GIẢI PHÁP TỔNG HỢP NGUỒN TÀI LIỆU KHOA HỌC CỦA ĐẠI HỌC ĐÀ NẴNG

SOLUTIONS INTEGRATED SCIENTIFIC LITERATURE RESOURCES OF THE UNIVERSITY OF DA NANG

Hồ Phan Hiếu, Trần Thanh Liêm

Văn phòng Đại học Đà Nẵng; Email:hophanhieu@ac.udn.vn, ttliem@ac.udn.vn

Tóm tắt -Đại học Đà Nẵng (ĐHĐN) đang phát triển theo định hướng đại học nghiên cứu nên việc công bố các thông tin khoa học rộng rãi trên website là điều rất cần thiết. Hiện nay, các nguồn tài liệu của các cơ sở giáo dục thành viên của ĐHĐN là rất lớn nhưng chưa tập trung và quy chuẩn; việc lưu trữ, thu thập, quản lý và phân phối các tài liệu khoa học, các ấn phẩm học thuật ở dạng số hóa còn rất ít và chưa đóng góp nhiều trong học tập, nghiên cứu của cán bộ và sinh viên.

Trong bài báo này, chúng tôi trình bày kết quả nghiên cứu thực trạng và đề xuất giải pháp để tổng hợp một nguồn tài liệu khoa học lớn bằng cách triển khai ứng dụng hệ thống DSpace trong toàn ĐHĐN. Nguồn tài liệu này sẽ được chúng tôi cập nhật vào kho dữ liệu để phục vụ hệ thống phát hiện sao chép. Bên cạnh đó, giải pháp chúng tôi triển khai đã góp phần nâng cao thứ hạng tên miền website ĐHĐN trên Webometrics và các hệ thống xếp hạng khác.

Từ khóa –tài liệu khoa học; kho dữ liệu; Webometrics; DSpace; tên miền; thứ hạng

Abstract- The University of Danang (UD) is developing to orient research university so the publication of scientific information on website is essential. Currently, the resources of the members of UD are great but have not focused and non standard; storage, collection, management and distribution of scientific literature, in digital form is very limited and don't contribute much in the academic, research of staff and students.

In this article, we present the results of the experimental research and suggest the solutions to synthesize large scientific literaturere sources by deploying DS pace applications system at the members of UD. This resources will be updated in our data warehouse to serve the detection copy system. Besides, the solution we implement to help improve rankings UD's domain name on Webometrics and other rating systems.

Key words –scientific literature; data warehorse; Webometrics; DSpace; domain name; rating

1. Đặt vấn đề

ĐHĐN là một đại học vùng trọng điểm quốc gia, hiện có 6 cơ sở giáo dục đại học thành viên và nhiều đơn vị trực thuộc, đang đào tạo khoảng 200 nghiên cứu sinh, 3.400 học viên cao học và hơn 78.000 sinh viên đại học và cao đẳng. Mỗi năm có hơn 20.000 học viên, sinh viên tốt nghiệp, với khóa luận tốt nghiệp, đồ án tốt nghiệp hằng năm khoảng 20.000 báo cáo. Bên cạnh đó, Tạp chí Khoa học và Công nghệ ĐHĐN đến nay đã xuất bản gần 100 quyển với hơn 2.000 bài báo, mỗi năm gần đây trung bình có khoảng hơn 300 bài báo được đăng.

Với sứ mệnh của mình, ĐHĐN đang dần định hướng để tiến tới đại học nghiên cứu trong những năm đến nên vấn đề nâng cao chất lượng đào tạo và nghiên cứu khoa học là nhiệm vụ hàng đầu, trong đó việc công bố các thông tin khoa học rộng rãi trên website là điều rất cần thiết. Tuy nhiên, việc lưu trữ, thu thập, quản lý và phân phối các tài liệu khoa học, các ấn phẩm học thuật ở dạng số hóa còn ít và chưa đóng góp nhiều trong học tập, nghiên cứu của cán bộ giảng viên và sinh viên ĐHĐN.Với nguồn tài liệu khoa học được công bố hằng năm rất lớn thì việc lưu trữ, tổng hợp để giúp cho sinh viên, cán bộ giảng viên dễ dàng truy cập, tìm kiếm, tra cứu… là việc cấp thiết.

Hệ thống DSpace được thiết kế để lưu trữ, bảo quản và cung cấp các tài liệu nghiên cứu học thuật đã được đa số các tổ chức nghiên cứu khoa học, các trường đại học, các thư viện sử dụng. Vì thế, việc nghiên cứu giải pháp và triển khai xây dựng các hệ thống DSpace để quản lý, tổng hợp nguồn tài liệu khoa học cho ĐHĐN là phù hợp.

Hiện nay, việc xếp hạng website các trường đại học đã trở nên phổ biến và là động lực để các trường cập nhật thông tin, công bố các công trình nghiên cứu khoa học, các tài liệu học thuật… nhằm làm tăng vị thế của trường đại học đối với cộng đồng khoa học trong và ngoài nước. Song song với việc tổng hợp tài liệu của ĐHĐN, chúng tôi tiến hành thực hiện hệ thống hóa tên miền của các website để có thể giúp tăng "Tần số xuất hiện" của websitevà việc xây dựng,tổng hợp kho học liệu có chất lượng sẽ giúp nâng cao chất lượng và thu hút lượng truy cập, làm tăng chỉ số "Độ phong phú tài liệu" và "Tính chất học thuật"… giúp tăng thứ hạng website ĐHĐN trên các bảng xếp hạng của thế giới, đặc biệt là trên hệ thống Webometrics.

Ngoài ra, nguồn tài liệu này sẽ được chúng tôi cập nhật vào kho dữ liệu để phục vụ cho hệ thống phát hiện sao chép nhằm mục đích hạn chế tình trạng "đạo văn" đang diễn ra khá phổ biến, nhất là trong các trường đại học ở Việt Nam.

Thực trạng tài liệu của ĐHĐN hiện nay là rất lớn nhưng không tập trung và chưa theo quy chuẩn. Vì vậy, triển khai Dspace kết hợp quy hoạch hệ thống tên miền trong toàn ĐHĐN sẽ tổng hợp được nguồn tài liệu khoa học lớn, giúp tăng thứ hạng website ĐHĐN trên hệ thống Webometrics. Bên cạnh đó, nguồn tài liệu tổng hợp này sẽ là dữ liệu rất cần thiết để phục vụ cho hệ thống phát hiện sao chép văn bản góp phần nâng cao chất lượng đào tạo và nghiên cứu khoa học của ĐHĐN.

2. DSpace

DSpace là một bộ phần mềm mã nguồn mở, cung cấp giải pháp lưu trữ, quản lý các bộ sưu tập số hóa trên internet. DSpace do HP và Viện Công nghệ Massachusetts (The MIT Libraries) phát triển vào năm 2002. Đến 4/2015, phần mềm đã ra đời phiên bản 5.1 với nhiều cải tiến nhằm đáp ứng mạnh mẽ các nhu cầu thiết yếu của các thư viện. Hiện nay, trên thế giới đã có hơn 1700 đơn vị sử dụng, trong đó đa số là các trường đại học, cao đẳng và các viện nghiên cứu.

Theo [4, 5], DSpace được sử dụng cơ bản như một phần mềm lưu trữ và phân phối tài liệu số với ba vai trò chính:

- Giúp cho việc thu nhận và quản lý tài liệu được dễ dàng, bao gồm siêu dữ liệu của tài liệu.

- Giúp cho việc truy cập tài liệu được dễ dàng, bằng cả việc liệt kê và tìm kiếm.

- Giúp cho việc bảo quản tài liệu lâu dài.

Những điểm nổi bật của phần mềm DSpace:

- Là một phần mềm mã nguồn mở, miễn phí, có một cộng đồng lớn người sử dụng và phát triển trên toàn thế giới

- Dễ dàng tùy chỉnh để phù hợp với nhu cầu của từng đơn vị; giao diện phần mềm thân thiệt trong môi trường web nên dễ dàng trong việc truy cập.

- Sử dụng được trên nhiều hệ điều hành như Windows, Linux, Unix,...

- Sử dụng hệ quản trị cơ sở dữ liệu độc lập nên đáp ứng tốt với thư viện có số lượng tài liệu lớn. Có thể dùng hệ quản trị cơ sở dữ liệu như Postgres SQL hoặc Oracle…

- Có thể quản lý và lưu giữ tất cả các loại tài liệu kỹ thuật số. Tài liệu được biên mục theo chuẩn DublinCoreMetadata.

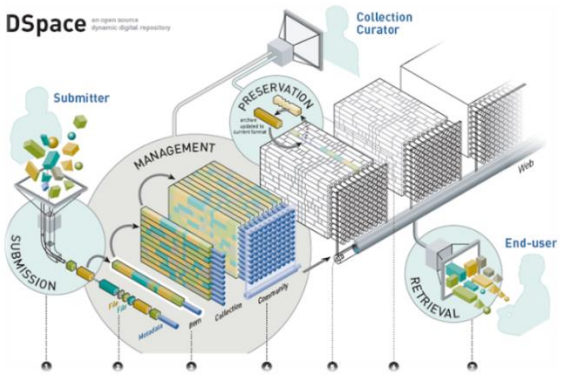
- Cung cấp cơ chế tìm kiếm toàn văn đối với các dạng tài liệu như: PDF, Word, Excel, Powerpoint, text, HTML,…

- Phân quyền và bảo mật mạnh. Có thể phân quyền đến từng tài khoản người dùng, đến từng Bộ sưu tập hoặc thậm chí đến từng tài liệu. Các quyền được cấu hình khá chi tiết như: Quyền xem biểu ghi thư mục, quyền xem toàn văn...

- Hỗ trợ nhiều kiểu báo cáo: Lượt truy cập, lượt xem biểu ghi thư mục, lượt tải tài liệu...

- Hỗ trợ đa ngôn ngữ, trong đó có tiếng Việt.

Cấu trúc DSpace: Có cấu trúc phân cấp theo dạng cây thư mục với các đơn vị và bộ sưu tập. Đơn vị dùng để quản lý các đơn vị con và các bộ sưu tập còn bộ sưu tập để quản lý tài liệu. Đồng thời với việc phân cấp các đơn vị và bộ sưu tập, DSpace còn cho phép phân cấp quản lý đối với các đơn vị và bộ sưu tập này.



Hình 1. *Mô hình phần mềm DSpace*

Với bộ phần mềm nguồn mở của DSpace, chúng tôi đã tiến hành Việt hóa và chỉnh sửa phần mềm thành bộ sản phẩm DSpace để có thể triển khai ở các đơn vị trong ĐHĐN.

Hình 2. *Giao diện DSpace đã đượcViệt hóa và chỉnh sửa*

3. Webometrics và các tiêu chí xếphạng

3.1. Giới thiệu

Hiện nay, đã có đến 15 hệ thống xếp hạng quốc tế, chưa kể các hệ thống xếp hạng quốc gia. Dưới đây là 5 hệ thống được xem là phổ biến và có ý nghĩa nhất đối với xếp hạng các trường đại học Việt Nam [6], đó là:

- *ARWU*, còn được biết dưới tên gọi Bảng xếp hạng Giao thông Thượng Hải (Trung Quốc), ra đời năm 2003.

- *QS World*, trước đây thường được gọi là THES hoặc THE-QS, ra đời năm 2004 tại Anh Quốc.

- *Webometrics*, chuyên xếp hạng trang web của các trường đại học, ra đời năm 2004 tại Tây Ban Nha.

- *QS Asia*, dành riêng cho khu vực châu Á dựa trên hệ thống xếp hạng toàn cầu của QS, ra đời năm 2009.

- *THE*, tách ra từ nhóm THE-QS và kết hợp với ThomsonReuters thành một hệ thống riêng, mới ra đời năm 2010 tại Anh Quốc.

Trong bài báo này, chúng tôi sẽ đề cập đến hệ thống xếp hạng Webometrics, đây là một trong những hệ thống xếp hạng website có độ bao phủ lớn trên thế giới nói chung và Việt Nam nói riêng. Kết quả xếp hạng Webometrics dựa trên website nhưng vẫn phản ảnh được phần nào chất lượng của một trường đại học, đặc biệt là thành tựu về nghiên cứu khoa học như: số lượng tài liệu khoa học được công bố, các trích dẫn tài liệu khoa học,...

Webometrics- "Bảng xếp hạng trang web các trường

đại học trên toàn thế giới" (Webometrics Ranking of World Universities) là một trong những bảng xếp hạng có uy tín hàng đầu về đánh giá, xếp hạng website các trường đại học trên toàn thế giới. Bảng xếp hạng Webometrics được công bố vào năm 2004 và được cập nhật định kỳ mỗi năm hai lần vào tháng 1 và tháng 7. Với mục tiêu "Khuyến khích và thúc đẩy việc đăng tải thông tin trên mạng internet" và hỗ trợ các sáng kiến "Tiếp cận mở" (Open Access Initiatives) nhằm nâng cao khả năng tiếp cận các thông tin về nghiên cứu khoa học và tài liệu học thuật xuất bản dưới dạng điện tử. Webometrics thể hiện sự "Đo lường trang web" của các trường đại học, theo đó những trường có trang web tốt (xét theo chỉ số tác động đối với cộng đồng), cung cấp dồi dào các thông tin khoa học sẽ có vị trí cao [7].

3.2. Các tiêu chí đánh giá xếp hạng

Trên thế giới, có gần 20.000 tổ chức giáo dục đại học được Webometrics đo lường và xếp hạng dựa trên các công cụ tìm kiếm tự động như: Google, Yahoo, LiveSearch...được công bố tại địa chỉ website *http://www.webometrics.info/*căn cứ vào bốn tiêu chí [2]:

-Kích thước (**PRESENCE – S**): Số lượng trang được truy hồi từ bốn công cụ tìm kiếm là Google, Yahoo, Live Search và Exalead.

- Mức độ hiện diện (**IMPACT– V**): Tổng số các đường dẫn ngoại biên đến một trang web (inlinks) được truy hồi từ công cụ tìm kiếm Yahoo Search.

- Độ phong phú tài liệu (rich files) (**OPENNESS– R**): Tổng số các tập tin liên quan đến các hoạt động học thuật công bố dưới các định dạng Adobe Acrobat (.pdf), Adobe PostScript (.ps), Microsoft Word (.doc) và Microsoft Powerpoint (.ppt).

- Tính chất học thuật (**EXCELLENCE– Sc**): Tổng số các tài liệu khoa học và các trích dẫn đối với các tên miền mang tính học thuật được truy xuất bằng công cụ tìm kiếm Google Scholar.

Đơn vị phân tích là các tên miền website (web domain) của các trường đại học, học viện. Vì vậy, chỉ có các trường đại học, học viện với một webdomain độc lập mới được xếp hạng. Nếu một đơn vị có nhiều hơn một tên miền chính, thì hai hoặc nhiều tên miền kia sẽ được sử dụng như là các địa chỉ độc lập. Hệ thống xếp hạng Webometrics sử dụng các máy tìm kiếm là phương tiện trung gian chính để tra cứu, đánh giá. Sự hiện diện của một tên miền trong cơ sở dữ liệu của chúng là một chỉ số về mức độ hiện diện (chỉ số V). Thông thườnghệ thống sử dụng các máy tìm kiếm như: Google, YahooSearch, Live, Exalead và Alexa.

Chuẩn hóa các số liệu thu về từ máy tìm kiếm theo công thức sau:

$$N_a = \frac{\log(n_a + 1)}{\log(\max(n_i) + 1)}$$

Trong đó:

+ N: Máy tìm kiếm (Google, Yahoo, Live, Exalead)

+ a: Tên miền website

Sau đó xác định các chỉ số:

+ Size (S):$S_a = 1/2 * ((G_a + Y_a + L_a + E_a) - \max(G_a, Y_a, L_a, E_a) - \min(G_a, Y_a, L_a, E_a))$

+ Rich file (R): $R_a = PDF_a + DOCa + PPT_a + Ps_a$

Xác định trọng số cho các chỉ số (Bảng 1).

Bảng 1. *Trọng số các tiêu chí xếp hạng Webometrics*

Chỉ số	Trọng số
Kích thước – S	20%
Mức độ hiện diện – V	50%
Độ phong phú tài liệu – R	15%
Tính chất học thuật – Sc	15%

Kết quả xếp hạng được xác định như sau:

Ra →Rank(R)

Sa →Rank(S)

Va →Rank(V)

Sc_a→Rank(Sc)

Rank (vị trí) = 5 * V + 2 * S + 1,5 * R + 1,5 * Sc

Theo [9], dưới đây là ví dụ về cách tính các chỉ số và đo độ xếp hạng như sau:

Chỉ số nhận diện (Visibility – V)

Xác định chỉ số V với 2 máy tìm kiếm: Yahoo và Alta vista. Kết quả lấy về sẽ chấp nhận với việc các trang từ chính domain đó trỏ tới.

Bảng 2. *Các câu truy vấn trong xác định chỉ số V*

Yahoo	link:udn.vn,link:www.udn.vn
AltaVista	linkdomain:udn.vn,linkdomain:www.udn.vn

Chỉ số V được xác định theo công thức:

V = 1/2 (V_yahoo | V_altavista)

Chỉ số kích thước (Size – S)

Chỉ số đại diện cho kích cỡ website của trường đại học. Mô hình chung để xác định chỉ số S cũng tương tự việc xác định chỉ số V dựa trên ý tưởng tận dụng khả năng của máy tìm kiếm. Trong đó, chúng ta xác định một tập các câu truy vấn cho từng máy tìm kiếm: Google, Yahoo, Alta vista. Ví dụ Google: site:udn.vn, Yahoo: site:udn.vn,… Chúng ta nhận kết quả từ máy tìm kiếm trả về để đưa ra giá trị cuối cùng cho chỉ số S theo công thức:

S = 1/3 (S_google + S_yahoo + S_altavista)

Bảng 3. *Câu truy vấn xác định chỉ số S*

Yahoo	site:udn.vn,link:www.udn.vn
Google	site:udn.vn,link:www.udn.vn
AltaVista	domain:udn.vn,domain:www.udn.vn

Chỉ số phong phú tài liệu (Rich files – R)

Chỉ số đại diện cho kết quả nghiên cứu, tài liệu học tập được các trường công bố trên internet.Chỉ số được xác định dựa trên máy tìm kiếm Google. Trong đó các kiểu file được đưa ra là: .pdf, .doc,.ppt.

Với các câu truy vấn như bảng 4.Ví dụ: filetype:.doc +

site:udn.vn. Chỉ số R sẽ là tổng hợp kết quả của từng loại file mà máy tìm kiếm Google trả về, theo công thức sau:

R = DOC + PPT + PDF

Bảng 4. *Câu truy vấn xác định chỉ số R*

Google	filetype:doc + (site:udn.vn,site:www.udn.vn)
	filetype:pdf + (site:udn.vn,site:www.udn.vn)
	filetype:ppt + (site:udn.vn,site:www.udn.vn)

Chỉ số bài báo khoa học (Scholar – Sc)

Chỉ số này cùng với chỉ số R đại diện cho kết quả nghiên cứu của từng trường. Chỉ số Sc được xác định dựa trên Google Scholar. Do không có khả năng lấy về chính xác chỉ số Sc từ cơ sở dữ liệu của Google Scholar, vì vậy dựa vào http://www.scholar.google.com cùng tập câu truy vấn thích hợp để lấy kết quả trả về.Để truy vấn tên miền ĐHĐN, ta dùng câu truy vấn: site:udn.vn

Kết quả xếp hạng được xác định như sau:

Ra = DOCa + PDFa + PPTa

Sa = 1/3 (Ga + Ya + Aa)

Va = 1/2 (Ya + Aa)

Sca

Theo Bảng 1 về trọng số các tiêu chí xếp hạng, tatính được vị trí xếp hạng theo công thức:

Rank (vị trí) = 5 * V + 2 * S + 1,5 * R + 1,5 * Sc

4. Giải pháp và kết quả triển khai

4.1. Quy hoạch tên miền

Để phù hợp với mô hình đại học vùng và tăng chỉ số về kích thước website ĐHĐN, chúng tôi đã tiến hành đồng bộ hóa tên miền website thống nhất trong toàn ĐHĐN. Với ĐHĐN có tên miền cấp 2; các cơ sở giáo dục đại học thành viên, các đơn vị trực thuộc ĐHĐN có tên miền cấp 3; các đơn vị trực thuộc của các cơ sở giáo dục đại học thành viên (khoa, phòng,…) có tên miền cấp 4…

Tên miền website ĐHĐN và các cơ sở giáo dục đại học thành viên, các đơn vị trực thuộc(Bảng 5).

Bảng 5. *Hệ thống tên miền website toàn ĐHĐN*

STT	Đơn vị	Tên miền
1	Đạihọc Đà Nẵng	http://www.udn.vn
2	TrườngĐạihọcBách khoa	http://dut.udn.vn
3	TrườngĐạihọc Kinh tế	http://due.udn.vn
4	TrườngĐạihọc Sư phạm	http://ued.udn.vn
5	TrườngĐạihọcNgoại ngữ	http://ufl.udn.vn
6	Trường Cao đẳng Công nghệ	http://dct.udn.vn
7	Trường Cao đẳngCNTT	http://cit.udn.vn
8	Phân hiệu ĐHĐN tạiKontum	http://kontum.udn.vn
9	ViệnNC&ĐTViệt - Anh	http://vnuk.udn.vn
10	Khoa Y Dược	http://smp.udn.vn
11	Khoa Đàotạoquốc tế	http://sie.udn.vn
12	Khoa Giáodục thể chất	http://ttgdtc.udn.vn
13	Trung tâm PT Phần mềm	http://sdc.udn.vn
14	Trung tâm TT Học liệu	http://lirc.udn.vn
15	Trung tâm ĐT Thường xuyên	http://cce.udn.vn

4.2. Triển khai hệ thốngDSpace

Như đề cập ở phần trên, chúngtôi đã Việt hóa và chỉnh sửa bộ phần mềm nguồn mở DSpace thành bộ phần mềm DSpace của ĐHĐN. Sau đó, triển khai cài đặt cho các cơ sở giáo dục đại học thành viên và đơn vị trực thuộc.Và để tập trung nguồn dữ liệu này, chúng tôi xây dựng trang tài nguyên số để tổng hợp tất cả các hệ thống DSpace của ĐHĐN như một cổng thông tin tổng hợp nguồn tài liệu khoa học.

Cổng thông tin về tài nguyên số tại địa chỉ: http://tainguyenso.udn.vn

Hình 3. *Trang tài nguyên số của ĐHĐN*

Sau thời gian ngắn triển khai Dspace tại các cơ sở giáo dục đại học thành viên, hiện đã có khoảng 31.752 tài liệu được cập nhật (Bảng 6). Các hệ thống Dspace hiện đang được sử dụngtạiđịa chỉ:

- Trường ĐH Kinh tế: http://data.due.udn.vn:8080/dspace/ hoặc http://thuvien.due.udn.vn:8080/dspace/
- Trường ĐH Sư phạm:http://thuvien.ued.udn.vn:8080/dspace/
- Trường ĐH Ngoại ngữ: http://data.ufl.udn.vn/
- Trường CĐ Công nghệ: http://data.dct.udn.vn:8080/
- Trường CĐ CNTT: http://data.cit.udn.vn/
- Trung tâm TTHL:http://tailieuso.udn.vn/

Bảng 6. *Số lượng tài liệu trên cácDSpace của ĐHĐN*

Thể loại	Số lượng	Tỉ lệ
Báo, tạp chí	12283	38,68%
Đề tài, công trình khoa học	667	2,10%
Giáotrình	176	0,55%
Luận văn, luậnánđạihọc	7591	23,91%
Tàiliệunướcngoài	169	0,53%
Tàiliệu tham khảo	7747	24,40%
Luận văn thạc sĩ	3093	9,74%
Luậnántiến sĩ	26	0,09%

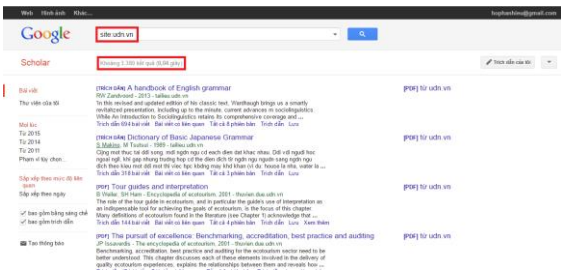
Các hệ thống Dspace tại Trường ĐH Kinh tế, Trường ĐH Sư phạm, Trường CĐ Công nghệ và Trung tâm Thông tin – Học liệu đã được tập huấn và đang được sử dụng. Các đơn vị còn lại chúng tôi sẽ tiếp tục triển khai trong thời gian tới. Chúng tôi dự kiến số lượng tài liệu sau khi hoàn tất và sử dụng DSpace là khoảng 100.000 tàiliệu. Hằng năm sẽ được bổ sung thêm khoảng 20.000 tài liệu.

4.3. Xếp hạng website ĐHĐN

4.3.1. Số lượng file tài liệu và tài liệu khoa học

Truy cập vào trang Google Scholar tại địa chỉ: https://scholar.google.com/để tìm tài liệu khoa học trên website ĐHĐN, nhập câu truy vấn: site:udn.vn

Số lượng tài liệu khoa học của ĐHĐN trên Google Scholar là 3.380 (tháng 6/2015).



Hình 4. *Số lượng tài liệu khoa học vào tháng 6/2015*

Truy cập vào trang Google để tìm tài liệu file PDF trên website ĐHĐN, nhập câu truy vấn:

filetype:pdfsite:udn.vn

Số lượng tài liệu file PDF của ĐHĐN trên Google là 24.500 (tháng 6/2015). Tương tự ta có thể tìm các loại file khác theo cú pháp câu truy vấn trên.

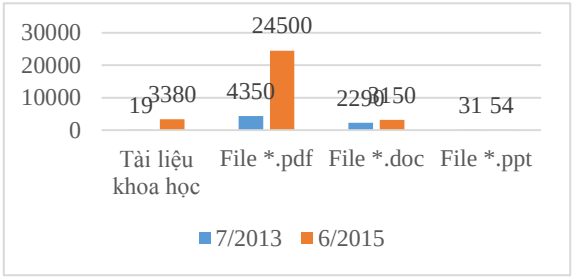
Hình 5. *Số lượng tài liệu file *.pdf vào tháng 6/2015*

Kết quả thực tế sau khi triển khai giải pháp quy hoạch tên miền và hệ thống DSpacecho thấy số lượng file tài liệu và tài liệu khoa học của ĐHĐN tăng lên rất nhiều. Bảng 7 và Hình 6 sẽ cho ta thấy được kết quả.

Bảng 7. *So sánh số lượng file và tài liệu khoa học của tên miền udn.vn*

Thời gian	Tài liệu khoa học	File *.pdf	File *.doc	File *.ppt
7/2013	19	4350	2290	31
6/2015	3380	24500	3150	54

Lượng file **.pdf, *.doc, *.ppt* vào tháng 7/2013 lần lượt là 4.350, 2.290, 31 đã tăng lên 24.500, 3.150, 54 vào tháng 6/2015. Như vậy, số lượng **.pdf, *.doc, *.ppt* đã tăng tương ứng là 20.150, 860, 23 file. Đặc biệt, số lượng tài liệu khoa học đã tăng lên đáng kể 3.380 vào tháng 6/2015 so với 19 vào tháng 7/2013.



Hình 6. *So sánh số lượng tài liệu và tài liệu khoa học*

Xếp hạng trên Webometrics

Trước đây, tên miền website ĐHĐN có thứ hạng thấp. Kết quả tháng 7/2012, website ĐHĐN xếp hạng 3.172 thế giới- World Rank (Hình 8) và không nằm trong Top 100 của Khu vực Đông Nam Á(Hình 7).

CONTINENT RANK	UNIVERSITY	COUNTRY	RANK WORLD RANK	RANK PRESENCE	RANK IMPACT	RANK OPENNESS	RANK EXCELLENCE
37	Vietnam National University Hanoi		1,050	500	1,756	954	2,168
39	Can Tho University		1,103	1,904	739	1,318	2,659
76	Hanoi University of Science & Technology		1,807	858	2,909	3,236	2,259
82	Ho Chi Minh City University of Economics		1,924	3,118	1,750	5,410	1,784
83	Water Resources University		1,950	6,229	240	4,559	4,603
85	Ho Chi Minh City University of Agriculture and Forestry (Nong Lam University)		1,999	2,156	2,595	2,844	2,787
100	Ho Chi Minh City University of Technology		2,228	1,438	2,047	1,994	5,228

Hình 7. *Thứ hạng website các trường ở Việt Nam trong khu vực Đông Nam Á (tháng 7/2012)*

WORLD RANK	UNIVERSITY	COUNTRY	RANK Presence	Impact	Openness	Excellence
2745	Hue University		5,998	3,379	2,894	3,317
3172	(1) University of Da Nang		3,171	4,412	2,539	4,603
4176	(1) Duy Tan University		1,204	6,530	3,854	5,228
5081	Nha Trang University		3,144	7,660	5,863	3,818
7934	Danang University of Technology		7,120	5,799	9,811	5,228
8843	Quy Nhon University		6,359	10,677	9,811	3,401
11674	College of Economics Da Nang University		10,077	11,167	9,811	5,228
12475	Tay Nguyen University		11,377	11,885	9,811	5,228

Hình 8. *Bảng so sánh WorldRank giữa website ĐHĐN và các trường trong khu vực miền Trung - Tây Nguyên (tháng 7/2012)*

Trong quá trình thử nghiệm và triển khai giải pháp đồng bộ tên miền và DSpace, thứ hạng tên miền website ĐHĐN đã tăng rất đáng kể. Hiện nay, theo bảng xếp hạng mới nhất của Webometrics (công bố vào tháng 1/2015), ĐHĐN đang xếp thứ 4 cả nước, có mặt trong Top 100 các trường của Đông Nam Á (xếp thứ 68) và World Rank là 2.222. Như vậy, so với tháng 7/2012, tên miền udn.vn đã tăng 950 bậc trong bảng xếp hạng thế giới của Webometrics.

60	2021	Yogyakarta State University		1251	1976	148	5414
61	2060	Petra Christian University		1080	3510	194	4158
62	2076	De La Salle University Manila		1319	3170	2775	2697
63	2120	Universiti Tunku Abdul Rahman		503	7246	1494	1894
64	2128	Universitas Riau		1312	3711	464	3826
65	2146	Assumption University of Thailand		1042	1839	1464	4808
66	2154	Hanoi National University of Education		4397	1272	5698	3084
67	2214	Universiti Teknikal Malaysia Melaka		478	7169	664	2630
68	2222	University of Da Nang		2077	1372	2297	4808
69	2257	Maejo University		735	3058	3142	3424
70	2268	(3) University of the Philippines Los Baños		3980	3754	1478	2891
71	2278	Rajamangala University of Technology Thanyaburi		633	2061	1115	5414
72	2281	Payap University		867	830	4497	5414

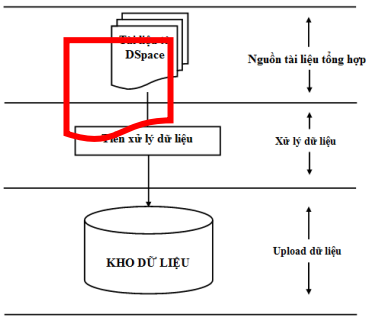
Hình 9. *Thứ hạng website các trường ở Việt Nam trong khu vực Đông Nam Á (tháng 1/2015)*

Vietnam

ranking	World Rank	University	Det.	Presence Rank*	Impact Rank*	Openness Rank*	Excellence Rank*
1	894	Vietnam National University Hanoi		551	1684	149	1822
2	1790	Can Tho University		1796	1142	4111	3272
3	2154	Hanoi National University of Education		4397	1272	5698	3084
4	2222	University of Da Nang		2077	1372	2297	4808
5	2781	Ho Chi Minh City University of Agriculture and Forestry (Nong Lam University)		3629	2611	5231	3500
6	3008	(1) Duy Tan University		1438	2263	3714	5414
7	3253	Hue University		3773	5801	1620	3980
8	3279	Ho Chi Minh City University of Technology		2202	2336	4424	5414
9	3527	International University National University HCM City		5081	3968	1409	5414
10	3536	Vietnam National University Ho Chi Minh City		7660	4844	10780	1735
11	3628	Hanoi University of Science & Technology		2629	8311	6371	2252

Hình 10. *Tên miền website ĐHĐN xếp thứ 4 trong cả nước(tháng 1/2015)*

4.4. Tổng hợp nguồn tài liệu khoa học



Hình 11. *Qui trình xử lý, upload tài liệu vào kho dữ liệu*

Với kết quả triển khai các hệ thống DSpace, chúng tôi đã tổng hợp được nguồn tài liệu lớn của ĐHĐN phục vụ cho việc học tập, nghiên cứu và hằng năm số lượng cập nhật dữ liệu sẽ tăng nhiều. Bên cạnh đó, chúng tôi lấy nguồn dữ liệu này để đưa vào kho dữ liệu (hệ thống này đã được xây dựng)phục vụ hệ thống phát hiện sao chép mà nhóm nghiên cứu đang thực hiện [3].

Dự kiến trong năm nay có khoảng 100.00 tài liệu được tổng hợp và mỗi năm ước tính khoảng 20.000 tài liệu được bổ sung sẽ là nguồn dữ liệu dồi dào, phong phú.

Chúng tôi đang tiến hành xây dựng công cụ để upload các tài liệu này vào kho dữ liệu để phục vụ hệ thống phát hiện sao chép đang nghiên cứu, góp phần hạn chế vấn nạn "đạo văn" hiện nay.

5. Kết luận

Về mặt khoa học, bài báo đã nghiên cứu đề ra giải pháp để tổng hợp nguồn tài liệu khoa học của ĐHĐN bằng cách xây dựng cổng thông tin tài nguyên số; ứng dụng, triển khai hệ thống DSapce; nghiên cứu các tiêu chí xếp hạng, đề ra giải pháp triển khai giúp tăng thứ hạng website ĐHĐN trên hệ thống Webometrics.

Về mặt ứng dụng, các giải pháp triển khai đã được ứng dụng thực tế, bước đầu nâng cao đáng kể thứ hạng website ĐHĐN trên hệ thống Webometrics. Hệ thống tên miền của tất cả các website trong toàn ĐHĐN được thống nhất theo quy định giúp cho việc truy cập được thuận lợi và thể hiện được vai trò phân cấp của mô hình đại học vùng. Số lượng file pdf, doc… và đặc biệt là các tài liệu khoa học của ĐHĐN tăng cao khi tìm kiếm trên Google và Google Scholar. Ngoài ra, nguồn tài liệu tổng hợp được sẽ được cập nhật vào kho dữ liệu phục vụ hệ thống phát hiện sao chép để góp phần nâng cao chất lượng đào tạo và nghiên của khoa học của cán bộ giảng viên, sinh viên ĐHĐN.

Tài liệu tham khảo

[1] Mike Thelwall, Liwen Vaughanand Lennart Björneborn, "Webometrics", *Annual Review of Information Science and Technology*, Volume 39, Issue 1, 2005, (81–135).

[2] R. K. Pandey, "Empirical Validation of Webometrics based Ranking of World Universities", *International Journal of Computer Science and Information Technologies (IJCSIT)*, Vol. 5 (1), 2014, (580-584).

[3] Châu Thùy Dương, Võ Trung Hùng, Hồ Phan Hiếu, "Xây dựng kho dữ liệu phục vụ hệ thống phát hiện sao chép", *Tạp chí Khoa học và Công nghệ ĐHĐN*, Số 1(86), 2015, (113-117).

[4] Phan Ngọc Đông,"DSpace - Giải pháp xây dựng thư viện số",*Tạp chí Thư viện Việt Nam*, Số 3(35), 2012, (39-41).

[5] Phan Ngọc Đông, "Ứng dụng phần mềm DSpace phiên bản 4.0 trong xây dựng thư viện số", *Tạp chí Thư viện Việt Nam*, Số 4(48), 2014, (31-33).

[6] Trần Nam Bình, Nguyễn Vĩnh Khương, "Tổng quan phương pháp luận xây dựng các tiêu chí đánh giá các trường đại học trên thế giới", *Tạp chí Khoa học Đại học Sư phạm TPHCM*, Số 3(68), 2015, (164-178).

[7] http://webometrics.info/

[8] http://www.dspace.org

[9] http://vnlp.net/?tag=webometrics