

DSPACE, GIẢI PHÁP PHẦN MỀM CHO THƯ VIỆN ĐIỆN TỬ QUẢN LÝ VÀ KHAI THÁC NGUỒN THÔNG TIN SỐ NỘI SINH Ở CÁC TRƯỜNG ĐẠI HỌC HIỆN NAY

(Bài đăng trong Tạp chí Thư viện Việt Nam số 1-2015)

*PGS.TS. Đoàn Phan Tân
Trường Đại học Văn hóa Hà Nội*

Yêu cầu quản lý các nguồn thông tin số nội sinh ở các trường đại học và sự ra đời của các phần mềm quản lý bộ sưu tập số

Hoạt động đào tạo và nghiên cứu khoa học ở các trường đại học tự nó đã tạo ra một khối lượng tài liệu đặc biệt có giá trị. Đó là các giáo trình, các tập bài giảng của giảng viên, các khóa luận tốt nghiệp của sinh viên, các luận văn thạc sĩ của học viên, các luận án tiến sĩ của nghiên cứu sinh, các công trình nghiên cứu khoa học của giảng viên và sinh viên đăng tải trên các tạp chí khoa học, các kỷ yếu hội nghị khoa học, v.v... Chúng được gọi chung là nguồn tài liệu nội sinh và thông tin mà các nguồn tài liệu này cung cấp được gọi là nguồn thông tin nội sinh.

Nguồn thông tin nội sinh của trường đại học là yếu tố phản ánh đầy đủ và hệ thống các thành tựu và tiềm năng khoa học của một trường đại học và có vai trò quan trọng trong công tác đào tạo của nhà trường. Các nguồn thông tin này ngày càng phong phú, đa dạng và luôn chứa đựng những thông tin mới nhất trong lĩnh vực mà nó xem xét, rất cần được quản lý và khai thác một cách hiệu quả.

Một thuận lợi là các nguồn thông tin này thường được lưu trữ dưới dạng tệp văn bản, tức là đã được số hóa và về cơ bản nhà trường có quyền sử dụng, không gặp nhiều khó khăn trong vấn đề bản quyền.

Ở nước ta, từ đầu những năm 2000, nhiều trường đại học ở nước đã tăng cường ứng dụng công nghệ thông tin để hiện đại hóa thư viện theo hướng xây dựng thư viện của mình theo mô hình thư viện điện tử. Các thư viện điện tử này trước hết có chức năng quản lý và khai thác nguồn thông tin số nội sinh toàn văn của trường.

Phần mềm được lựa chọn là các hệ quản trị thư viện tích hợp, như: LIBOL của Cty Tinh Vân, ILIB của Cty CMC hay Virtua của Cty VTLIS Hoa Kỳ. Tuy nhiên qua thực tế ứng dụng, các phần mềm này còn bộc lộ nhiều hạn chế, nhất là trong khâu quản lý và khai thác nguồn thông tin số toàn văn.

Trong thực tế các phần mềm này mới chỉ chú ý tới quản lý các CSDL thư mục, với việc áp dụng chuẩn biên mục đọc máy MARC21 và thực hiện tìm tin trên OPAC, tức là tìm các thông tin thư mục trên mạng. Và như vậy thư viện chưa thể trở thành thư viện điện tử đúng nghĩa. Bởi vì thư viện chỉ trở thành thư viện điện

từ khi người đọc có thể đọc trực tiếp toàn văn một bộ phận quan trọng tài liệu của thư viện trên máy tính.

Yêu cầu quản lý và khai thác các nguồn thông tin số đang phát triển bùng nổ, mang tính tính toàn cầu trong những năm gần đây, cùng với những yêu cầu phát triển của thư viện điện tử, mà các bộ sưu tập số toàn văn là thành phần cốt lõi, đã dẫn đến sự ra đời và phát triển các *phần mềm quản lý bộ sưu tập số*, trong đó đáng kể nhất là hai phần mềm Greenstone và DSpace.

Greenstone, có tên đầy đủ là GREENSTONE DIGITAL LIBRARY (GSDL), là phần mềm mã nguồn mở hỗ trợ việc xây dựng và phân phối các bộ sưu tập số của thư viện trên Internet hoặc trên CD-ROM. Các bộ sưu tập số có thể bao gồm các tài liệu số dưới nhiều dạng thức khác nhau: văn bản, âm thanh, hình ảnh tĩnh và động. Những tài liệu không thuộc dạng chữ được liên kết với các tài liệu dạng chữ hoặc được mô tả dạng chữ (ví dụ: chú thích cho các ảnh) để hỗ trợ việc tìm kiếm theo nội dung.

Greenstone là kết quả của dự án thư viện số tại trường đại học Waikato, NewZealand (NewZealand Digital Library Project) với sự hợp tác của hai tổ chức UNESCO và Human Info NGO.

Phiên bản Greenstone đầu tiên phát hành vào tháng 8 năm 2000. Mục đích của phần mềm Greenstone là trao quyền cho người sử dụng, đặc biệt là thư viện các trường đại học để xây dựng thư viện số cho riêng mình và chia sẻ nguồn lực thông tin trong cộng đồng. Greenstone được sử dụng khá rộng rãi trong các trường đại học và nhiều tổ chức nghiên cứu trên thế giới.

DSpace là phần mềm mã nguồn mở hỗ trợ giải pháp xây dựng và phân phối các bộ sưu tập số trên Internet. DSpace do thư viện của học viện công nghệ Massachusetts (Massachusetts Institute of Technology Libraries – MIT Libraries) và phòng thí nghiệm của Hewlett-Packard (HP Labs) phát triển. Phiên bản DSpace đầu tiên phát hành vào tháng 11 năm 2002, với chức năng ban đầu là đáp ứng yêu cầu quản lý các kết quả nghiên cứu, các tài liệu giảng dạy và học tập đã số hóa của MIT.

DSpace cung cấp một công cụ hoàn chỉnh để quản lý các tài liệu khoa học của MIT trong một kho lưu trữ kỹ thuật số chuyên nghiệp, luôn được duy trì và tái tạo, dễ dàng truy cập và hiển thị tại bất cứ thời điểm nào.

Tháng 7 năm 2007, do cộng đồng người sử dụng DSpace ngày càng mở rộng MIT và HP Labs thành lập ra DSpace Foundation để lãnh đạo và hỗ trợ sự phát triển của DSpace. Ngày nay DSpace và cộng đồng người dùng nhận sự lãnh đạo và hướng dẫn từ DuraSpace (Một tổ chức phi lợi nhuận ra đời tháng 5 năm 2009, kết quả của sự hợp tác của DSpace Foundation và tổ chức Common)

Trong hai phần mềm nói trên, phần mềm DSpace tỏ ra có nhiều ưu điểm nổi trội. Hiện nay trên thế giới đã có hơn 1100 trường đại học, thư viện và các tổ

chức sử dụng DSpace để quản lý, khai thác và chia sẻ nguồn tài nguyên thông tin số của mình.

Tại Việt Nam, trong những năm gần đây đã có nhiều đơn vị ứng dụng DSpace để xây dựng thư viện điện tử, quản lý và khai thác các nguồn thông tin số dưới dạng các bộ sưu tập số cho thư viện của mình. Trong các đơn vị ứng dụng DSpace thành công có thể kể: Trung tâm Thông tin - Thư viện Đại học Đà Lạt, Thư viện Đại học Quốc gia Hà Nội, Thư viện Đại học Quốc gia Tp. Hồ Chí Minh, Thư viện Đại học Công nghiệp Tp. Hồ Chí Minh, Trung tâm Thông tin - Thư viện Đại học Nguyễn Tất Thành, Thư viện Đại học Lạc Hồng, Thư viện Đại học Nội vụ Hà Nội, v.v.

Ở trường Đại học Văn hóa Hà Nội, trang thư viện số của trường được xây dựng trên cơ sở ứng dụng DSpace đã khai trương ngày 26 tháng 3 năm 2014, nhân kỷ niệm 55 năm ngày thành lập trường. Các bộ sưu tập số được quản lý và khai thác ở đây là các nguồn thông tin số nội sinh toàn văn của trường, bao gồm:

- Bộ sưu tập các bài giảng,
- Bộ sưu tập các bài tạp chí,
- Bộ sưu tập các khóa luận tốt nghiệp, và
- Bộ sưu tập các luận văn thạc sĩ, tiến sĩ.



The image shows a screenshot of the Hanoi University of Culture Digital Library website. The header is blue and yellow, featuring the university's logo on the left and the text "THƯ VIỆN SỐ TRƯỜNG ĐẠI HỌC VĂN HÓA HÀ NỘI" in large yellow and white letters. Below this, it says "Hanoi University of Culture Digital Library". On the right side of the header, there is contact information: "ĐC: 418 Đường La Thành, Đống Đa, Hà Nội" and "Tel: (84-4) 38511971 Fax: (84-4) 385141629". Below the header is a search bar with the text "Tìm kiếm" and a "Thực hiện" button. Below the search bar, there is a "Tìm kiếm nâng cao" link and a "Trang chủ" link. Below the search bar, there is a "Duyệt theo" section with a "Bộ sưu tập" link. The main content area is white and features a large grey box with the text "CHÀO MỪNG 55 NĂM NGÀY THÀNH LẬP TRƯỜNG ĐẠI HỌC VĂN HÓA HÀ NỘI" and "26 Tháng 3 Năm 1959 - 26 Tháng 3 Năm 2014". On the right side of the main content area, there is a "Liên kết thư viện số" section with links to "Trường Đại học Bách khoa Hà Nội" and "Trường Đại học Quốc gia".

Qua thực tiễn ứng dụng, DSpace tỏ ra là giải pháp hiệu quả cho việc xây dựng thư viện điện tử, quản lý và khai thác nguồn thông tin số toàn văn nội sinh ở các trường đại học nước ta hiện nay.

Nghiên cứu những đặc trưng tính năng và công nghệ của DSpace là một công việc cần thiết và có ý nghĩa. Những tính năng nổi trội của DSpace sẽ góp phần làm sáng tỏ và mở rộng những kết quả đã đạt được trong thực tiễn ứng

dụng DSpace để quản lý và khai thác các nguồn thông tin số nội sinh ở các trường đại học nước ta hiện nay.

Những đặc trưng tính năng và công nghệ của DSpace

Phần mềm quản lý bộ sưu tập số DSpace có những đặc trưng tính năng và công nghệ sau:

(1) DSpace là phần mềm quản lý bộ sưu tập số mã nguồn mở, giúp người sử dụng có thể tự xây dựng các bộ sưu tập số cho thư viện của mình.

Các bộ sưu tập số được xây dựng riêng lẻ, thông qua sự giống nhau nổi bật của các tài liệu, thường xuyên được duy trì, được cập nhật bổ sung và tự động tái tạo. Các tài liệu đưa vào bộ sưu tập có thể chọn từ máy tính hay tải về từ Internet.

Để tạo thuận lợi cho người sử dụng, mô hình hệ thống thông tin trong DSpace được xây dựng trên ý tưởng: một hệ thống thông tin bao gồm nhiều đơn vị thành viên. Có thể tạo nhiều đơn vị cùng cấp. Mỗi đơn vị thành viên lại có nhu cầu riêng trong việc tổ chức thông tin trong các bộ sưu tập. Như vậy các bộ sưu tập phải được tạo ra bên trong một đơn vị. Mỗi bộ sưu tập quản lý một loại tài liệu số cụ thể của đơn vị đó.

Ví dụ: Trong thư viện số của một trường đại học có các đơn vị: KHOA, PHÒNG BAN, TRUNG TÂM.

Trong đơn vị KHOA chẳng hạn lại có các bộ sưu tập: GIÁO TRÌNH – BÀI GIẢNG, BÀI TẬP CHÍ, LUẬN VĂN, ...

Với cấu trúc như trên, DSpace cho phép xây dựng các bộ sưu tập theo cấu trúc nhiều cấp, giúp việc tổ chức các bộ sưu tập khoa học hơn so với Greenstone.

Các bộ sưu tập số trong DSpace có khả năng lưu trữ thông tin với dung lượng lớn, hàng vạn, hàng triệu tài liệu.

(2) DSpace có khả năng xử lý các tài liệu đa phương tiện với nhiều định dạng tệp tin khác nhau, trong đó có các định dạng tài liệu văn bản (doc, txt, rtf, pdf, html, xml,...), định dạng tài liệu về hình ảnh (gif, jpg,...), định dạng các tài liệu âm thanh (wav, flv, mp3, mp4,...)

Với khả năng nhận biết được 64 định dạng tệp tin, có thể nói DSpace tương thích với hầu như tất cả các định dạng tệp tin, từ các định dạng đối với tài liệu văn bản đến các định dạng đối với tài liệu âm thanh, hình ảnh, video ca nhạc,....

(3) DSpace sử dụng sơ đồ siêu dữ liệu Dublin Core Metadata để mô tả tài liệu trong các bộ sưu tập.

Dublin Core Metadata là một trong những sơ đồ siêu dữ liệu phổ biến được hình thành lần đầu tiên vào năm 1995. Tập hợp các yếu tố siêu dữ liệu này được coi là cốt lõi (core) vì nó được thiết kế đơn giản và chỉ bao gồm 15 trường cốt lõi nhất (trong khi MARC21 gồm hơn 200 trường và rất nhiều trường con).

Dublin Core Metadata đơn giản trong tạo lập và bảo trì, được thiết kế phục vụ những người không chuyên, dễ sử dụng nhưng mang lại hiệu quả lớn.

Mục đích đầu tiên và yêu cầu cốt yếu nhất của siêu dữ liệu là góp phần mô tả và tìm lại các tài liệu điện tử trên mạng Internet, vốn là những tài liệu khó xác định loại hình và nội dung các yếu tố cần thể hiện.

Khi mô tả tài liệu trong DSpace, ba trường bắt buộc là: nhan đề, tác giả và năm xuất bản. Các yếu tố khác như nhà xuất bản, tóm tắt, từ khóa chủ đề, ngôn ngữ, loại hình tài liệu,... là tùy chọn, phụ thuộc vào tài liệu khi cập nhật. Trong DSpace các trường này được hiển thị trong biểu ghi thư mục của tài liệu và được định chỉ mục, phục vụ cho việc duyệt xem thông tin và tìm tin.

(4) Toàn bộ các thao tác như biên mục, bổ sung, duyệt xem và tìm kiếm tài liệu, quản trị hệ thống, v.v... trong DSpace đều được thực hiện trên nền giao diện web (web-based interface). Có giao diện dành cho người nhập tài liệu vào hệ thống, có giao diện dành cho người dùng tin để tìm kiếm thông tin, có giao diện dành cho người quản trị hệ thống.

Giao diện dành cho người nhập tài liệu vào hệ thống giúp việc biên mục và bổ sung tài liệu vào các bộ sưu tập dễ dàng. Khi cần bổ sung tài liệu vào các bộ sưu tập không cần phải xây dựng lại từ đầu như Greenstone.

Vì DSpace là phần mềm mã nguồn mở nên giao diện dành cho người quản trị hệ thống cho phép cải tiến cách trình bày và mở rộng khả năng ứng dụng của phần mềm. Ví dụ, có thể tạo các đường link để kết nối với các thư viện điện tử của các trường đại học khác.

Đối với tất cả các bộ sưu tập, DSpace đều cung cấp một giao diện đồng nhất cho phép người dùng tin có thể duyệt xem và tìm kiếm dễ dàng các tài liệu. Có thể duyệt xem và tìm trong một bộ sưu tập. Cũng có thể duyệt xem và tìm trong một đơn vị, tức là trong nhiều bộ sưu tập của đơn vị. Nhưng khi đó diện tìm sẽ rộng hơn.

Kết quả tìm kiếm sẽ hiển thị đầu tiên dưới dạng một danh sách các tài liệu tìm được, với ba cột thông tin là năm xuất bản, nhan đề và tác giả. Thông tin đầy đủ về mỗi tài liệu được hiển thị dưới dạng một biểu ghi thư mục. Để xem nội dung của tài liệu, cần tải tài liệu về và sử dụng một phương tiện thích hợp với dạng thức của tài liệu.

Ví dụ:

Muốn hiển thị một văn bản lưu ở dạng PDF (Adobe's Portable Document Format) thì cần sử dụng phần mềm Adobe Acrobat hoặc Nitro PDF Professional

Muốn xem một bộ sưu tập ảnh thì cần tích hợp chương trình Windows Picture and Fax Viewer.

Muốn xem một clip video ca nhạc phải sử dụng phần mềm KMPlayer hay Windows Media Player.

(5) DSpace sử dụng Unicode là bộ phông chữ chuẩn quốc tế để trình bày nội dung tài liệu. Unicode được dùng để hỗ trợ chuyển đổi ngôn ngữ. Có thể tạo chỉ mục cho các ngôn ngữ khác nhau.

DSpace sử dụng bộ phông chữ tiếng Việt của Unicode trong giao diện cũng như trong sử dụng.

(6) DSpace khai thác thông tin trong các bộ sưu tập theo hai hình thức: duyệt xem thông tin và tìm tin.

DSpace có khả năng duyệt xem thông tin trong các bộ sưu tập theo bốn dấu hiệu: tác giả, nhan đề, chủ đề và năm xuất bản. Trong mỗi dấu hiệu lại có thể duyệt xem theo vần chữ cái.

Ví dụ: Duyệt xem thông tin theo nhan đề trong bộ sưu tập BÀI TẬP CHÍ của thư viện số của trường Đại học Văn hóa Hà Nội, bắt đầu bằng chữ E, ta có kết quả sau:

The screenshot shows the DSpace search results page. The search criteria are: 'Duyệt "BÀI TẬP CHÍ" theo Tên tài liệu' and 'Chuyển đến: 0-9 A B C D E F G H I J K L M N O P Q R S T U V W X Y Z'. The results are sorted by 'Tên tài liệu' and show 57 results out of 228. The table below lists the results:

Ngày ban hành	Tên tài liệu	Tác giả
2014	Entropy và Thông tin	Đoàn, Phan Tân
2010	Gia đình- từ cách tiếp cận văn hóa	Nguyễn, Hồng Mai
2010	Gia đình và hôn nhân của người Raqlai - Nhìn từ góc độ luật tục	Nguyễn, Thị Thanh Vân
2010	Gia đình Việt Nam trong cơn bão của thời đại	Nguyễn, Hồng Mai
2010	Giải mã văn hóa địa danh rồng bay	Dương, Văn Sáu

DSpace tìm kiếm thông tin theo các trường đã được chỉ mục. Có khả năng tìm kiếm toàn văn theo từng từ (từ khóa, từ chuẩn) và đặc biệt có thể tìm theo một cụm từ trong văn bản.

Ví dụ: Trong bộ sưu tập BÀI TẬP CHÍ của thư viện số Đại học Văn hóa Hà Nội, tìm bài viết mà trong nhan đề có cụm từ “đổi mới toàn diện” ta có kết quả sau:

The screenshot shows a search interface with a search bar containing 'BÀI TẬP CHỈ'. Below the search bar, there are options for 'Kết quả/Trang' (10), 'Sắp xếp tài liệu theo' (Mức độ liên quan), 'Theo trật tự' (Giảm dần), 'Tác giả/biểu ghi' (Tất cả), and a 'Cập nhật' button. The search results are displayed in a table with columns for 'Ngày ban hành', 'Tên tài liệu', and 'Tác giả'.

Ngày ban hành	Tên tài liệu	Tác giả
1999	40 năm Đại học Văn hóa Hà Nội - Khắc phục khó khăn, đổi mới toàn diện, vững bước đi lên	Đoàn, Phan Tân

Có hai phương thức tìm tin là:

- Tìm đơn giản, là tìm theo một từ khóa, từ chuẩn hay theo một cụm từ trong văn bản.
- Tìm tin nâng cao, là tìm với biểu thức tìm được thiết lập bằng cách liên kết các thuật ngữ tìm trong các trường bằng các toán tử logic AND, OR, NOT.

(7) Với khả năng phân quyền mạnh, DSpace cho phép phân quyền đến từng tài khoản người dùng và từng bộ sưu tập. Các quyền được cấu hình khá chi tiết như quyền cập nhật các tài liệu vào bộ sưu tập, quyền xem biểu ghi thư mục, quyền xem toàn văn v.v. Phần mềm Greenstone không cho phép thực hiện các thao tác này.

(8) Về mặt công nghệ, DSpace là một tập hợp các hợp tác ứng dụng của Java web và các chương trình tiện ích nhằm duy trì một kho siêu dữ liệu của nguồn thông tin số. DSpace là phần mềm mã nguồn mở, tất cả các mã gốc đều được viết bằng ngôn ngữ lập trình Java.

Các siêu dữ liệu về tài liệu được lưu trữ trong các CSDL được xây dựng theo mô hình quan hệ và được quản lý bởi phần mềm quản trị CSDL PostgreSQL.

Các ứng dụng web sử dụng phần mềm Web server và Java servlet engine (Apache và Tomcat, cả hai đều từ Apache Foundation) cung cấp giao diện để tạo lập, quản lý, lưu trữ và tìm kiếm các nguồn thông tin số. DSpace hiện hỗ trợ hai giao diện web chính là JSPUI (Java Server Page User's Interface) và XMLUI (eXtension Markup Language User's Interface). Tất cả các phần mềm này đều là mã nguồn mở. Nhiều phiên bản DSpace gần đây cũng hỗ trợ tìm kiếm và duyệt chức năng nhờ sử dụng server đánh chỉ mục Apache Solr.

(9) DSpace vận hành trong môi trường Internet với giao diện Web, đáp ứng các yêu cầu công nghệ sau:

- Tuân theo các chuẩn công nghệ về truyền thông của mạng Internet theo mô hình Client/Server,
- Sử dụng trình duyệt Internet Explorer hoặc FireFox.
- Phần mềm có thể được cài đặt và hoạt động trên một trong số các hệ điều hành phổ biến như Unix, Windows XP, Windows 7.

Vì DSpace là một tập hợp các hợp tác ứng dụng của Java web và các chương trình tiện ích, nên việc cài đặt DSpace phải qua nhiều công đoạn.

Để để cài đặt DSpace trên môi trường Windows:

- Trước hết ta phải cài đặt các phần mềm cần thiết.
- Tiếp theo tạo CSDL trong PostgreSQL cho DSpace.
- Sau đó mới tiến hành cài đặt DSpace vào thư mục đã được cấu hình từ trước.

Các phần mềm cần thiết là:

1) **Java SDK 6** hoặc phiên bản mới hơn, chứa các công cụ giúp sử dụng tốt nhất các chương trình, bảo đảm có những chương trình cần thiết hỗ trợ cho máy tính.

2) **ProstgreSQL 8.x for Windows**, là hệ quản trị CSDL theo mô hình quan hệ, được phát triển bởi khoa điện toán của trường đại học Berkeley bang California, có lịch sử phát triển trên 15 năm, có nhiều ưu điểm nổi trội, được tạp chí Linux Journal Editors đánh giá là DBMS tốt nhất hiện nay.

3) **Apache Maven 2.x**, là phần mềm quản lý việc xây dựng và lập tài liệu cho dự án, dựa trên khái niệm Project Object Model.

4) **Apache Ant 1.7.x**, là công cụ hỗ trợ đặc biệt cho lập trình bằng Java, dùng định dạng XML làm cơ chế hoạt động cho công cụ dưới dạng các lệnh (instructions)

5) **Apache Tomcat 6.x**, là một máy chủ web, mã nguồn mở, được phát triển bởi Apache Software Foundation (ASF).

Ngoài ra để chương trình có thể chạy với giao diện tiếng Việt cần file giao diện tiếng Việt **Messages_vi.properties**.

Nhiều lớp huấn luyện cài đặt và sử dụng phần mềm DSpace được mở ra trong mấy năm gần đây đã góp phần quảng bá và thúc đẩy việc sử dụng phần mềm DSpace. Nhiều thư viện các trường đại học đã sử dụng DSpace xây dựng các thư viện điện tử, quản lý và khai thác có hiệu quả các nguồn thông tin số nội sinh của mình.

Đó là tín hiệu mới của quá trình hiện đại hóa thư viện các trường đại học ở nước ta hiện nay.

TÀI LIỆU THAM KHẢO

- 1) DSpace, An Open Source Dynamic Digital Repository . D-Lib Magazine, January 2003, Volume 9 Number 1, ISSN 1082-9873
- 2) About DSpace. Truy cập từ website <http://www.DSpace.org/introducing>
- 3) DSpace. From Wikipedia, the free Encyclopedia. Truy cập từ website <http://www.en.wikipedia.org/wiki/DSpace>
- 4) DSpace on Windows. Truy cập từ website <http://wiki.durapace.org/DSPACE/DSpaceOnWindows>