

Giải pháp chuyển đổi dữ liệu từ ISIS sang Greenstone để chia sẻ nguồn lực thông tin trên mạng Internet của Thư viện Viện nghiên cứu Hán - Nôm



I. Nguồn lực và sản phẩm thông tin

Tiền thân Thư viện Viện Nghiên cứu Hán Nôm là Thư viện của một viện nghiên cứu chuyên ngành được thành lập ngay sau khi Ban Hán Nôm ra đời năm 1970, lúc đầu có tên gọi là Tổ tư liệu, sau là phòng Thông tin-Tư liệu-Thư viện và hiện nay là Thư viện Viện Nghiên cứu Hán Nôm.

Thư viện có vốn tài liệu tuy khiêm tốn, nhưng lại rất đa dạng và phong phú về mặt chủng loại, trong số đó có nhiều tài liệu quý và được lưu trữ duy nhất tại Thư viện. Vốn tài liệu của Viện Nghiên cứu Hán Nôm hiện có, được hình thành từ 3 nguồn chính:

1. Mua trực tiếp của nhà nước; mua lại của các thư viện tư nhân; qua con đường trao đổi, biếu tặng gồm: Tài liệu tra cứu, tham khảo (tiếng Việt và các ngôn ngữ khác) trong đó bao gồm 20.170 đơn vị sách, bản đồ và trên 7.000 đơn vị tạp chí các loại;

2. Ngoài việc tiếp nhận nguồn tài liệu Hán Nôm của Trường Viễn đông Bác cổ Pháp, (do Viện Thông tin KHXH chuyển giao vào năm 1980), từ năm 1991 đến nay Viện thực hiện công tác sưu tầm tài liệu Hán Nôm trong cả nước bằng nguồn kinh phí do Nhà nước tài trợ, nâng vốn tài liệu hiện có của Viện Nghiên cứu Hán Nôm lên một cách đáng kể:

- Tài liệu Hán Nôm: Kho sách tổng hợp có khoảng 34.000 đơn vị;

 - Kho thác bản văn khắc có trên 56.000 đơn

 - Kho ván khắc in cổ có khoảng 20.000 đơn vị;

1. Tài liệu sưu tầm ở nước ngoài: Gần đây trong khuôn khổ hợp tác giữa Viện Nghiên cứu Hán Nôm với các cá nhân và tổ chức quốc tế, Viện đã thực hiện một cách hiệu quả công tác sưu tầm tư liệu ở Mỹ và Pháp, kết quả đã thu được tổng cộng 33.218 trang ảnh tài liệu Châu bản và sách Hán Nôm, cụ thể như sau:

- 64 cuộn microfilm tổng cộng 13.960 trang ảnh lưu giữ hầu như toàn bộ kho Châu bản Triều Nguyễn ở Việt Nam, do Viện Harvard-Yenching Hoa Kỳ gửi tặng Viện tháng 12 năm 2007;

- 6 cuộn microfilm tổng cộng 1.400 trang ảnh lưu giữ những tài liệu Hán Nôm mà Viện chưa hề có, được GS. Liam Kelley thay mặt Đại học Hawaii, Mỹ tặng Viện;

- 16.858 trang ảnh của hơn 54 tên sách đã được số hóa với sự tài trợ của Viện Harvard- Yenching Hoa Kỳ và Học viện Viễn đông Bác cổ Pháp, đây là toàn bộ tài liệu Hán Nôm tại Thư viện của Học viện Viễn đông Bác cổ Pháp và một phần sách Hán Nôm tại Thư viện Hiệp hội châu Á, Paris.

- 1.000 trang ảnh của 30 tài liệu chữ Nôm do Ngài Alexandre Lê, nguyên quản thủ Thư viện Á Châu Viện Viễn đông Bác cổ Pháp gửi tặng.

- Sản phẩm thông tin: Với sự đầu tư tài chính có hạn, Thư viện Viện Nghiên cứu Hán Nôm đã từng bước chuyển từ phục vụ thủ công sang tự động hoá các hoạt động của thư viện. Hầu hết vốn tài liệu và các tư liệu chuyên ngành của Viện đã được quản lý ở trên máy, hiện Thư viện đã làm được tổng số 39.678 biểu ghi thể hiện ở các CSDL sau đây:

1. SACH: Quản lý toàn bộ vốn tài liệu hồi cổ tiếng Việt, với 6.073 biểu ghi;

2. KSHN: Quản lý sách Hán Nôm các loại, với 10.664 biểu ghi;

3. NVDD: Giới thiệu về các nhà khoa bảng Việt Nam thời phong kiến, với 3.126 biểu ghi;

4. TCMN: Quản lý các bài trích trong các tạp chí xuất bản ở miền Nam trước năm 1975 và các tạp chí Nam Phong, Tri Tân..., thuộc lĩnh vực KHXH, với 3.618 biểu ghi;

5. TCHN: Quản lý các bài trích trong Tạp chí Hán Nôm (từ khi thành lập cho đến nay), với trên 1.512 biểu ghi, cho phép bạn đọc tiếp cận với thông tin thuộc lĩnh vực nghiên cứu Hán Nôm theo nhiều chiều khác nhau;

6. Tmục: Quản lý bài trích trong các tạp chí chuyên ngành có liên quan đến lĩnh vực Hán Nôm như: Tạp chí Văn học, Tạp chí nghiên cứu Lịch sử, Tạp chí Văn hoá dân gian,... tổng cộng 2.895 biểu ghi;

7. TBHN: Quản lý bài trích trong toàn bộ các số Thông báo Hán Nôm xuất bản từ năm 1995 đến nay, bao gồm 877 biểu ghi;

8. HPCD: Quản lý vốn thông tin khoa học nội sinh là các tư liệu điều tra điền dã và hoành biểu đối liên

của cán bộ trong Viện sưu tầm trong nhiều năm qua, với 2.531 biểu ghi;

9. TDDH: Quản lý thông tin về tên tự, tên hiệu của các nhân vật lịch sử và các tác gia Hán Nôm, với 1.277 biểu ghi;

10. STQ: Quản lý vốn tài liệu tiếng Trung Quốc, với 2.351 biểu ghi;

11. BTCTQ: Quản lý các bài trích trong các loại tạp chí xuất bản bằng tiếng Trung Quốc, với 1.990 biểu ghi;

12. HN54: Quản lý vốn tư liệu xuất bản về Hà Nội, với 797 biểu ghi, kết nối với một số hình ảnh Hà Nội xưa;

13. BTC: Quản lý vốn tạp chí các loại, với 117 tên tạp chí;

14. NHO: Quản lý 2.005 biểu ghi các công trình và bài viết về vấn đề Nho giáo ở Việt Nam;

- Thư viện đã phối hợp với Tạp chí Hán Nôm hoàn thành việc đưa Tạp chí Hán Nôm toàn văn lên mạng để phục vụ nhu cầu nghiên cứu và khai thác của bạn đọc trong và ngoài nước trong bối cảnh hội nhập và giao lưu văn hoá.

- Thư viện đang chuẩn bị các bước cần thiết để thực hiện việc quản lý vốn tài liệu và phục vụ độc giả bằng công nghệ mã vạch trong thời gian sớm nhất.

II. Giải pháp lựa chọn phần mềm và xây dựng chương trình chuyển đổi dữ liệu

Các CSDL trên đều được xây dựng dựa trên phần mềm ISIS với chữ Việt là font ABC, chữ Hán là font TwinBridge đã đáp ứng được yêu cầu đảm bảo thông tin và tạo các sản phẩm thông tin tại Viện trong thời gian qua. 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 internet. 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, các phần mềm khác thì còn nhiều bất cập đối với một thư viện chuyên ngành với vốn tài liệu không lớn như Thư viện Viện Nghiên cứu Hán Nôm. Với số biểu ghi như của Viện Nghiên cứu Hán Nôm và giá các phần mềm đang (30.000 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 Thư viện Viện Hán Nôm không thể đáp ứng được. Để giải quyết mâu thuẫn giữa nhu cầu công việc và nguồn kinh phí được cấp, Thư viện đã quyết định sử dụng phần mềm mã nguồn mở Greenstone.

Greenstone là một 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 internet hoặc trên CD-ROM. Greenstone là sản phẩm của dự án New Zealand Digital Library của trường Đại học University of 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 trên <http://greenstone.org> theo thỏa thuận đăng kí GNU 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 UNESCO 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 Nôm còn CSDL Greenstone dùng cấu trúc Dublin Core Metadata và font chữ Unicode. Như 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.

Thư viện 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 Nôm và Trung Quốc thì phức tạp hơn nhiều, vì phải tách riêng phần chữ Việt và phần chữ Hán Nôm hoặc chữ Trung Quốc, sau đó đổi font chữ Việt từ ABC sang Unicode, tiếp đó đổi chữ Hán Nôm 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 các cán bộ Thư viện Viện Nghiên cứu Hán Nôm thực hiện, đã được chuyển sang CSDL Greenstone để đưa lên mạng internet như sau:



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

Biểu ghi ISIS:



Biểu ghi Greenstone tương ứng:



Còn đây là phần minh họa biểu ghi Greenstone chữ Hán:

- Kết quả tìm kiếm thông tin như sau:



Chi tiết một biểu ghi:



Trong xu thế phát triển hiện nay của hoạt động thông tin – thư viện, việc đưa dữ liệu lên internet để chia

sẽ thông tin là việc làm cần thiết và hữu ích đặc biệt đối với các thư viện có vốn tài liệu chuyên ngành quý và đặc thù như Viện Nghiên cứu Hán Nôm. Đa số các thư viện nhỏ hiện đang dùng phần mềm ISIS để quản trị dữ liệu, nhưng nếu muốn đưa dữ liệu lên internet mà dùng winisis 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 từ Thư viện Viện Nghiên cứu Hán Nôm, 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. Khi nháy vào liên kết “Toàn văn”, ta sẽ đọc được toàn văn như trang minh họa dưới đây:



Việc chuyển đổi dữ liệu từ phần mềm ISIS sang phần mềm mã nguồn mở Greenstone và đưa thông tin lên mạng internet là thành công không nhỏ của một Thư viện chuyên ngành – Thư viện Viện Nghiên cứu Hán Nôm. Sự thành công này đã giúp cho các nhà nghiên cứu Hán Nôm và những người có quan tâm đến văn hóa Việt Nam.

Chu Tuyết lan

GD. Thư viện Viện Nghiên cứu Hán Nôm

Nguồn: Tạp chí Thư viện Việt Nam số 2(22) – 2010 (tr.10-14)