

Kinh nghiệm xây dựng Thư viện số với Phần mềm mã nguồn mở GREENSTONE

ĐẶNG ĐỨC NGUYỄN

Chuyên viên Thư viện số

Phòng Công tác Kỹ thuật – Thư viện ĐH Khoa học Tự nhiên

Uớc cách mạng kỹ thuật số đã làm thay đổi toàn diện cuộc sống chúng ta. Các thư viện và các trung tâm thông tin cũng không ngoại lệ. Công nghệ thông tin làm thay đổi hình thức xuất bản như các tài nguyên thông tin được phát hành dưới dạng số rất hấp dẫn, lôi cuốn. Trong sự đổi thay đó, các thư viện cũng như các thủ thư phải thể hiện vai trò quản lý nguồn tài nguyên số. Phần mềm thư viện số mã nguồn mở Greenstone đóng vai trò quan trọng trong việc tổ chức nguồn tài nguyên này, nó hỗ trợ người dùng xây dựng, lưu hành và truy cập các bộ sưu tập số với đầy đủ các chức năng quản lý tài nguyên thư viện thông thường và số hóa. Greenstone là một phần mềm đa ngôn ngữ, hiện nay đã được dịch sang hơn 40 thứ tiếng, và được dùng rộng rãi trên thế giới. Do đó, Thư viện đại học Khoa học Tự nhiên đã chọn Greenstone để xây dựng cho mình một thư viện số hiện đại.

Trong bài viết này, chúng tôi sẽ trình bày ngắn gọn về thư viện số và các tiện ích mà phần mềm Greenstone mang lại. Ngoài ra, chúng tôi sẽ mô tả các tập tin và thư mục có liên quan trong một bộ sưu tập, bàn luận về tiến trình cập nhật các bộ sưu tập sẵn có cũng như tạo mới một bộ sưu tập. Cuối cùng là công việc xây dựng Thư viện số với những công việc cụ thể liên quan đến Greenstone mà Thư viện Đại học Khoa học Tự nhiên đã làm trong thời gian qua.

Greenstone - Tại sao gọi là phần mềm thư viện số

Thư viện số được xem như là nơi trình bày những bộ sưu tập thông tin có tổ chức. Đối tượng của những bộ sưu tập đó là nguồn tài nguyên thông tin số hóa cùng với những phương thức: truy hồi, chọn lọc, truy cập, tổ chức và bảo trì sưu tập đó. Greenstone là một phần mềm mã nguồn mở đáp ứng được công việc đó. Nó tạo các bộ sưu tập hiệu quả trong việc tìm kiếm, thể hiện toàn văn và dễ dàng trình duyệt metadata rất hay và dễ sử dụng. Hơn nữa, các bộ sưu tập dễ dàng được bảo trì.

Dùng Greenstone để xây dựng bộ sưu tập số

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 ngày hay nhiều hơn.

➤ Lưu ý trong khi xây dựng bộ sưu tập, cảnh báo sẽ xuất hiện khi gặp một trong những vấn đề sau:

- Không tồn tại tập tin đưa vào;
- Không có plugin đọc được tập tin hoặc;
- Không tìm thấy các tập tin liên đới – như hình ảnh kèm theo trong tài liệu html.

Các plugin sẵn có của GSDL

Danh sách các plugin gồm nhiều định dạng khác nhau nhằm phù hợp với nhiều tài liệu, như là ZIPPlug, GMLPlug, TEXTPlug, HTMLPlug, EMAILPlug, RecPlug (for "recursive"), MS WordPlug, PDFPlug, v.v..

Tập tin và thư mục

Mỗi bộ sưu tập sẽ chứa trong một thư mục riêng, thư mục đó sẽ gồm có năm thư mục con: *Import* (chứa tài liệu thô), *Archives* (chứa các tập tin tự tạo dạng GML), *Index* (chứa các tập tin cuối cùng phục vụ người dùng), *Building* (thư mục này được dùng trong quá trình xây dựng bộ sưu tập) và một thư mục bổ trợ - *etc*, chứa các tập tin điều khiển quá trình kiến tạo bộ sưu tập. Ngoài ra, có thể có một số tập tin để cấu trúc hệ thống cấp bậc theo yêu cầu phân loại dữ liệu.

Cập nhật sưu tập

Từ khi bộ sưu tập được hình thành, bản thân thư mục của nó đã chứa các thư mục con như *import*, *archives*, *building*,

index, và *etc*; thư mục *etc* sẽ chứa tập tin cấu hình bộ sưu tập phù hợp.

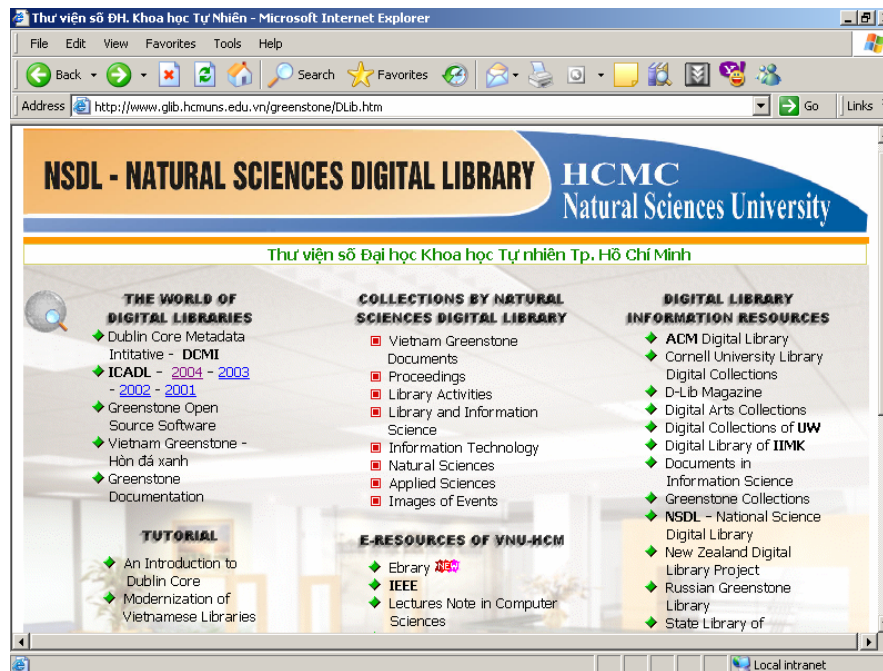
Để cập nhật bộ sưu tập, tài liệu gốc mới được đặt ở thư mục

Import, với bất kỳ hình thức nào sẵn có. Sau đó, tiến trình *import* được thực hiện, nó chuyển các tập tin sang dạng GML

dùng các plugin chỉ định. Các tập tin cũ mà các tập tin GML đã được tạo ra trong lần *build* trước thì sẽ không nhập lại. Kế đến, tiến trình *build* được kích hoạt để xây dựng các chỉ mục cần thiết cho bộ sưu tập. Cuối cùng, nội dung của thư mục *building* được chuyển sang thư mục *index*. Như vậy phiên bản mới của bộ sưu tập đã được hình thành.

Quy trình cập nhật này dường như công kênh. Nhưng tất cả các bước đều cần thiết, thao tác hiệu quả với các bộ sưu tập lớn. Tiến trình *import* có thể thực hiện trơn tru trong suốt quá trình xây dựng – nhưng vì thiết lập các chỉ mục là một thao tác nhiều khê, thường thì việc nhập vào dài dòng có thể lặp lại nhiều lần. Việc xây dựng có thể mất nhiều thời gian, một vài ngày đối với các bộ sưu tập lớn. Do đó, những gì đã có trong thư mục *building* của các bộ sưu tập sẵn có sẽ tiếp tục được dùng lại dạng cũ trong việc xây dựng.

Thư viện số Đại học Khoa Học Tự Nhiên Tp. HCM



Trên cơ sở xây dựng thư viện số từ nền tảng thư viện truyền thống, thư viện Đại học Khoa Học Tự Nhiên đã từng bước xây dựng các bộ sưu tập số từ các tài liệu được

số hóa cũng như các tài liệu số sẵn có bằng nhiều nguồn khác nhau.

Cùng với quá trình xây dựng, bổ túc các bộ sưu tập. Thư viện cũng đã có thêm một trang web riêng về thư viện số – <http://www.glib.hcmuns.edu.vn/greenstone/Dlib.htm>. Đây là trang web tập hợp thông tin liên quan đến thư viện số, các bộ sưu tập thông tin của Thư viện, các nguồn tài nguyên điện tử liên kết khác.

Trang Web Thư viện số ĐH Khoa học Tự nhiên bao gồm:

- **The World of Digital Library – Thế giới Thư viện số:** Giúp bạn đọc và đồng nghiệp tìm hiểu về Thư viện số qua các Web site chuyên ngành như DCMI (Dublin Core Metadata Intitative); ICADL (Hội nghị Thư viện số Châu Á); Greenstone Open Source Software; Hòn đá xanh (Greenstone Việt Nam); và

Greenstone Documentation (bao gồm các bộ sưu tập về văn kiện, bài báo và tài liệu liên quan đến Greenstone).

- **Tutorial – Phụ đạo:** Gồm những bài giảng và tham luận dưới hình thức powerpoint về Thư viện số, Greentone, và Dublin Core Metadata, vv...
- **Digital Library Information Resources – Tài nguyên thông tin Thư viện số:** Tập hợp nhiều nguồn tài nguyên thông tin thư viện số miễn phí từ các đại học và viện nghiên cứu trên thế giới, chẳng hạn như: ACM Digital library; D-Lib Magazines; NSDL – National Science Digital Library; Russian Greenstone Library, vv...
- **Collections by Natural Sciences Digital library – Những Bộ Sưu tập của Thư viện số ĐH KHTN:** Bao gồm những bộ sưu tập do chính Chuyên viên Thư viện số phối hợp với thầy cô trong trường tạo nên bằng công nghệ Greenstone, chẳng hạn như: Sưu tập tài liệu hội nghị; Khoa học Thông tin và Thư viện; Tin học; Khoa học Tự nhiên; Khoa học Ứng dụng; vv... Ngoài những bộ sưu tập khoa học kinh điển, Thư viện cũng luôn kịp thời đưa ra các sưu tập mang tính thời sự như sưu tập Sóng thần – tập hợp các bài báo, hình ảnh liên quan đến sự kiện Sóng thần ngày

26/12/2004; sưu tập các Kỷ yếu Hội nghị – gồm các thông tin, bài viết, tài liệu hội nghị, hội thảo.

- **E-Resources of VNU-HCM – Tài nguyên điện tử của ĐHQG-HCM:** Bao gồm đường kết nối đến những CSDL trực tuyến thương mại do ĐHQG TP. HCM mua quyền sử dụng, chẳng hạn như: Ebrary; IEEE; Lectures Note in Computer Sciences. Hiện nay Những CSDL này chỉ được truy cập vào từ Thư viện trường.

Mục tiêu của Thư viện ĐH Khoa Học Tự Nhiên là xây dựng thư viện số ngày càng hiện đại, nội dung phong phú với các bộ sưu tập về nhiều lĩnh vực khác nhau nhằm đáp ứng được nhu cầu học tập nghiên cứu ngày càng cao của độc giả. Đây là công việc đòi hỏi phải có thời gian, nhân lực và cả sự hỗ trợ từ các thư viện bạn cũng như từ phía độc giả. Do đó, Thư viện cũng đã kêu gọi sự cộng tác từ các thầy cô của các khoa trong trường.

Thư viện Đại học Khoa Học Tự Nhiên luôn chia sẻ, trao đổi kinh nghiệm với các thư viện bạn để mạng lưới thư viện các trường đại học ngày càng phát triển. Ngoài ra, Thư viện cũng thường xuyên mở các lớp tập huấn chuyên đề cho các nhân viên thư viện trong đó có cả chuyên đề “**Thư viện số – Digital Library**”.

TÀI LIỆU THAM KHẢO

1. **B. Maruthu Pandian, Sharad Kuma Sonker and R. Moorthy.** “*Creating Digital Libraries: An experiment with Greenstone Digital Library Open Source Software*”
2. **Ian H. Witten, Rodger J. McNab, Stefan J. Boddie, David Bainbridge.** “*Greenstone: A Comprehensive Open-Source Digital Library Software System*”
3. <http://www.glib.hcmuns.edu.vn/greenstone/Dlib.htm>