

PHẦN MỀM THƯ VIỆN SỐ GREENSTONE VÀ VẤN ĐỀ ỨNG DỤNG TRONG THỰC TIỄN

NGUYỄN THỊ LOAN^(*)

Quá trình ứng dụng công nghệ thông tin trong hoạt động thông tin - thư viện đã có bước phát triển mạnh mẽ từ giữa những năm 1980. Mở đầu quá trình này là việc xây dựng và phát triển phần mềm quản lý tài liệu trong thư viện. Năm 1986 UNESCO đã tạo lập phần mềm quản trị cơ sở dữ liệu (CSDL) CDS/ISIS (Computerized Documentation System/Integrated Set of Information System) để cung cấp cho các nước đang phát triển, trong đó có Việt Nam.

Cùng với sự phát triển của công nghệ thông tin và truyền thông, vào khoảng những năm 1990 cho tới nay, hoạt động thông tin - thư viện đã có những bước phát triển rất rõ rệt. Đồng thời, vấn đề phát triển các phần mềm được sử dụng trong lĩnh vực này cũng rất được chú trọng. Ở nước ta, điều này được phản ánh khá rõ nét, mà biểu hiện là sự xuất hiện của các phần mềm thư viện số, bao gồm cả phần mềm thương mại (Libol của Công ty Tinh Vân, Ilib của Công ty CMC, Vebrary của Công ty Lạc Việt,...) và phần mềm nguồn mở (Greenstone, Zope, Drupal,...). Các phần mềm này đã giúp việc chuyển đổi từ thư viện truyền thống sang thư viện số, cũng như giải quyết các bài toán trong hoạt động thông

tin - thư viện (không chỉ quản lý tài liệu, mà còn quản lý cả việc luân chuyển tài liệu, việc khai thác thông tin...) trở nên thuận tiện hơn. Tuy nhiên, mỗi phần mềm lại có những đặc trưng và tiện ích khác nhau. Bài viết này tập trung vào những nét khái quát về phần mềm Thư viện số Greenstone (Greenstone Digital Library) và khả năng ứng dụng phần mềm này trong thực tiễn hoạt động thông tin thư viện ở Việt Nam.

1. *Greenstone* là sản phẩm của Dự án thư viện số New Zealand tại trường đại học Waikato cùng với sự hợp tác của UNESCO và Human Info NGO. Greenstone là phần mềm mã nguồn mở dùng để xây dựng các bộ sưu tập số. Theo David Wheeler, “Phần mềm mã nguồn mở sẽ cho phép người dùng sử dụng theo bất kỳ mục đích nào; được phép nghiên cứu và sửa đổi, được phép sao chép và tái phát hành phần mềm gốc hoặc phần mềm đã sửa đổi mà không phải trả tiền cho những người lập trình trước”. Greenstone là phần mềm đa ngôn ngữ. Hiện nay, Greenstone được dịch sang 60 ngôn ngữ (3). Trong đó 4 ngôn ngữ nòng cốt là: Anh, Pháp, Tây Ban Nha và Nga. Ở Việt Nam, nếu muốn sử

^(*) Viện Thông tin Khoa học xã hội.

dụng phiên bản tiếng Việt thì cơ quan thông tin-thư viện phải trả phí cho người biên soạn phiên bản tiếng Việt. Ngoài ra, Greenstone cũng cung cấp một phương thức mới để tổ chức và xuất bản thông tin dưới dạng số, kể cả trên Internet cũng như được bao gói trên CD- ROM.

2. Một số ưu điểm của Greenstone

- Truy cập rộng rãi. Bộ sưu tập được truy cập qua một trình duyệt Web chuẩn.

- Đa hệ (có thể sử dụng trên nhiều hệ điều hành). Bộ sưu tập có thể chạy trên Windows, Unix/ Linux và Mac OS-X.

- Hướng siêu dữ liệu. Các chỉ số lướt tìm được tạo nên từ các siêu dữ liệu. Các siêu dữ liệu có thể kết hợp với mỗi tài liệu hoặc một phần của tài liệu.

- Đa ngôn ngữ. Unicode được dùng để hỗ trợ việc chuyển đổi ngôn ngữ. Chỉ mục riêng biệt có thể tạo ra cho những ngôn ngữ khác nhau.

- Thang độ (Scale) lớn. Những bộ sưu tập chứa hàng triệu tài liệu, nhiều Gigabytes có thể được tạo ra. Tìm kiếm toàn văn nhanh chóng. Có thể nén để giảm kích thước văn bản và các chỉ số được gắn kèm.

- Tương thích Z39.50. Giao thức Z39.50 hỗ trợ việc truy cập máy chủ bên ngoài cũng như giới thiệu bộ sưu tập Greenstone cho người sử dụng bên ngoài.

- Tìm kiếm linh hoạt. Người dùng có thể tìm toàn văn tài liệu với những chỉ số liên kết thích hợp.

- Lướt tìm linh hoạt. Ngoài khả năng tìm kiếm theo các thuộc tính đã định, người dùng có thể lướt tìm theo quy định của việc thiết kế như: danh mục tác giả, chủ đề, từ khoá, năm xuất bản của tài liệu,...

- Đa phương tiện. Bên cạnh thông tin dạng văn bản, bộ sưu tập có thể chứa hình ảnh, âm thanh, băng ghi âm, ghi hình.

- Xuất ra CD-ROM. Bộ sưu tập có thể xuất ra CD-ROM một cách tự động.

- Greenstone là một bộ phần mềm mã nguồn mở, do đó người sử dụng dễ dàng chỉnh sửa phù hợp với công việc của mình.

- Một ưu điểm nổi bật của phần mềm mã nguồn mở nói chung và Greenstone nói riêng là không phải trả tiền bản quyền. Đặc biệt, các sản phẩm trên phần mềm nguồn mở có thể được thay đổi theo nhu cầu của người dùng và có tính tương thích cao.

3. Các tính năng của Greenstone

- Có thể sử dụng để xây dựng các bộ sưu tập tài liệu số từ Internet và các CSDL trực tuyến dạng đa phương tiện: các bộ sưu tập thông tin dạng âm thanh, tranh ảnh, hình ảnh động, hoạt hình, đồ họa, toàn văn. Lúc này, thông tin đã tồn tại dưới dạng số - bởi vậy không cần thiết phải tiến hành việc số hoá.

- Có thể sử dụng để xây dựng các sưu tập về các chuyên ngành bằng cách số hoá các tài liệu hiện có tại thư viện: sách, tạp chí, luận văn, báo cáo khoa học, đề tài nghiên cứu khoa học, bài giảng, giáo trình, với sưu tập toàn văn. Những bộ sưu tập số này được xây dựng trên cơ sở nguồn tài liệu dạng truyền thống (in trên giấy) - bởi vậy, cần tiến hành số hoá các tài liệu.

- Có thể sử dụng để xây dựng CSDL thư mục theo các chuẩn Dublin Core^(*)

^(*) Dublin Core là chuẩn chủ yếu dùng cho việc mô tả các tài liệu dạng số hoặc các thư mục của các nguồn thông tin trên Internet.

hay MARC 21. Ngoài ra, phần mềm Greenstone có khả năng hỗ trợ xây dựng sưu tập, biên mục theo Dublin Core và MARC 21 của Greenstone bằng công cụ Librarian Interface.

- Greenstone có thể tích hợp vào phần mềm quản lý thư viện có sẵn.

- Greenstone có thể được phát triển thành một phần mềm quản lý thư viện hoàn chỉnh theo yêu cầu của từng thư viện - tức là có thể được thiết kế để thực hiện các chức năng khác nhau, giải quyết các bài toán trong thư viện.

....

Với những tính năng nêu trên, Greenstone đã được sử dụng để tạo lập và xuất bản các bộ sưu tập trên Web ở nhiều tổ chức của các nước trên thế giới. Ở Việt Nam, một số nơi đã sử dụng Greenstone để xây dựng các bộ sưu tập số của mình. Đầu năm 2004, Thư viện Đại học Khoa học Tự nhiên - Đại học Quốc gia Tp. Hồ Chí Minh đã sử dụng Greenstone xây dựng thư viện số với sự hỗ trợ phiên bản tiếng Việt của Integrated e-Solution, ITD Việt Nam (Ies). Tháng 6/2004, Thư viện trường Đại học Ngân hàng Tp. Hồ Chí Minh cũng đã sử dụng Greenstone. Tháng 10/2004, Trung tâm Thông tin - Tư liệu Đại học Đà Nẵng cũng sử dụng Greenstone để tạo lập các bộ sưu tập số. Hiện nay, các thư viện trường đại học, các trung tâm thông tin cũng đã bắt đầu nghiên cứu và sử dụng Greenstone như: Trung tâm Thông tin KH&CN Quốc gia, Viện Thông tin Khoa học xã hội, Thư viện trường Đại học Hà Nội,...

So sánh các tính năng của CDS/ISIS và Greenstone ở bảng (trang bên) theo một số tiêu chí cụ thể sẽ cho thấy rõ những ưu điểm của Greenstone so với CDS/ISIS. Và như vậy, so với CDS/ISIS,

Greenstone là một công cụ rất hữu ích đối với thư viện số. Ngoài ra, Greenstone còn là một phần mềm được chuẩn hoá cao, điều này giúp chúng ta tiếp cận với các chuẩn thư viện quốc tế, từ đó nâng cao khả năng hợp tác và hội nhập với các thư viện trong nước và quốc tế.

4. Xây dựng bộ sưu tập số dựa trên phần mềm Greenstone

Xây dựng bộ sưu tập số là một nội dung đang thu hút sự quan tâm của các thư viện trên thế giới cũng như ở Việt Nam. Đối với các tệp tài liệu được lưu giữ trên máy tính điện tử thì việc sưu tầm, tổ chức và xử lý dữ liệu đơn giản hơn. Song, đối với nguồn thông tin truyền thống (dạng in trên giấy) thì việc số hóa khá phức tạp và tốn kém, trên thực tế, khó có đủ khả năng để số hóa tất cả các tài liệu đang có trong thư viện. Vì vậy, chúng ta cần phải xác định mức độ cần thiết và độ ưu tiên của một bộ sưu tập để tiến hành việc xây dựng bộ sưu tập số: tài liệu quý hiếm, chỉ có một bản duy nhất; tài liệu viết tay dễ hư hỏng, rách nát; tài liệu có tần suất sử dụng cao; tài liệu lưu hành và phục vụ nhu cầu nội bộ,... Ngoài ra, số hoá một tài liệu cũng là công việc tạo ra cho tài liệu gốc một khả năng phổ biến mới, vì thế tất ảnh hưởng đến vấn đề bản quyền hay quyền sở hữu trí tuệ của tài liệu được số hoá. Đó là một thực tế mà quá trình xây dựng bộ sưu tập số cần phải tính đến, đặc biệt khi số hoá những tài liệu chưa được xuất bản, như: luận án khoa học, báo cáo kết quả nghiên cứu, tư liệu dịch chưa xuất bản... Đây là những vấn đề phức tạp và đòi hỏi phải được giải quyết một cách có hệ thống từ khía cạnh chính sách cụ thể, vì vậy bài viết không đề cập và phân tích vấn đề này tại đây.

STT	Tiêu chí	CDS/ISIS	Greenstone
1	Hỗ trợ tiếng Việt	- Hỗ trợ ASCII 16 bit (TCVN 5712 - 1999) - Cần phát triển bổ sung các chức năng cho phép hiển thị thanh công cụ bằng tiếng Việt - Để truy cập Internet cần có máy chủ Webasis, một máy chủ Proxy đặc biệt để thực hiện chuyển đổi từ mã ASCII sang Unicode (TCVN 6909 -2001) và ngược lại	- Hỗ trợ Unicode (TCVN 6909 - 2001) - Có sẵn giao diện tiếng Việt - Không cần chuyển đổi bộ mã ký tự, Greenstone mặc định sử dụng Unicode - Dữ liệu bên trong Greenstone lưu trữ theo UTF-8
2	Hỗ trợ xuất bản CSDL trên Web	CDS/ISIS tự thân không có tính năng cài đặt chạy trên Internet mà cần phải có công cụ đặc biệt (Webasis)	Có thể cài đặt chạy trên Internet
3	Quản lý CSDL	- Số lượng tối đa các biểu ghi trong 1 CSDL là 16 triệu (dưới 500 MB) - Một biểu ghi CDS/ISIS có tối đa 200 trường, mỗi trường chứa tối đa 32.000 ký tự - Không có sẵn khả năng quản lý toàn văn	- Không giới hạn số lượng ký tự trong một biểu ghi - Có khả năng quản lý toàn văn, tạo chỉ mục toàn văn. Do đó có thể tiến hành tìm kiếm toàn văn
4	Tạo cấu trúc CSDL	Người dùng phải tạo Bảng định nghĩa các trường (FDT) để định nghĩa cấu trúc (các yếu tố mô tả của biểu ghi thư mục)	- Có sẵn một số bộ siêu dữ liệu để người dùng lựa chọn (ví dụ: Dublin Core với 15 yếu tố mô tả)
5	Tìm kiếm thông tin	- Cho phép sử dụng các toán tử như: toán tử Bool, tìm lân cận, chặt cụt - Hai hình thức tìm kiếm: Tìm nâng cao và tìm đơn giản có hướng dẫn - Đối với Webasis cần tạo mẫu tìm kiếm, điều này không đơn giản đối với những người không phải là lập trình viên	- Khả năng tìm kiếm mạnh, sử dụng toán tử Bool và ký hiệu chặt cụt - Hai hình thức tìm kiếm là tìm đơn giản và tìm theo mẫu - Dễ dàng tạo mẫu tìm kiếm sử dụng các công cụ sẵn có
6	Hiển thị và xem CSDL trên màn hình	- Người dùng có thể xem lần lượt các biểu ghi hoặc xem bất kỳ một biểu ghi nào - Không có chức năng duyệt (xem) theo hệ thống thứ bậc hoặc theo chủ đề	- Greenstone sử dụng trình duyệt Web (Ví dụ: Internet Explorer, Netscape) để duyệt bộ sưu tập - Cho phép duyệt bộ sưu tập theo các cách khác nhau (Sắp xếp theo danh sách, theo chủ đề, theo thời gian,...)
7	Lưu và in ấn kết quả tìm kiếm	- In trực tiếp ra bất kỳ máy in nào có cài trong Windows - Có thể in trực tiếp trong chế độ xem hoặc chuyển kết quả sang Word để in	Sử dụng tính năng lưu và in ấn kết quả mặc định của trình duyệt Internet.

Để tiến hành xây dựng bộ sưu tập số với Greenstone chúng ta cần thực hiện các bước sau:

1. Xác định chủ đề nội dung của bộ sưu tập. Chủ đề này được phản ánh một cách hàm súc qua tên của bộ sưu tập, và được dùng để phân biệt giữa các bộ sưu tập khác nhau.

2. Xác định nguồn tài liệu cần được bao quát để đưa vào bộ sưu tập. Nguồn tài liệu cụ thể gồm: các tài liệu in trên giấy hoặc các vật mang tin khác, các tệp dữ liệu được lưu trữ trên máy tính điện tử, các tài liệu được lấy về từ Web hay PDF.

3. Xác định cấu hình của bộ sưu tập. Đây là một công việc đóng vai trò đặc biệt quan trọng vì nó quyết định bộ sưu tập có hoạt động được hay không. Trong tập tin cấu hình, tài liệu được tự động chuyển sang định dạng XML^(*), đồng thời các Plugin^(**) thích hợp sẽ được chọn trong quá trình tạo lập CSDL. Sau đó các chỉ số tìm kiếm và cấu trúc trình duyệt tài liệu được quy định trong tập tin cấu hình được tạo ra.

Đến đây, quá trình xây dựng bộ sưu tập số đó hoàn thành và có thể tiến hành tìm kiếm thông tin theo các dấu hiệu khác nhau được quy định trong cấu hình của bộ sưu tập. Tuy nhiên, để xây dựng bộ sưu tập số thì vấn đề tạo siêu dữ liệu là một công việc hết sức cần thiết. Đối với Greenstone, việc xác định siêu dữ liệu được thực hiện theo chuẩn *Dublin Core*.

Chuẩn Dublin Core bao gồm 15 yếu tố mô tả: Nhan đề, Đề mục, Mô tả, Loại

hình, Nguồn gốc, Liên kết, Nơi chứa, Tác giả, Tác giả phụ, Thông tin về xuất bản, Bản quyền, Ngày tháng, Mô tả vật lý, Địa danh, Ngôn ngữ. Dublin Core có ưu điểm là: Đơn giản trong tạo lập và bảo trì (được thiết kế nhằm phục vụ những người không chuyên), dễ sử dụng, rẻ và mang lại hiệu quả cao; Ngữ nghĩa thông dụng (khắc phục những khó khăn trong việc hiển thị các thuật ngữ, ví dụ: yếu tố <tác giả> - Creator - được gán cho người tạo lập ra tài liệu như: tác giả tài liệu, nhà soạn nhạc, đạo diễn,...) ; Khả năng mở rộng và đa ngôn ngữ.

Chuẩn Dublin Core giúp ta xác định siêu dữ liệu theo 3 cách:

Thứ nhất: Xây dựng thủ công bằng công cụ Organizer;

Thứ hai: Xây dựng bán tự động bằng công cụ Librarian Interface;

Thứ ba: Xây dựng tự động bằng công cụ Greenstone Collector.

Cách thứ nhất đòi hỏi người xây dựng phải có khả năng lập trình tốt. Cách thứ ba thường áp dụng với các tài liệu đơn giản và kết quả tạo ra các bộ sưu tập với giao diện định trước. Cách thứ hai là cách làm phù hợp với khả năng hiện nay của chúng ta. Bởi vì giao diện Librarian Interface trình bày 15 yếu tố của Dublin Core giúp cho việc tạo lập những điểm truy cập theo nhan đề, tác giả, từ khóa, chủ đề...

5. Tìm kiếm thông tin số dựa trên phần mềm Greenstone

Greenstone có giao diện đơn giản và thân thiện với màu xanh lá cây đặc trưng; ngôn ngữ giao tiếp mặc định là tiếng Anh. Người dùng có thể chuyển đổi giao diện sang ngôn ngữ mình cần bằng cách: Từ cửa sổ chính của Greenstone → nhấn chuột trái vào biểu tượng *thuộc tính* → lựa chọn ngôn ngữ

(*) XML là một định dạng tệp dữ liệu tổng hợp và chứa đựng nhiều tính năng, cho phép thu nhỏ kích thước tệp dữ liệu, phục hồi các tệp dữ liệu đã bị hỏng, đồng thời tăng cường độ an toàn.

(**) Plugin là công cụ dùng để mở rộng các chức năng của WordPress.

→ Enter. Người dùng sẽ được hướng dẫn chi tiết hơn nếu nhấn vào biểu tượng giúp đỡ.

Từ cửa sổ tìm kiếm của Greenstone, chúng ta có thể tiến hành tìm kiếm thông tin theo các cách khác nhau như: Nhập từ khoá vào hộp tìm kiếm hoặc tìm theo kiểu duyệt lướt. Tùy thuộc vào cấu trúc CSDL do người tạo lập thiết kế, người dùng có thể tiến hành tìm kiếm theo các dấu hiệu khác nhau như: nhan đề, tác giả, từ khóa, tên tác giả, chủ đề,...

Sau khi tìm kiếm, kết quả được hiển thị dưới dạng thư mục. Để đọc được toàn văn, người dùng tin có thể kích vào biểu tượng file đính kèm để kết nối tới toàn bộ văn bản dưới dạng PDF hay HTML.

Ngày nay, vấn đề xây dựng thư viện điện tử/thư viện số đang diễn ra rất mạnh mẽ. Các thành tựu của khoa học, công nghệ hiện đại không ngừng nâng cao khả năng trong các quá trình lưu trữ, tổ chức và truy cập thông tin. Xu thế hội nhập đã tạo ra những cơ hội, đồng thời cũng đặt ra không ít thách thức cho ngành thông tin - thư viện. Chính xu thế đó và nhu cầu thông tin của xã hội nói chung đòi hỏi cần phải có những đổi mới trong hoạt động của các thư viện. Việc xây dựng các bộ sưu tập số tại các thư viện là một động thái tích cực nhằm bảo quản lâu dài nguồn tài liệu, đáp ứng một cách hiệu quả nhu cầu tin của người dùng tin. Để thực hiện mục tiêu này thì việc sử dụng phần mềm thư viện số Greenstone sẽ là một giải pháp thiết thực góp phần phát triển hoạt động thông tin thư viện theo hướng hiện đại hoá.

TÀI LIỆU THAM KHẢO

1. Allison Zhang and Don Gourley. A Digital Collections Management

System Based On Open Source Software. Washington Research Library Consortium Poster/Demo Presentation at JCDL 2003 (28/5/2003).

2. About Greenstone.
<http://greenstone.org>
3. Greenstone language support.
http://greenstone.sourceforge.net/wiki/index.php/Greenstone_language_support
4. Ian H. Witten, David Baibridge, Stefan J. Boddie. Greenstone Open - SourceDigital Library Software. *D-Lib Magazine*, October 2001, Volume 7, Number 10.
www.dlib.org
5. Diane Hillmann. Using Dublincore.
<http://uk.dublincore.org/documents/usageguide/#whatis>
6. Cao Minh Kiểm, Đào Mạnh Thắng. Lựa chọn phần mềm và khổ mẫu dữ liệu số phục vụ nông thôn miền núi (Kỷ yếu Hội nghị thư viện số châu Á lần thứ 10). H., 2007, tr. 106-118.
7. Kỷ yếu Hội thảo khoa học: Tăng cường công tác tiêu chuẩn hoá trong hoạt động thông tin - tư liệu. H.: Trung tâm Thông tin KH&CN quốc gia, 2006.
8. Kỷ yếu Hội thảo khoa học: Tiếp cận xây dựng thư viện số ở Việt Nam - Hiện trạng và vấn đề. H.: Trung tâm Thông tin KH&CN Quốc Gia, 2007, 171 tr.
9. Nguyễn Minh Hiệp. Sử dụng phần mềm nguồn mở thư viện số Greenstone để xây dựng kho tài nguyên học tập (Kỷ yếu Hội thảo Xây dựng và phát triển nguồn học liệu để phục vụ đào tạo và nghiên cứu, tr.93-103).
10. Sử dụng Greenstone để xây dựng Bộ sưu tập thư viện số. *Bản tin Thư viện - Công nghệ thông tin*, số 1/2006, tr. 22-33.