

KỸ THUẬT LÀM NỘI DUNG SỐ THEO TIÊU CHUẨN WEBOMETRICS

I. XẾP HẠNG WEBSITE ĐẠI HỌC QUA CHỈ SỐ WEBOMETRICS

Website của các đại học được xem như bộ mặt của đại học trên mạng Internet.

Website không chỉ là phương tiện cung cấp các thông tin tổ chức, hành chính, hoạt động của đại học mà còn là phương tiện cung cấp các tài liệu học tập và nghiên cứu - đây là một đặc thù của website đại học. Chính vì vậy, để đánh giá website của các đại học, ngoài quy mô (đo bằng số trang khác nhau trên cùng một tên miền), tính hấp dẫn (đo bằng số các tham chiếu từ website khác) thì cần tính đến số các file tài liệu trong các định dạng thường dùng và các tài liệu khoa học có thể nhận diện qua chuẩn mực trình bày.

Webometrics là một cách thức đánh giá các website của các đại học do nhóm nghiên cứu Cybermetrics Lab thuộc tổ chức Spanish National Research Council (CSIC) có trụ sở tại Madrid đề xuất và được thừa nhận rộng rãi, được các tổ chức đảm bảo chất lượng sử dụng. Đến nay, tổ chức này đã thực hiện xếp hạng cho hơn 20.000 đại học. Việc xếp hạng được thực hiện định kỳ 1 năm hai lần vào tháng 1 và tháng 7, bắt đầu từ 2004.

Chỉ số webometrics là một chỉ số tổng hợp được tính từ 4 loại chỉ số sau:

1. **Chỉ số quy mô website gọi là S** do viết tắt từ Size, tính theo số trang độc lập của website xuất hiện dưới tên miền tương ứng.

2. **Chỉ số nhận diện V** do viết tắt từ Visibility, tính bằng tổng tất cả các đường link ở các website khác đến một trang nào đó trong website. Theo ý nghĩa này, nó cũng gọi là chỉ số tham chiếu inlink. Chỉ số này được xem là chỉ số có ý nghĩa nhất về ảnh hưởng của một website trên Internet vì nó thể hiện sự hấp dẫn của website.

3. **Chỉ số độ phong phú nội dung R** do viết tắt từ Rich file đo bằng tổng số các file trong các định dạng tài liệu phổ biến như Microsoft Word (doc), Adobe Acrobat (pdf), Microsoft Powerpoint (ppt), Adobe PostScript (ps)

4. Chỉ số thư tịch khoa học Sc viết tắt từ Scholar, đo số tài liệu khoa học viết dưới định dạng html hay pdf có cấu trúc chuẩn của một tài liệu khoa học.

Việc đo các chỉ số trên thực hiện qua các công cụ tìm kiếm (search engine) của một số nhà cung cấp dịch vụ tìm kiếm lớn trên Internet như Google, Yahoo, Microsoft, Altavista... Các nhà cung cấp này sử dụng các phần mềm chuyên dụng (gọi là spider hay robot) lần vết theo các liên kết web trên mạng Internet để lấy các trang web đưa về máy chủ của họ, sau đó dùng các thuật toán tính đặc trưng để lập chỉ dẫn (index) các thông tin lấy được từ các trang web rồi đưa vào cơ sở dữ liệu chỉ dẫn của mình để phục vụ tìm kiếm.

Cần lưu ý rằng các thuật toán do mỗi nhà cung cấp sử dụng khác nhau nên kết quả đưa ra cũng không giống nhau. Vì thế các tổ chức đánh giá thường sử dụng một công thức phối hợp các kết quả đo của một số công cụ tìm kiếm để tạo ra một số đo tổng hợp. Ở đây, chúng ta sẽ không đi sâu vào chi tiết này.

Cách đo như sau:

Đối với chỉ số S: lấy số lượng trang nội dung trên 4 công cụ tìm kiếm Google, Yahoo, Live Search và Exalead với các lệnh có cú pháp dưới đây:

Yahoo	site:tên_miền OR site:www.tên_miền
Google	site:tên_miền OR site:www.tên_miền
Alta vista	domain: tên_miền OR domain:www.tên_miền

Thông thường các tổ chức thường sử dụng cả hai tên miền, một không bắt đầu từ "www", một có cả "www" (ví dụ ĐHQGHN có sử dụng cả hai tên miền vnu.edu.vn và www.vnu.edu.vn) nên các lệnh trên áp dụng đồng thời cho cả hai tên miền. Nhớ rằng phải gõ OR chữ hoa đừng gõ or.

Đối với chỉ số V, sử dụng các lệnh có cú pháp như sau:

Yahoo	link:tên_miền OR link:www.tên_miền
Alta vista	linkdomain: tên_miền OR linkdomain:www.tên_miền

Chỉ số R sẽ được đo theo từng kiểu file rồi cộng lại

Google	filetype:doc (site:tên_miền OR site:www.tên_miền)
	filetype:pdf (site:tên_miền OR site:www.tên_miền)
	filetype:ppt (site:tên_miền OR site:www.tên_miền)

	filetype:ps (site:tên_miền OR site:www.tên_miền)
--	--

Đối với chỉ số Sc, ta dùng công cụ tìm kiếm của Google Scholar. Công cụ này căn cứ vào các tài liệu pdf và html, tuân thủ định dạng chuẩn của tài liệu khoa học. Cú pháp tính chỉ số Sc như sau:

Google Scholar	site:tên_miền OR site:www.tên_miền
----------------	------------------------------------

Trong các lệnh trên để tính các chỉ số cho ĐHQGHN chỉ cần thay tên_miền bởi vnu.edu.vn

Để xếp hạng một nhóm website (ví dụ có thể thực hiện theo từng khu vực như Đông Nam Á, châu Á hay toàn cầu) trước hết người ta xếp hạng theo từng chỉ số S, V, R và Sc. Website nào có chỉ số lớn hơn sẽ được xếp trước (có thứ tự nhỏ hơn). Website có chỉ số lớn nhất sẽ được xếp thứ 1. Nếu thứ tự xếp hạng theo các chỉ số S, V, R và Sc lần lượt là R(S), R(V), R(R) và R(Sc) thì chỉ số webometrics sẽ được tính bằng công thức

$$W = 25\% R(S) + 50\% R(V) + 12,5\% R(R) + 12,5\% R(Sc)$$

Sau đó ta xếp các website theo thứ tự tăng dần của W. Thứ tự theo W này được lấy làm thứ hạng của website trong nhóm website được đưa ra xếp hạng.

Nhận xét rằng theo công thức này thì chỉ số ảnh hưởng có giá trị nhất là V, sau đó mới đến S và cuối cùng là R và Sc.

Đặc biệt lưu ý rằng, các công cụ tìm kiếm không chỉ tính số lượng trang, số lượng inlink và số lượng file của chính website được chỉ ra mà còn tính cả số lượng các website là tên miền con. Theo đó, các số liệu đo của các website như hus.vnu.edu.vn, ussh.vnu.edu.vn, ueb.vnu.edu.vn v.v. không những được tính cho chính nó mà còn được tính cho cả website vnu.edu.vn. Tuy nhiên nếu các website hus.edu.vn, ussh.edu.vn, ueb.edu.vn... không đồng thời là các trang tương ứng hus.vnu.edu.vn, ussh.vnu.edu.vn, ueb.vnu.edu.vn thì chúng sẽ không đóng góp được vào tổng chung của website vnu.edu.vn. Vì thế cách sử dụng tên miền của các tổ chức thành viên hay trực thuộc có ảnh hưởng quan trọng đến chỉ số xếp hạng của website.

Trên thực tế các robot phải đề phòng trường hợp gặp vòng lặp (tức là sau một số bước lần vết lại quay về chính một trang đã được thăm) nên sẽ chỉ đếm đến

một độ sâu nhất định chứ không truy vô hạn. Vì thế độ sâu lần vết trong một website con tính từ gốc là website mẹ thường nhỏ hơn độ sâu lần vết tính từ gốc là một website con. Điều này giải thích cho việc chỉ số đo từ website mẹ thường nhỏ hơn tổng các chỉ số tương ứng của tất cả các site con. Vấn đề là cần phải tổ chức như thế nào để các công cụ tìm kiếm có thể tiếp cận đến nhiều file, nhiều trang và nhiều inlink nhất.

Về cấu trúc và điều kiện của một file được nhận diện như một tài liệu khoa học.

Do các công cụ tìm kiếm không đủ thông minh để nhận diện nội dung của một tài liệu được xem là tài liệu khoa học, nên Google Scholar đã chọn các nhận diện thông qua phân tích cấu trúc.

Việc nhận dạng cấu trúc của Google Scholar có hai khía cạnh:

1. Theo cách trình bày, theo đó có dấu hiệu các phần: tên tài liệu, tác giả và phần tài liệu tham khảo phân biệt bằng cỡ chữ và cách viết phần tài liệu tham khảo.
2. Sử dụng siêu dữ liệu (metadata) để thiết lập các thông tin tìm kiếm bằng phương pháp lập chỉ dẫn (index) trong cơ sở dữ liệu tìm kiếm của công cụ tìm kiếm Google Scholar. Siêu dữ liệu là các loại dữ liệu dùng để định nghĩa dữ liệu khác. Các siêu dữ liệu được viết trong những đoạn mã lồng ghép trong bài dưới dạng các thẻ (tag).

Google Scholar có hướng dẫn chi tiết cách viết các tài liệu khoa học trong các địa chỉ <http://scholar.google.com/intl/en/scholar/publishers.html> và <http://scholar.google.com/intl/en/scholar/inclusion.html>

Những điểm chính được tóm tắt như sau:

- Các tài liệu được đặt trong file html hay pdf với độ lớn không quá 5 MB
- Với cách nhận dạng thứ nhất, cần lưu ý:
 - + Tiêu đề của bài báo xuất hiện trong dòng đầu tiên với một cỡ lớn của trang đầu tiên. Đối với file Pdf, cỡ chữ được khuyến cáo từ 24 trở lên, còn đối với file html cỡ chữ được quy định bằng thẻ <h1> hoặc <h2>.
 - + Dòng tiếp theo ngay theo tên tiêu đề là dòng các tác giả của tài liệu. Đối với file pdf cỡ chữ phải bé hơn tiêu đề nhưng lớn hơn phần nội dung bài, thường từ 16 đến 23, còn trong định dạng html sẽ được quy

định bằng thẻ <h3>. Các đồng tác giả được tách bằng các dấu phẩy và bỏ qua các thông tin bằng cấp của các tác giả. Khuyến cáo file pdf chỉ dùng 3 loại cỡ chữ.

- + Phân liệt kê tài liệu tham khảo (Biography hay Reference) đặt ở phần cuối cùng của bài. Sau dòng có chỉ dẫn Biography hay Reference là các tài liệu tham khảo. Đối với file pdf có số thứ tự đánh theo kiểu "1 - 2 -.. 3." hoặc "[1] - [2] - [3]", đối với file html dùng thẻ .

- Với việc quy định cấu trúc bằng siêu dữ liệu, có rất nhiều phần mềm quản lý tài liệu điện tử dùng các kiểu siêu dữ liệu được Google Scholar hỗ trợ như Eprints, DSpace, Common Digital hoặc OJS. Khuyến cáo cấu hình các phần mềm đó để phần mềm có thể tự động tạo các thẻ siêu dữ liệu phục vụ cho Google Scholar. Google Scholar hỗ trợ có khả năng hỗ trợ các loại thẻ:

- Thẻ Highwire Press (ví dụ, citation_title),
- Thẻ Eprints (ví dụ, eprints.title),
- Thẻ BE Press (ví dụ, bepress_citation_title),
- Thẻ PRISM (ví dụ, prism.title).

- Trong trường hợp không sử dụng các thẻ trên có thể sử dụng thẻ Dublin Core (ví dụ, DC.title) một loại chuẩn mô tả tài liệu rất phổ biến, quy định 15 yếu tố thông tin chuẩn của tài liệu bao gồm: Nhan đề (Title), Tác giả (Creator), Chủ đề (Subject), Mô tả (Description), Tổ chức xuất bản (Publisher), Tác giả phụ (Contributor), Ngày tháng (Date), Loại hình (Type), Mô tả vật lý (Format), Định danh tài liệu (Identifier), . Nguồn gốc (Source), Ngôn ngữ (Language), Liên kết (Relation), Nơi chứa (Coverage) và Bản quyền (Rights)

Sau đây là một vài ví dụ về các thẻ siêu dữ liệu được lồng ghép trong các tài liệu giúp cho việc xử lý các tài liệu.

```
<meta name="citation_title" content="The testis isoform of the phosphorylase kinase catalytic subunit (PhK-T) plays a critical role in regulation of glycogen mobilization in developing lung">
<meta name="citation_author" content="Liu, Li">
<meta name="citation_author" content="Rannels, Stephen R.">
<meta name="citation_author" content="Falconieri, Mary">
<meta name="citation_date" content="1996/05/17">
<meta name="citation_journal_title" content="Journal of Biological Chemistry">
<meta name="citation_volume" content="271">
<meta name="citation_issue" content="20">
<meta name="citation_firstpage" content="11761">
<meta name="citation_pdf_url"
content="http://www.example.com/content/271/20/11761.full.pdf">
```

Sau đây là ví dụ một số thẻ đi với chuẩn mô tả theo Dublin Core

```
<META NAME= "DC.Title" CONTENT="Kita Yam (Japan)">
```

```
<META NAME= "DC.Creator" CONTENT="Kertesz, Andre">  
<META NAME= "DC.Date" CONTENT="1968">  
<META NAME= "DC.Type" CONTENT="image">  
<META NAME= "DC.Format" CONTENT="image/gif">  
<META NAME= "DC.Identifier"  
CONTENT="http://foo.bar.zaf/kertesz/kyama">
```

Những người dùng bình thường có thể không cần biết sâu cấu trúc bên trong của tài liệu điện tử như ví dụ trên. Khi được cấu hình, các phần mềm quản trị tài liệu điện tử có khả năng tự động tạo các thẻ này. Điều quan trọng là nếu có điều này thì Google Scholar sẽ hỗ trợ việc lập chỉ dẫn tìm kiếm tài liệu khoa học dễ dàng, tiện lợi và nhanh chóng.

Một số vấn đề có liên quan có thể tìm thấy trong các tài liệu hướng dẫn về D-Space mà ĐHQGHN đã giao cho Trung tâm Thông tin Thư viện hỗ trợ các đơn vị triển khai.

II. THỰC TRẠNG XẾP HẠNG WEBSITE CỦA ĐHQGHN

Ngay từ khi khai trương mạng VNUnet (năm 1999), ĐHQGHN đã xây dựng website. Từ đó đến nay, website ĐHQGHN đã được nhiều lần nâng cấp, được duy trì thường xuyên, đáp ứng được việc công bố thông tin của ĐHQGHN ra công chúng. Cùng với các thông tin của website, các phần mềm quản trị nội dung CMS (Content Management System) và cơ sở dữ liệu tin tức, sự kiện và tổ chức của ĐHQGHN cũng đã được xây dựng. Đây là một cơ sở dữ liệu (CSDL) lưu trữ tốt các thông tin về hoạt động và phát triển của ĐHQGHN.

Hầu hết các trường, các khoa trực thuộc, các Viện và Trung tâm trực thuộc ĐHQGHN đã xây dựng website của đơn vị mình, góp phần đáng kể vào việc quảng bá hình ảnh của đơn vị và của ĐHQGHN.

Năm 2010, xếp hạng webometrics của ĐHQGHN đứng thứ 58 trong top 100 của Đông Nam Á và đứng thứ 3 ở Việt Nam sau ĐH Cần Thơ và ĐH Bách Khoa TPHCM. Sang đầu năm 2011, ĐHQGHN chỉ còn đứng thứ 85 trong top 100 của Đông Nam Á và đứng thứ 7 ở Việt Nam, sau cả ĐH An Giang, ĐH BK Hà Nội, ĐH Nông lâm TP HCM, ĐH Nông nghiệp 1. Nguy cơ lọt ra khỏi top 100 của Đông Nam Á là cận kề do sự tiến bộ nhanh của các đại học khác nếu ĐHQGHN không kịp thời hành động (Xem bảng 1)

Bảng 1. Xếp hạng của một số trường hàng đầu của khu vực Đông Nam Á và của các trường Đại học của Việt Nam trong top 100 Đông Nam Á

REGION RANK	UNIVERSITY	WORLD RANK	R(S)	R(V)	R(R)	R(Sc)
1	National University of Singapore	92	149	160	72	80
3	Chulalongkorn University	418	303	560	391	337
5	Nanyang Technological University	466	376	516	945	258
6	Mahidol University	499	278	773	738	98
53	Ho Chi Minh City University of Technology	1,455	1,420	1,713	1,425	2,055
58	Can Tho University	1,600	1,447	1,958	916	3,243
59	Nong Lam University	1,682	1,475	2,858	1,919	1,817
62	An Giang University	1,785	2,467	6,859	2,742	539
78	Hanoi Agricultural University	2,268	3,308	7,730	2,507	1,219
81	Hanoi University of Technology	2,320	1,756	1,613	4,202	2,501
85	Vietnam National University	2,380	2,054	1,670	2,439	3,720
94	Ho Chi Minh City University of Economics	2,702	1,458	2,283	2,055	5,624

Liên quan đến chỉ số xếp hạng, hệ thống website của ĐHQGHN còn một số bất cập sau đây:

- Hệ thống tên miền chưa thống nhất. Một số đơn vị hosting website của mình bên ngoài và dùng tên miền không nằm trong hệ thống của ĐHQGHN. Một số đơn vị còn sử dụng đồng thời nhiều tên miền khác nhau, khiến các máy tìm kiếm không nhận diện được các website của cùng một tổ chức. Trước tháng 3/2011 chỉ có 1/3 số đơn vị trực thuộc dùng tên miền trong hệ thống của ĐHQGHN, 1/3 có website dùng tên miền và địa chỉ ở ngoài ĐHQGHN và 1/3 chưa có website. ĐHQG HN đã chỉ đạo quyết liệt việc thống nhất tên miền. Trung tâm UDCNTT đã cấp toàn bộ tên miền trong hệ thống của ĐHQG cho các đơn vị kể cả các đơn vị cấp dưới. Đến nay chỉ còn một vài đơn vị chưa thực hiện được điều này, có thể do các thủ tục đối với các nhà cung cấp dịch vụ hosting.

- Chỉ số rich file, chỉ số inlink và chỉ số thư tịch khoa học đều rất thấp, không tương xứng với thực lực của ĐHQG HN. Cho tới nay chỉ có website của ĐHQGHN, Trường ĐH Công nghệ và Trường ĐH Kinh tế có số lượng rich fie (chỉ số R) và số trang web trong cùng tên miền (chỉ số S) là tương đối lớn. Các đơn vị khác đóng góp hầu như không đóng góp đáng kể vào chỉ số chung của ĐHQGHN.

BGD ĐHQGHN đã thi hành nhiều biện pháp để nâng cao các chỉ số này. Trường ĐHCN (phòng thí nghiệm công nghệ tri thức KTLab), Trung tâm TTTV, TT UDCNTT đã đầu tư nhiều công sức để làm dữ liệu nạp vào website cũ trước

đây và nạp trước vào cổng thông tin điện tử đang được triển khai song song. Trong một tháng gần đây, các chỉ số phục vụ xếp hạng đã được cải thiện rõ rệt đối với các site thành phần là site của Trung tâm Thông tin Thư viện và chính site của ĐHQGHN. Ở các nơi khác tình hình chưa khác biệt nhiều.

Bảng 2A là các chỉ số phục vụ xếp hạng đo vào đầu tháng 3/2011; Bảng 2B là các kết quả đo ngày 7/4/2011, một tháng sau khi triển khai hàng loạt công việc thống nhất tên miền, nạp rich file và chuẩn hóa cấu trúc tài liệu khoa học.

Bảng 2A. Chỉ số Sc, V, S của một số website trong ĐHQGHN đo vào đầu tháng 3/2011

Đơn vị	Sc	V	S	R (doc/pdf/ppt)
Đại học Quốc gia Hà Nội www.vnu.edu.vn	NA	1.226	63.100	1.750/1.860/ 50
Trường ĐH KHXH&NV www.ussh.edu.vn	NA	282	5.510	266/ 398/ 0
Trường ĐHKHTN www.hus.edu.vn	NA	402	4.173	132/422/ 0
Trường ĐH Ngoại ngữ www.ulis.vnu.edu.vn	NA	36	1.950	1/ 1/ 37
Trường ĐH Công nghệ www.coltech.vnu.edu.vn	NA	165	15.800	246/ 402/ 37
Trường ĐH Kinh tế www.ueb.vnu.edu.vn	NA	51	17.630	104/ 71/ 01
Trường ĐH Giáo dục www.education.vnu.edu.vn	NA	20	3.250	131/ 23/ 0
Khoa Quốc tế www.is.vnu.edu.vn	NA	61	0	0/ 0/ 0
Trung tâm Thông tin Thư viện www.lic.vnu.edu.vn	NA	99	87	0/ 6/ 0

Tạp chí khoa học VNU www.js.vnu.edu.vn	NA	10	2.540	2/ 552/ 0
---	----	----	-------	-----------

Bảng 2B. Các chỉ số Sc, V, S, R của ĐHQG HN và một số đơn vị đào tạo và nghiên cứu tính đến ngày 7/4/2011.

Đơn vị	Sc	V	S	R Doc/Pdf/Ppt	Đóng góp vào VNU			
					Sc	V	S	R
Đại học Quốc gia Hà Nội www.vnu.edu.vn	238	1109	195000	3160/4480/64				
Trường ĐH KH&NV ussh.edu.vn www.ussh.vnu.edu.vn	0	179	4950	262/126/0		1	3150	147/67/0
Trường ĐHKHTN hus.vnu.edu.vn www.hus.vnu.edu.vn	0	401	7270	150/42/1		1	64	8/7/0
Trường ĐH Ngoại ngữ www.ulis.vnu.edu.vn	0	38	3200	1/1/0		38	3200	1/1/0
Trường ĐH Công nghệ www.coltech.vnu.edu.vn	6	165	92100	345/722/43	0	165	92100	345/722/43
Trường ĐH Kinh tế www.ueb.vnu.edu.vn www.ueb.edu.vn	0	60	252500	823/284/6		31	52700	79/97/1
Trường ĐH Giáo dục www.education.vnu.edu.vn	0	1	4730	222/48/0		1	4730	222/48/0
Khoa Quốc tế www.is.vnu.edu.vn	0	61	906	1/39/3		61	906	1/39/3
Khoa Luật Khoaluatdhqg.edu.vn	0	0	115	3/0/0		0	115	0/0/0
Khoa Sau Đại học Sgs.vnu.edu.vn	0	0	33500	0/0/0			33500	0/0/0
Khoa Quản trị Kinh doanh Hsb.edu.vn Bs.vnu.edu.vn	0	8	1070	0/0/0		6	1490	0/0/0
Trung tâm TT Thư viện www.lic.edu.vn www.lic.vnu.edu.vn	16	4	223	895/875/0		102	9530	895/875/0
Tạp chí khoa học VNU www.js.vnu.edu.vn	64	12	1420	1/233/0		12	1420	1/233/0
Viện Vi sinh & CNSH www.Biotechvnu.edu.vn	0	443	3130	2/4/0		0	0	0/0/0
Viện Việt Nam học và Khoa học Phát triển www.ivides.edu.vn	0	4	390	7/6/0		0	0	0/0/0

Viện Công Nghệ thông tin iti.vnu.edu.vn	0	37	364	0/0/0		37	364	0/0/0
--	---	----	-----	-------	--	----	-----	-------

3. CÁC KỸ THUẬT TỔ CHỨC DỮ LIỆU PHỤC VỤ CHO NÂNG CÁC CHỈ SỐ WEBOMETRICS.

Việc nâng hạng thực chất là nâng các chỉ số cơ sở S, R, V, Sc. Để nâng các chỉ số này chỉ còn 3 vấn đề:

1. Tăng số lượng các rich file, chú trọng việc viết đúng chuẩn tài liệu khoa học giúp nhận diện được các tài liệu khoa học;
2. Tăng độ hấp dẫn của tin tức để có nhiều inlink;
3. Làm thế nào để các robot tìm thấy các tài liệu.

Hai nội dung 1 và 2 phụ thuộc vào khối lượng và nội dung tài liệu và tin tức mà mỗi đơn vị phải có trách nhiệm thực hiện.

Việc thứ 3 là vấn đề kỹ thuật, đồng thời là nội dung của đợt tập huấn này.

Sau đây là các khuyến cáo:

1. Tổ chức tên miền đúng quy cách, tất cả phải nằm trong hệ thống tên miền của ĐHQGHN, kể cả trang của các đơn vị cấp dưới. Để không làm mất mối liên hệ của các tên miền của các đơn vị chưa dùng tên miền của ĐHQGHN, ĐHQGHN cho phép các đơn vị được dùng tới hết năm 2011. Sau thời gian này, phải hủy bỏ các tên miền cũ không nằm trong hệ thống tên miền của ĐHQGHN. Việc dùng song song hai tên miền phải là song song thực sự, hai tên miền đều được khai báo cho cùng một website vật lý, chứ không phải sử dụng tên miền của ĐHQG để chuyển hướng đến tên miền cũ vì cách làm này sẽ không làm cho các site của các đơn vị có đóng góp thực sự vào các chỉ số cơ sở để tính webometrics.

Ngoài ra phải chuyển tên miền của cả các đơn vị cấp dưới theo hệ thống tên miền của ĐHQGHN.

2. Các tài liệu phải có một link hiển (tức là link từ một hyperlink dạng text có sẵn hoặc được tự động tạo ra trong quá trình tìm kiếm của robot. Phải loại trừ mọi khả năng cản trở việc tìm ra tài liệu của robot như:

- Bắt phải gõ tham số tìm kiếm rồi mới làm hiển thị liên kết
- Website đòi hỏi đăng nhập
- Đòi hỏi phải cài đặt một phần mềm riêng
- Đòi hỏi xác nhận bằng cách gõ vào một tổ hợp ký tự để chống spam

- Chặn popup
- Đòi hỏi quảng cáo xen kẽ
- Yêu cầu bấm vào các nút (chứ không phải link bằng chữ)
- Bắt phải cuộn trang để làm xuất hiện link

3. Tổ chức link thích hợp để các robot có thể tìm được hết tài liệu

- Không để không gian tìm kiếm cục bộ quá lớn. Ví dụ việc tạo ra các trang mà trong đó có số link quá lớn sẽ hạn chế việc tìm ra các file.
- Khuyến cáo số địa chỉ trung gian để đi đến tài liệu không vượt quá 10% số địa chỉ tài liệu.
- Một số robot có thể nhận diện các dấu hiệu chặn tìm kiếm ở một địa chỉ nào đó. Khi đó rất nên đặt ngăn chặn để các robot khỏi mất công vào tìm các tài liệu ở các địa chỉ không có ích. Ví dụ website của một đại học có thể có link đến một trang mua sắm trực tuyến với lệnh mở ra vùng mua hàng là `add_cart.php`. Khi đó có thể đặt lệnh trong một file cấu hình `robot.txt` để robot không đi tìm trang do `add_cart.php` dẫn đến. Để làm điều này, ta chỉ cần đưa vào file `robot.txt` một lệnh như dưới đây.

User-agent: *

Disallow: / `add_cart.php`

Chi tiết về điều này có thể tham khảo ở địa chỉ <http://www.robotstxt.org/>

- Không để sâu quá, phải đi qua rất nhiều link trung gian mới tới file. Khuyến cáo độ sâu tìm kiếm không quá 10

Việc đưa các file lên mạng phải là một việc làm thường xuyên và không thấy hiệu quả ngay. Đối với các site phức tạp, thời gian để thấy sự thay đổi với các chỉ số Sc có thể từ vài tuần đến vài tháng.

4. KẾT LUẬN

Nâng hạng website không phải là một vấn đề hình thức. Ngoài việc tạo ảnh hưởng lớn hơn, thì việc làm giàu có các tư liệu của website sẽ giúp ích cho chính cán bộ và sinh viên của ĐHQG trong hoạt động học tập và nghiên cứu.

Để nâng hạng website, ngoài việc chuẩn bị nội dung, cách thức cấu trúc tin tức và kỹ năng đưa tin về mặt kỹ thuật là rất quan trọng để các công cụ tìm kiếm có thể nhận diện được. Ý nghĩa của việc nhận diện không phải ở chỗ làm tăng các chỉ số một cách hình thức mà chính là ở chỗ: khi đã nhận diện được, các công cụ tìm

kiếm mới lập được chỉ dẫn và từ đó có thể tìm kiếm được tài liệu trên Internet qua các công cụ tìm kiếm

Việc nâng hạng website của ĐHQG HN không phải là một chiến dịch có thời hạn mà là sự mở đầu của một quá trình liên tục đòi hỏi trách nhiệm của tất cả các đơn vị trong ĐHQG HN, ý thức được sự hiện diện của ĐHQGHN trên internet nhằm phục vụ các hoạt động khai thác tư liệu cho cán bộ và sinh viên của ĐHQGHN.