vụ công tác địa chí.

Những tài liệu địa chí của địa phương bao gồm đủ mọi hình thức từ những di chỉ, mẫu vật đến công trình kiến trúc, tài liệu văn bản cổ, vv... đều có thể số hóa; rồi biên mục trên từng tài liệu và tổ chức trong một hay nhiều bộ sưu tập theo đề tài. Thật dễ dàng trong việc lưu trữ phục vụ công tác nghiên cứu cũng như quảng bá rộng rãi phục vụ công tác du lịch.

Kết luận

Thư viện luôn luôn đóng vai trò quan trọng trong việc hỗ trợ công tác học tập, nghiên cứu và giảng dạy trong một trường đại học; cũng như đáp ứng nhu cầu thông tin cho quần chúng. Nhu cầu này ngày càng cao do đó giá trị nghiệp vụ thư viện cũng thay đổi từ chỗ thư viện là nơi tập trung thông tin chờ người đến sử dụng, người cán bộ thư viện đóng vai trò của người giữ sách thụ động; ngày nay thư viện là nơi đáp ứng tức thì nhu cầu của người sử dụng, người cán bộ thư viện có vai trò chủ động của người cung cấp thông tin và hợp tác với người sử dụng để làm phong phú nguồn tri thức vì lợi ích chính người sử dụng.

Phần mềm nguồn mở thư viện số Greenstone là một trong những công nghệ mới giúp người quản lý thông tin thực hiện được vai trò chủ động trên.

TÀI LIỆU THAM KHẢO

1. FOX, Adward A., SULEMAN, Hussein, LUA. Ming. *Building Digital Libraries Made Easy: Toward Open Digital Libraries*. Proceedings. – 5th ICADL 2002. – Singapore, 11-14/12/2002.
2. LOOT, Michel, CAMARZAN, Dan, WITTEN, Ian, BODDIE, Stefan. *Từ giấy đến bộ sưu tập*. <http://prdownloads.sourceforge.net/greenstone/Paper-2.39-vi.pdf>.
3. PANDIAN, B. Maruthu, SONKER, Sharad Kumar và MOORTHY R. *Creating Digital Libraries: An Experiment with Greenstone Digital Library Open Source Software*
4. WITTEN, Ian H., BAIBRIDGE, David và BODDIE, Stefan J.. *Greenstone Open-Source Digital Library Software*, D-Lib Magazine, October 2001, Volume 7, Number 10 (www.dlib.org)
5. WITTEN, Ian H. và BAIBRIDGE, David. *How to Build a Digital Library*. – New York : Morgan Kaufmann, 2003.