

TRẮC LƯỢNG THƯ MỤC: CÁC CHỈ SỐ PHỔ BIẾN- VIỆC ỨNG DỤNG VÀ VẤN ĐỀ ĐÀO TẠO NGÀNH THÔNG TIN, THƯ VIỆN

ThS Trần Mạnh Tuấn

Viện Thông tin Khoa học xã hội

Tóm tắt: Trên cơ sở giới thiệu về các chỉ số trắc lượng thư mục phổ biến và việc ứng dụng trắc lượng thư mục tại các doanh nghiệp thông tin và xuất bản để tạo ra các loại CSDL có khả năng thực hiện các bài toán về thống kê thư mục để thấy rõ những yêu cầu về sự đổi mới chương trình đào tạo ngành thông tin, thư viện. Nghiên cứu, đề xuất các giải pháp cụ thể để tích hợp nội dung trắc lượng thư mục vào chương trình đào tạo từ bậc Cao đẳng, Đại học và Sau đại học tại các trường đại học về chuyên ngành thông tin, thư viện.

Từ khóa: Trắc lượng thư mục; thông tin-thư viện; thống kê thư mục; đào tạo thông tin-thư viện.

Bibliometrics: key indicators - their application and training in information and library faculties

Summary: On basis of introducing key bibliometric indicators and their applications at information and publishing enterprises for creating databases, capable to solve bibliostatistical problems, makes clear requirements on improvement of training programmes for library and information specialties; studies and sets forth specific solutions to integrating the content of bibliometrics into training programmes for high-school, undergraduate and post-graduate levels at universities specialized in library and information science.

Keywords: Bibliometrics; library and information; bibliographical statistics; training in library and information science.

Mở đầu

Các số liệu thống kê về mối quan hệ giữa các tài liệu khoa học, phân tích văn bản tài liệu,... với tư cách là tiền đề của trắc lượng thư mục đã xuất hiện ngay từ những năm 1920-1930 với các phát kiến của G.K. Zipf, S.C. Bradford¹,... Cho tới khoảng đầu những năm 1960, các nghiên cứu về mối quan hệ này được quan tâm với nhiều mục đích khác nhau, từ chỗ làm cơ sở để xác định trong số các tài liệu có liên quan, tài liệu nào được đánh giá là tốt hơn, có giá trị hơn, đến chỗ dần hình thành phương pháp đánh giá khoa học dựa trên các xuất bản phẩm khoa học của mọi chủ thể: cá nhân, tổ chức, cộng đồng, quốc gia. Lúc này, sự ra đời của chỉ số tác động (Impact Factor - IF) với công lao trước hết của E. Gardfield, Viện trưởng sáng lập Viện Thông tin khoa học (ISI) Mỹ, và các nghiên cứu tiếp theo đã làm cho cộng đồng khoa học thế giới dành sự quan tâm

đặc biệt đến *Danh sách Philadelphia*, danh sách các tạp chí khoa học có uy tín trên thế giới được ISI công bố hàng năm và ngày nay được biết đến là các tạp chí có trong Journal Master List do Thomson Reuters xây dựng. Cho tới những năm đầu thế kỷ XXI, vấn đề đánh giá khoa học đối với mọi đối tượng đang rất được quan tâm trên phạm vi toàn cầu, và một lần nữa, chỉ số IF cùng hàng loạt các chỉ số được sử dụng để đánh giá khoa học luôn thu hút sự quan tâm của giới nghiên cứu. Bài viết này giới thiệu một số vấn đề của trắc lượng thư mục, với các nghiên cứu cơ bản và ứng dụng của nó.

Bài viết tập trung vào các nội dung sau:

- Những chỉ số phổ biến của trắc lượng thư mục;
- Ứng dụng trắc lượng thư mục tại các doanh nghiệp thông tin và xuất bản;
- Nội dung trắc lượng thư mục trong chương trình đào tạo về thông tin, thư viện.

¹ Khi còn là sinh viên, chúng tôi đã biết được các tri thức đầu tiên về các định luật mà G.K. Zipf và S.C. Bradford xây dựng từ một giáo trình khá nổi tiếng về lĩnh vực thông tin-thư viện, được xuất bản tại Liên-xô vào cuối những năm 1960: Lý luận và thực tiễn của thông tin khoa học và kỹ thuật, bản tiếng Nga. Điều đó cho thấy: Các vấn đề của trắc lượng thư mục đã thực sự là một trong số các nội dung/hướng nghiên cứu của các khoa học thư viện và thông tin (LIS).

1. Những chỉ số phổ biến của trắc lượng thư mục

Trắc lượng thư mục cho phép xác định hiệu suất và/hoặc tác động của mỗi tài liệu khoa học. Mỗi đại lượng trắc lượng thư mục hay mỗi loại chỉ số của trắc lượng thư mục có tính năng của riêng mình, nhưng nói chung, tất cả đều nhằm thực hiện được việc xếp hạng tài liệu và cung cấp sự đánh giá sâu sắc về tài liệu dựa trên các phân tích trích dẫn. Các chỉ số này được xây dựng dựa trên những tiền đề cơ bản: khi diễn ra một trích dẫn đến một tài liệu tức là một hình thức xác nhận giá trị/tầm quan trọng của tài liệu đó; và phân tích cơ bản nhất được thực hiện bằng việc đếm số lượt trích dẫn đến một tài liệu. Nếu một tài liệu được trích dẫn nhiều hơn một tài liệu khác thì có nghĩa có nhiều người hơn quan tâm đến tài liệu đó, và vì thế nó được coi là quan trọng hơn so với tài liệu khác. Từ đó, tiếp tục với *quy mô lớn hơn*, các trích dẫn đến các tài liệu nguồn (thông qua các tài liệu được xuất bản trên tài liệu nguồn (tạp chí, sách, kỷ yếu ...)) cũng có thể được tính toán, và do đó, có thể xác định được tầm quan trọng của một tài liệu nguồn đối với cộng đồng của nó trong việc so sánh với chỉ số trắc lượng thư mục của các tài liệu nguồn khác. Các chỉ số trắc lượng thư mục đối với tài liệu nguồn khác nhau được hình thành trên cơ sở sử dụng các phương pháp luận và các nguồn dữ liệu khác nhau, do đó đưa ra các quan điểm khác nhau về công bố khoa học. Các chuyên gia trắc lượng thư mục sẽ sử dụng các số liệu khác nhau tùy thuộc vào những mục đích và quan điểm nghiên cứu, quan điểm đánh giá nghiên cứu khác nhau mà mình theo đuổi.

Dưới đây là phần giới thiệu khái lược về một số chỉ số trắc lượng thư mục phổ biến đối với tài liệu khoa học.

a) Chỉ số tác động

(IF) được sử dụng như một chỉ số phản ánh hiệu quả khoa học thông qua tầm ảnh hưởng của một tạp chí. IF được xác định như sau:

$$IF(T) = \frac{C}{A}$$

Trong đó **C** là tổng số lượt trích dẫn đến các bài báo được xuất bản trên tạp chí **T** trong một khoảng thời gian xác định (chuẩn là 2 năm liên tiếp trước năm hiện tại); **A** là tổng số bài báo được công bố trên tạp chí **T** trong khoảng thời gian tương ứng.

Người ta có thể tính chỉ số IF cho các khoảng thời gian khác nhau khác là 1 năm, 3 năm, 4 năm [2], [3], [12]... Trên các trang của mỗi tạp chí khoa học đã được Thomson Reuters lựa chọn thường có giới thiệu chỉ số IF của tạp chí đó. Ví dụ, tại địa chỉ <http://www.journals.elsevier.com/the-journal-of-academic-librarianship> giới thiệu tạp chí The Journal of Academic Librarianship, tại thời điểm 29/9/2014, có các thông tin sau:

Journal Metrics:

...

Impact Factor: 0,574

5-Year Impact Factor: 0,838

“Impact Factor: 0,574” phản ánh có trung bình 0,574 lượt trích dẫn đến 1 bài báo của tạp chí trên trong 2 năm 2012-2013; “5-Year Impact Factor: 0,838” phản ánh có trung bình 0,838 lượt trích dẫn đến 1 bài báo của tạp chí trên trong 5 năm, từ năm 2009 đến năm 2013.

Trên cơ sở chỉ số tác động (IF), còn có một số chỉ số khác được sử dụng để đánh giá, xếp hạng tạp chí khoa học như: Chỉ số

Nghiên cứu - Trao đổi

tác động loại bỏ số lượt tự trích dẫn (IF*); Chỉ số nửa vòng đời được trích dẫn của tạp chí (Journal Cited Half Life); Chỉ số nửa vòng đời trích dẫn của tạp chí (Journal Citing Half Life); Chỉ số Journal Immediacy Index (JImI); Chỉ số Journal Influence Index (JII); Chỉ số Paper Influence Index (PII) [13, 62-74]².

b) Chỉ số SNIP³

SNIP - chỉ số đo lường tạp chí được GS Henk Moed, Đại học Tổng hợp Leiden đề xuất, xác định tác động trích dẫn theo ngữ cảnh dựa trên xác định giá trị của trích dẫn trong tổng số các trích dẫn trong một lĩnh vực chủ đề xác định. Giá trị của của một trích dẫn đơn lẻ sẽ được gán cao hơn khi nó được trích dẫn từ các bài viết thuộc một lĩnh vực khác. SNIP tiếp tục phát triển quan điểm của E. Garfield về “tiềm năng trích dẫn” của một lĩnh vực nghiên cứu: Tiềm năng trích dẫn của một lĩnh vực nghiên cứu được xác định là độ dài trung bình danh sách các tài liệu tham khảo trong lĩnh vực đó...

Các chỉ số trắc lượng thư mục đầu vào được sử dụng để tính toán SNIP là RIP, DCP và RDCP:

- RIP_j (Raw Impact for Paper) của tạp chí j trong năm hiện tại là tỷ số giữa tổng số các trích dẫn của năm hiện tại trên số các bài báo

được xuất bản trong 3 năm liền trước năm hiện tại của tạp chí j⁴;

- DCP_j (Database Citation Potential) của tạp chí j trong năm hiện tại là số lượng trung bình các tham khảo trong một bài báo được xuất bản trong 3 năm liền trước năm hiện tại của tạp chí j⁵;

- RDCP (Relative Database Citation Potential) trong năm hiện tại của một tạp chí là tỷ số giữa DCP của tạp chí đó trên giá trị DCP trung bình của CSDL chứa tạp chí đó.

$$\text{Giá trị trung bình DCP} = \frac{\sum r_j, j=1 \div N}{N}$$

trong đó, r_j là DCP của tạp chí j có trong CSDL, và CSDL này bao quát N tạp chí.

Khi đó, RDCP trong năm hiện tại của tạp chí được xác định theo công thức:

$$\text{RDCP}_j = \frac{\text{DCP}_j}{\text{Giá trị trung bình DCP}}$$

Giá trị của chỉ số SNIP trong năm hiện tại của tạp chí j được xác định theo công thức sau:

$$\text{SNIP}_j = \frac{\text{RIP}_j}{\text{RDCP}_j} \quad [14]$$

² Các chỉ số JII và PII được CJR Journal-Ranking.com sử dụng để xếp hạng tạp chí. Có thể tìm hiểu các chỉ số này qua các trang web tương ứng.

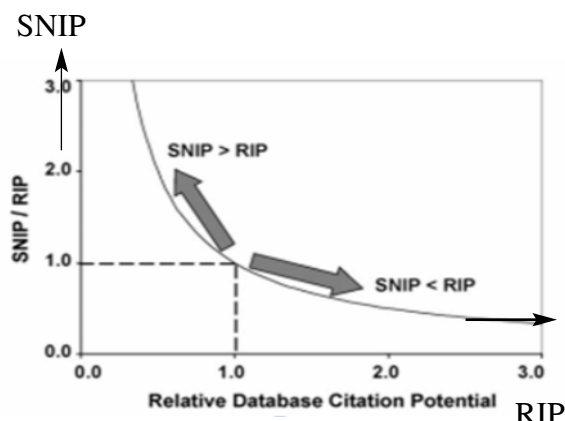
³ Tạm dịch là Tác động chuẩn từ nguồn trên một bài báo của tạp chí (Source Normalized Impact per Paper – SNIP). Các thông tin giới thiệu tại đây được trích dẫn từ [14].

⁴ Ví dụ, tạp chí Journal of Informetrics trong các năm 2010, 2011 và 2012 xuất bản tổng cộng 100 bài báo. Trong năm 2013 có tổng cộng 200 lượt trích dẫn đến 100 bài báo này. Khi đó, RIP năm 2013 của Journal of Informetrics là $200/100 = 2$.

⁵ Tiếp ví dụ trên: Giả sử trong 100 bài báo trên của Journal of Informetrics, có tổng số 2.000 tài liệu tham khảo, khi đó, DCP của tạp chí này năm 2013 là $2.000/100 = 20$.

Nghiên cứu - Trao đổi

Hình 1 sẽ phản ánh mối quan hệ giữa các chỉ số trắc lượng thư mục trên⁶.



Hình 1: Tỷ lệ giữa SNIP và RIP là một hàm số của RDCP. Đồ thị của hàm số này có dạng một đường hyperbol với trục hoành là RIP và trục tung là SNIP [10].

c) Chỉ số H

Chỉ số H (H-index) được nhà vật lý học Jorge Hirsch (Đại học California San Diego)

xây dựng năm 2005. Chỉ số H được tính toán dựa vào số công trình công bố và số lần trích dẫn. Mục tiêu của chỉ số H là đo lường mức độ ảnh hưởng tích lũy của một nhà khoa học. Chỉ số H được định nghĩa như sau: Chỉ số H của một nhà khoa học là H công trình trong số N công trình của nhà khoa học đó được trích dẫn ít nhất là H lần, và (N – H) được trích dẫn dưới H lần. Ví dụ, nếu một nhà khoa học có chỉ số H = 20 có nghĩa là nhà khoa học này có 20 công trình nghiên cứu với mỗi công trình được trích dẫn ít nhất là 20 lần. Nếu phản ánh quan hệ giữa số lượt trích dẫn đến các công trình của một tác giả và số lượng các công trình được thực hiện bởi tác giả đó dưới dạng các điểm đồ thị, thì các điểm phản ánh mối tương quan này có hình dạng của một đường hyperbol [13]⁸. Ví dụ, GS Nguyễn Văn Tuấn đã giới thiệu chỉ số H của một số nhà khoa học nổi tiếng trên thế giới (tài liệu đã dẫn).

Vật lý	Hóa học	Khoa học máy tính
Ed Witten (Princeton) : 110	George Whitesides (Harvard) : 135	Hector Garcia Molina (Stanford) : 70
Martin Cohen (Berkeley) : 94	Elias James Corey (Harvard) : 132	Deborah Estrin (UCLA) : 68
Philip Anderson (Princeton) : 91	Martin Karplus (Harvard) : 129	Ian Foster (Illinois) : 67
Manuel Cardona (Max Planck) : 86	Alan Heeger (California) : 114	Scott Shenker (Berkeley) : 65
Frank Wilczek (MIT) : 68	Kurt Wurthrich (Switzerland) : 113	Don Towsley (Massachusetts) : 65 Jeffrey Ullman (Stanford) : 65

d) Chỉ số SJR

Chỉ số SJR (SCImago Journal Rank) được SCImago phát triển từ thuật toán Google PageRankTM. Chỉ số này cho thấy mức độ ảnh

hưởng của các tạp chí có trong cơ sở dữ liệu Scopus® từ năm 1996.

SCImago Journal Rank (chỉ số SJR) là một phép đo xác định ảnh hưởng khoa học của tạp

⁶ Đến đây ta dễ liên tưởng đến chỉ số trắc lượng NSP mà Marek Kosmulski xây dựng năm 2010 (sẽ được trình bày ở phần dưới).

⁷ Về SNIP và các chỉ số có liên quan có thể tham khảo thêm tại [13], [14] và tại Scopus.com

⁸ Nguồn tham khảo: Nguyễn Văn Tuấn. Đánh giá ảnh hưởng trong nghiên cứu khoa học qua chỉ số H. , <http://vietsciences.free.fr/khaocuu/congtrinhkhoaoc/H-index.htm>, <29/9/2014>

Nghiên cứu - Trao đổi

chí khoa học, trong đó, ngoài việc dựa vào các số liệu thống kê về trích dẫn (Chỉ số IF mặc định, chỉ số IF được xác định qua số liệu thống kê của 2 năm liền trước năm hiện tại), còn quan tâm tới các yếu tố mang tính tham biến phản ánh giá trị khoa học như tầm quan trọng, uy tín của các tạp chí công bố các bài viết đã trích dẫn đến. Nhìn chung, cách tính toán chỉ số SJR là phức tạp, hơn nữa, gần đây lại xuất hiện chỉ số SJR2 - loại chỉ số mới dựa trên SJR có bổ sung tham biến là các số liệu thống kê về cùng trích dẫn giữa các tạp chí được khảo sát. Việc trình bày hệ thống các công thức để xác lập chỉ số SJR là rất cồng kềnh và cần được thực hiện trong bối cảnh khác. Những người cần quan tâm sâu về chỉ số SJR có thể tham khảo giới thiệu về SJR tại SCImagojr.com hoặc các công trình có liên quan [6].

Kết quả xếp hạng các tạp chí thuộc lĩnh vực các khoa học thư viện và thông tin (LIS) dựa theo các **chỉ số SJR** và **IF** là khác nhau.

đ) Chỉ số Journal Immediacy Index (JImI)

Chỉ số JImI của một tạp chí, tính cho năm hiện tại N, được xác định theo công thức sau:

$$JImI(N) = \frac{A(N)}{B(N)}$$

Trong đó:

- A (N): Tổng số các trích dẫn trong năm hiện tại N đến các bài báo được công bố trong năm hiện tại N trên tạp chí.

- B(N) Số các bài báo được công bố trong năm hiện tại N trên tạp chí.

Chỉ số JImI của tạp chí sẽ phản ánh mức độ nhanh chóng trung bình được trích dẫn đến một bài báo của tạp chí trong một năm cụ thể. Chỉ số JImI của tạp chí có rất nhiều ý nghĩa. Trước hết nó chứng tỏ tạp chí đã được phổ biến rất nhanh chóng, kịp thời đến người đọc.

Sau đó, nó lại chứng tỏ tạp chí luôn thu hút được một cách rất hiệu quả sự quan tâm của người đọc. Một tạp chí như vậy được đánh giá là có uy tín.

e) Số bài báo thành công (NSP)

Năm 2011, Marek Kosmulski (Đại học Tổng hợp Công nghệ Lublin, Ba Lan) đã xây dựng chỉ số là Số các bài báo thành công (Number of Successful Papers - NSP) để đánh giá tạp chí khoa học [9]. Chỉ số này được xây dựng như sau:

Score_j = 1 khi c_j > r_j

Score_j = 0 trong các trường hợp khác, trong đó:

- c_j : Số lượt các tài liệu khác trích dẫn đến bài báo j.

- r_j : Số tài liệu mà bài báo j trích dẫn đến.

Thông thường chỉ số này được xác lập trong các quan hệ sau:

- Được thống kê trong một khoảng thời gian xác định (1 năm, 2 năm,... 5 năm hay thậm chí trong suốt thời gian tồn tại) của tạp chí đó.

- Xây dựng tỷ lệ giữa các bài báo thành công với tổng số các bài báo thuộc diện có thể trích dẫn được (Citable Paper) của tạp chí đó.

Tác giả đưa ra bốn lưu ý khi sử dụng chỉ số NSP như sau:

+ Các chỉ số trung bình của c_j và r_j giữa các lĩnh vực khoa học khác nhau sẽ khác nhau. Các chỉ số trung bình thường là cao tại các ngành, lĩnh vực có nhiều cán bộ khoa học và mang tính phổ biến trong xã hội. Trong trường hợp ngược lại, các chỉ số trung bình này là thấp hơn.

+ Các chỉ số trung bình của c_j và r_j giữa các quốc gia khác nhau. Các chỉ số trung bình này thường là cao đối với những tạp chí được xuất bản tại các nước giàu, có trình độ KH&CN phát triển. Trong trường hợp ngược lại, các chỉ số trung bình này là thấp hơn.

Nghiên cứu - Trao đổi

+ Các chỉ số c_j và r_j thường là cao đối với các bài báo dạng tổng quan (reviews), còn thường là thấp đối với các bài báo nghiên cứu gốc.

+ Khi xem xét các chỉ số c_j và r_j cần đặc biệt quan tâm tới loại tự trích dẫn (self-citation): Bài báo mà tỷ lệ tự trích dẫn cao thường tạo ra các giá trị ảo, thiếu chính xác, và vì vậy khi đòi hỏi về ngưỡng các chỉ số cho những bài thuộc loại này cần đặc biệt chú ý và cần được *trừ hao* một cách thỏa đáng.

Năm 2013, tại Hội nghị thường niên Hội sinh học tế bào Mỹ tổ chức tại San Francisco, 150 nhà nghiên cứu và 75 tổ chức khoa học tham gia Hội nghị đã ra tuyên bố DORA với khuyến cáo không tiếp tục sử dụng IF trong việc đánh giá khoa học⁹. Tuy vậy, theo chúng tôi, trên thực tế, rất khó bác bỏ được các thành tựu mà ISI (Mỹ) đã xây dựng trên nửa thế kỷ nay, những thành tựu đã nhận được sự hưởng ứng của đại đa số các cộng đồng khoa học trên thế giới, đặc biệt là các tổ chức thực hiện chức năng đánh giá khoa học có uy tín trên thế giới như Thomson Reuters, SCImago, Journal-Ranking... Ở đây, trong nhiều trường hợp, nếu như các IF không được trực tiếp sử dụng, thì người ta lại sử dụng các chỉ số được hình thành trên cơ sở là hàm số của IF. Vì vậy, về bản chất, các tổ chức này cũng sử dụng IF. Và vì thế, việc xuất bản các công trình khoa học trên các tạp chí khoa học có trong danh mục tạp chí được đưa vào CSDL của ISI vẫn luôn được đánh giá cao, và thậm chí trong không ít trường hợp, được xem là điều kiện cần và đủ của một nghiên cứu có chất lượng, vì thế, để được Thomson Reuters lựa chọn hàng năm vẫn là mục đích lớn và quan trọng của mọi tạp chí khoa học trên thế giới.

2. Ứng dụng trắc lượng thư mục tại các doanh nghiệp thông tin và xuất bản

Nguồn thông tin KH&CN dưới dạng các

CSDL đã có một bước phát triển rất mạnh mẽ về số lượng, quy mô và tính chất trong khoảng từ cuối những năm 1990, khi mà Internet trở thành môi trường lưu giữ, trao đổi và cung cấp thông tin của mọi lĩnh vực hoạt động, của mọi khu vực trên thế giới. Cũng vào thời kỳ này, đã xuất hiện những doanh nghiệp thông tin và xuất bản có quy mô thị trường rộng khắp trên thế giới. Thị trường thông tin khoa học, nhất là phân khúc các nguồn tin trực tuyến tăng trưởng với gia tốc lớn và xâm nhập vào mọi nơi có người làm việc trong lĩnh vực khoa học và công nghệ. Về cơ cấu, đa số các nguồn tin trực tuyến đều có thể được chia thành hai bộ phận: các CSDL thư mục và các tệp toàn văn tương ứng (dưới dạng tệp văn bản hay tệp pdf) và đến nay, hầu như chúng đều được lưu giữ và truyền trên Internet. Hai bộ phận này được tích hợp với nhau nhằm mục đích tạo sự thuận tiện cho người sử dụng. Các CSDL thư mục (chủ yếu là nguồn tin mở, khai thác miễn phí) thường thực hiện chức năng tiếp thị, cung cấp các thông tin thiết yếu giúp người sử dụng quyết định có truy cập, khai thác phần toàn văn của tài liệu (phần này chủ yếu là thu phí) hay không. Sự phát triển này còn được thể hiện ở chỗ, ngày nay, hầu hết các CSDL khoa học do các doanh nghiệp thông tin và xuất bản có uy tín trên thế giới đưa ra thị trường, ngoài chức năng kiểm soát nguồn tin khoa học phục vụ việc quản lý, khai thác, tìm kiếm thông tin, thì chúng đều thực hiện được chức năng *thống kê khoa học*, làm cơ sở cho việc đánh giá và xếp hạng khoa học. Lutz Bornmann đã khẳng định, theo ông và nhiều nhà khoa học khác: *Khoa học thực hiện việc đánh giá và bản thân khoa học cũng là đối tượng để đánh giá* [1]. Trong quá trình nghiên cứu để tìm ra phương pháp đánh giá của mình, thông qua việc khảo sát các số liệu về xuất bản tài liệu khoa học tại

⁹ Tham khảo: <http://am.ascb.org/dora/> và Tạp chí Khoa học và Công nghệ Việt Nam, các số 12, năm 2013 và số 16, năm 2014..

Ngô nghiên cứu - Trao đổi

một số trường đại học, tác giả đã đi đến kết luận: *Trắc lượng thư mục đã trở thành công cụ không thể thiếu trong việc đánh giá các trường đại học...*[1].

2.1. Ứng dụng tại Thomson Reuters¹⁰

Thomson Reuters là một doanh nghiệp thông tin, đánh giá khoa học lớn và có uy tín bậc nhất trên thế giới và được biết đến là người kế tục công việc mà ISI thực hiện trước đây với một quy mô rộng lớn hơn: Tính đến tháng 9 năm 2014, tổng số biểu ghi của CSDL Web of Science của Thomson Reuters là 90 triệu, với hơn một tỷ biểu ghi tài liệu tham khảo trích dẫn, mỗi năm gia tăng khoảng 65 triệu biểu ghi: Đây chính là CSDL

trích dẫn lớn nhất trên thế giới hiện nay¹¹.

Để cung cấp được các số liệu trích dẫn đối với các tài liệu trong một CSDL thông tin khổng lồ, đương nhiên cấu trúc CSDL ở đây cũng rất phức tạp. Sản phẩm đánh giá khoa học được Thomson Reuters đưa ra thị trường hàng năm là Journal Citation Report. Cấu trúc CSDL được mô tả dưới đây sẽ giúp hình dung được tính chất phức tạp của nó, một sự khác biệt căn bản so với các loại CSDL thư mục mà ta đã từng gặp [8].

Dưới đây là minh họa về thông tin trắc lượng thư mục mà CSDL do Thomson Reuters cung cấp phục vụ cho xây dựng báo cáo trích dẫn tạp chí (JCR)¹².

Mark	Journal Title	ISSN	Total Cites	Impact Factor	5-Year Impact Factor	Immediacy Index	Citable Items	Cited Half-life	Citing Half-life
	CA-CANCER J CLIN	0007-9235	16130	162,5	107,74	27,76	25	3,1	5,4
Cited Journal  Citing Journal  Source Data Journal Self Cites									
CITED JOURNAL DATA CITING JOURNAL DATA  IMPACT FACTOR TREND RELATED JOURNALS									

Hình trên cho thấy, Tạp chí CA-A CANCER JOURNAL FOR CLINICIANS có chỉ số IF (năm 2013) là cao nhất (162,5) theo số liệu của Journal Citation Report 2013.

2.2. Ứng dụng tại SCImagojr.com

SCImago Journal&Country Rank là một cổng thông tin bao gồm các tạp chí và các chỉ số quốc gia khoa học được phát triển từ những thông tin chứa trong CSDL Scopus (Elsevier BV).

Dữ liệu về tài liệu trên SCImagojr.com cung cấp các dữ liệu sau đây về mỗi tài liệu: Nhan đề, Quốc gia xuất bản, Phạm vi bao quát (xếp theo 4 hạng về độ gần gũi với chủ đề mà

tạp chí bao quát Qi, i = 1,4), H-Index, So sánh các chỉ số IF và SJR của tạp chí, So sánh các chỉ số trích dẫn và tự trích dẫn, So sánh trích dẫn trung bình và trích dẫn trung bình mở rộng của tài liệu, So sánh chỉ số IF trong 2, 3 và 4 năm, Số liệu về tính chất quốc tế của các tài liệu được công bố, So sánh tỷ lệ tài liệu có thể trích dẫn được và các tài liệu không thể trích dẫn được, So sánh tỷ lệ tài liệu được trích dẫn và tài liệu không được trích dẫn. Khoảng thời gian thống kê từ 1999-2013. Các số liệu so sánh trên được thể hiện dưới dạng đồ thị/biểu đồ.

¹⁰ Các thông tin phản ánh ứng dụng trắc lượng thư mục được tổng hợp từ các trang chủ của các doanh nghiệp này.

¹¹ <http://wokinfo.com/citationconnection/>

¹² Nguồn <http://admin-apps.webofknowledge.com/JCR/JCR?RQ=RECORD&rank=1&journal=CA-CANCER+J+CLIN,<7/10/2014>>

Nghiên cứu - Trao đổi

Bảng xếp hạng tài liệu

CSDL của SCImago ngoài việc cung cấp các thông tin thư mục về tài liệu, đã thực hiện việc xếp hạng tài liệu dựa trên các dữ liệu phản ánh: Chỉ số H-Index; Tổng số các bài báo trong năm liền trước năm hiện tại; Tổng số các bài báo trong 3 năm liền trước năm hiện tại; Tổng số các trích dẫn mà tài liệu đã trích dẫn trong năm liền trước năm hiện tại; Tổng số các trích dẫn đến tạp chí trong 3 năm liền trước năm hiện tại; Tổng số các tài liệu có thể trích dẫn được trong 3 năm liền trước năm hiện tại; Chỉ số IF mặc định tính cho năm hiện tại; Tỷ lệ trích dẫn đi và trích dẫn đến; Quốc gia xuất bản tài liệu.

Dữ liệu về quốc gia

CSDL của SCImago cung cấp các dữ liệu về mỗi quốc gia bao gồm: Chỉ số H, Tổng số tài liệu được xuất bản, số tài liệu có thể trích dẫn được, tổng số trích dẫn, Tổng số tự trích dẫn, Số trích dẫn trung bình của mỗi tài liệu, Tỷ lệ trung bình tự trích dẫn của tài liệu, Số các tài liệu được trích dẫn, Số các tài liệu không được trích dẫn, Các số liệu về Hợp tác quốc tế, So sánh giữa quốc gia với các nước thuộc vùng (mà SCImago phân nhóm), So sánh giữa quốc gia với thế giới, So sánh giữa tài liệu có thể trích dẫn và không thể trích dẫn, Tổng số các sản phẩm có liên quan (so sánh với thế giới), Số liệu phân bố tài liệu theo lĩnh vực khoa học. Các số liệu được thống kê từ 1996-2013 và được liệt kê theo từng năm.

Bảng xếp hạng các quốc gia

CSDL của SCImago cũng cung cấp bảng xếp hạng các quốc gia dựa trên các chỉ số sau đây: Tổng số tài liệu, Tổng số tài liệu có thể trích dẫn được, Tổng số trích dẫn, Tổng số tự trích dẫn, Chỉ số IF mặc định và Chỉ số H.

Ngoài Thomson Reuters và SCImago, còn có một số tổ chức khác như Journal-Ranking.com, Annual Review,... cũng triển khai dịch vụ cung cấp các số liệu để xếp hạng khoa học đối với các chủ thể khác nhau. Nguyên lý để triển khai dịch vụ này là dựa trên các phân tích trích dẫn theo các dạng, loại

khác nhau. Do khuôn khổ có hạn nên chúng tôi xin được chuyển phần giới thiệu các nội dung chi tiết này đến một dịp khác.

3. Nội dung trắc lượng thư mục trong chương trình đào tạo về thông tin- thư viện

Các trình bày nêu trên cho thấy, rõ ràng các vấn đề thư mục, phương pháp thư mục được ứng dụng rất phổ biến hiện nay. Nhờ vậy, đã tạo nên diện mạo mới cũng như những kết quả hết sức ấn tượng dựa trên *việc xử lý thư mục đối với tài liệu ở các mức độ khác nhau* mà các doanh nghiệp thông tin và xuất bản mang lại. Như đã nêu, các vấn đề của trắc lượng thư mục và ứng dụng của nó còn chưa nhận được sự quan tâm của đông đảo những người hoạt động trong lĩnh vực TT-TV nước ta. Các khoa, bộ môn giảng dạy về TT-TV tại các trường đại học có vai trò quan trọng để thay đổi tình trạng này, đưa các vấn đề liên quan tới thư mục trở thành một tâm điểm trên các diễn đàn về nghiên cứu thư mục nói riêng và các khoa học về thư viện và thông tin nói chung.

Hiện tại, trong chương trình đào tạo bậc đại học chuyên ngành TT-TV tại nhiều nơi, các môn học (hay chủ đề của môn học) có liên quan trực tiếp đến trắc lượng thư mục và các ứng dụng của nó có thể là: Cơ sở thông tin học; Thư mục học; Thư mục chuyên ngành (khoa học và công nghệ; khoa học xã hội...); Thư viện số; Công nghệ nội dung và công nghiệp nội dung; Sản phẩm và dịch vụ TT-TV; Các mạng thông tin, các nguồn tin trực tuyến, các nguồn thông tin dạng số; Marketing trong hoạt động thư viện

Trên cơ sở có các dấu hiệu liên quan về nội dung tri thức được cung cấp của các môn học trên, có thể tìm thấy các điểm/phần chung với trắc lượng thư mục. Từ những góc độ tiếp cận khác nhau, mức độ liên quan của trắc lượng thư mục với mỗi một trong số các môn học trên cũng được thể hiện khác nhau, ở những nội dung khác nhau, từ nghiên cứu cơ bản đến nghiên cứu ứng dụng, từ quá trình *sản xuất* đến công đoạn phổ biến và cung cấp sản phẩm, dịch vụ thông tin... Khi lồng ghép nội dung của trắc lượng thư mục vào mỗi

Nghiên cứu - Trao đổi

môn học, thì những bộ phận kiến thức của nó cũng có những trọng số khác nhau, để đảm bảo sự tích hợp được nói ở đây được nhuần nhuyễn, các nội dung mới được bổ sung mới không trở nên khiên cưỡng. Việc xây dựng chương trình có bổ

sung nội dung trắc lượng thư mục, ứng với mỗi vị trí (mỗi môn học), các trọng tâm được giới thiệu cũng có những khác biệt nhất định. Bảng 1 phản ánh sự khác biệt đó từ các kết quả nghiên cứu các tài liệu [4], [5], [7], [10] và đúc kết của chúng tôi.

Bảng 1: Những nội dung liên quan đến trắc lượng thư mục trong một số môn học

Môn học	Nội dung trọng tâm của TLTM cần được quan tâm trong môn học
Cơ sở Thông tin học	- Lịch sử TLTM; - Các định luật cơ bản trong TLTM; - Nội dung của TLTM; - Quan hệ với Scienmetrics, Informetrics.
Thư mục học	- Ứng dụng TLTM để phát triển phương pháp biên mục; - Chỉ dẫn trích dẫn, chỉ dẫn tham khảo; - Các phương pháp biên mục có ứng dụng TLTM.
Thư mục chuyên ngành (*)	- Các chức năng mới của thư mục; - CSDL thư mục; - Các hệ thống chỉ dẫn tham khảo.
Thư viện số	- Xây dựng và phát triển nguồn tin, nội dung số, TV số tích hợp vào Web of Science, Scopus.
Công nghiệp nội dung	- Các mô hình liên kết dữ liệu; - Vấn đề siêu dữ liệu; - Các phương pháp xử lý thông tin có ứng dụng TLTM.
SP&DV (*)	- Cách tạo lập các sản phẩm phục vụ thống kê khoa học, đánh giá khoa học
Mạng thông tin...(*)	- Khai thác các tính năng thống kê của các CSDL trên mạng. Nhấn mạnh khía cạnh trắc lượng web
Marketing	- Xây dựng chiến lược marketing phục vụ đa mục đích: kiểm soát nguồn tin; - Quản lý và đánh giá khoa học tại thư viện đại học và cơ quan thông tin, tại các doanh nghiệp thông tin và xuất bản

Chú ý 1. Đối với chương trình đào tạo bậc Thạc sỹ và Nghiên cứu sinh, thích hợp nhất là nội dung trắc lượng thư mục được lồng ghép vào môn học Cơ sở thông tin học hoặc Thư mục học. Tại đây, cũng cần mở rộng ra các chuyên ngành nghiên cứu mới và có liên quan mật thiết là Scienmetrics và Informetrics.

Chú ý 2. Các môn học được đánh dấu (*) bổ sung có thể được thực hiện đối với chương trình đào tạo bậc Cao đẳng. Khi đó nội dung của trắc lượng thư mục có thể được

giới thiệu chủ yếu dưới góc độ tạo cho sinh viên các kỹ năng cần thiết để hiểu và khai thác, sử dụng các chức năng thông kê của các CSDL loại này.

Kết luận

Hiện nay ở nước ta, việc đổi mới giáo dục và đào tạo một cách toàn diện, có hệ thống đang thu hút sự quan tâm của toàn xã hội, là một trong số các nội dung được phản ánh sâu đậm trong các văn bản nghị quyết của Đảng và Nhà nước... Công tác đào tạo tại các trường đại học vì thế đang có những

Nghiên cứu - Trao đổi

thay đổi sâu sắc và đa dạng, từ hình thức (phương thức đào tạo theo tín chỉ, việc học, việc nghiên cứu, đào tạo từ xa, e-learning...) cho tới nội dung (chú trọng cập nhật những thông tin mới nhất, theo hướng hội nhập từng bước vào các chương trình đào tạo tiên tiến trên thế giới, chú trọng cung cấp các tri thức, kinh nghiệm có khả năng ứng dụng cao vào thực tiễn, sự gắn kết giữa kiến thức và kỹ năng ...).

Để thực hiện được những đổi mới với các yêu cầu và chuẩn mực rất cao này, các trường đại học phải không ngừng đổi mới cách thức và chương trình giảng dạy của mình. Các khoa/bộ môn đào tạo trong lĩnh vực thông tin, thư viện tại các trường đại học, cao đẳng cũng cần phải bước vào guồng máy đổi mới, đi trên con đường đổi mới chung này. Lý luận và thực tiễn trong lĩnh vực thông tin, thư viện trên mọi phạm vi đã có những bước nhảy vọt từ cuối thế kỷ XX. Trên thực tế, các chương trình đào tạo

tại các khoa/bộ môn này cũng đã có nhiều thay đổi rõ rệt. Tuy nhiên, so với yêu cầu thực tiễn, giữa chúng còn có một khoảng cách đáng kể. Điều đó được phản ánh qua trường hợp các vấn đề của trắc lượng thư mục chưa được chúng ta quan tâm ở mức cần phải có, đã làm cho người sinh viên chưa đủ điều kiện để thực sự đảm trách được các vị trí chuyên môn của mình theo hướng hội nhập. Công việc của các tổ chức nghiên cứu, đào tạo ở đây là thu hẹp lại khoảng cách đó bằng nhiều giải pháp và cách thức khác nhau, và việc không ngừng đổi mới chương trình đào tạo là nội dung không thể thay thế. Tích hợp các tri thức và thực tiễn ứng dụng trắc lượng thư mục theo các phương án đề xuất trên đây là một trong các kiến nghị mà chúng tôi hy vọng sẽ thu hút sự quan tâm của các đồng nghiệp. Chúng tôi mong nhận được các ý kiến trao đổi thảo luận về vấn đề trên qua địa chỉ tmtuanissi@gmail.com. Xin cảm ơn.

Tài liệu tham khảo

1. Bornmann L. The Problem of Citation Impact Assessments for Recent Publication Years in Institutional Evaluations.// Journal of Informetrics. 2013. Vol. 7. pp. 722-729.
2. Bouyssou D., Marchant T. Bibliometric rankings of journals based on Impact Factors: An axiomatic approach.// Journal of Informetrics. 2011. Volume 5, Issue 1, January. pp. 75-86.
3. Cheng Su ... PrestigeRank: A new evaluation method for papers and journals.// Journal of Informetrics. 2011. Volume 5, Issue 1, January. pp. 1-13
4. Chuyên đề về trắc lượng thư mục.// Thông tin & Tư liệu. 2009. số 2. tr. 1-25.
5. The Concept Citation Indexing: Unique and Innovative Tool for Navigating the Research Literature
http://www.thomsonreuters.com/business_units/scientific/free/essays/history/,
6. Guerrero-Bote V.P., Moya Anegón F. A Further Step forward in Measuring Journals' Scientific Pertige: The SJR2 Indicator.// Journal of Informetrics. 2012. Vol. 6. pp 674-688.
7. Hahn K.L. Research Library Publishing Services: New Options for University Publishing.- Washington: ARL. 2008. 41 p.
8. Journal Citation Reports on the Web 4.0
http://www.sris.com.tw/download/workbook/en/jcr4_sem_workbook0305.pdf
9. Kosmulski M. Successful Papers: A New Idea in Evaluation of Scientific Output.// Journal of Informetrics. 2011, Vol 5. pp. 481-485.
10. Moed H.F. Measuring contextual Citation Impact of Scientific Journals. Journal of Informetrics. 2010. Vol. 4. pp. 265-277.
11. Thelwall M. Bibliometrics to webometrics
<http://jis.sagepub.com/cgi/content/abstract/34/4/605>,
12. The Thomson Reuters Impact Factor .
http://thomsonreuters.com/products_services/science/free/essays/impact_factor/
13. Trần Mạnh Tuấn. Tiêu chuẩn quốc tế đối với tạp chí khoa học và việc áp dụng tại Viện Khoa học Xã hội Việt Nam./ Chủ nhiệm Trần Mạnh Tuấn: Chương trình cấp Bộ nghiên cứu và thông tin khoa học năm 2011-2012.- H.: Viện Khoa học Xã hội Việt Nam. 2012. 207 tr.
14. Waltman L., etc. Some Modifications to the SNIP Journal Impact Indicator.// Journal of Informetrics. 2013. Vol. 7. Issue 2. pp 272-285.

(Ngày Tòa soạn nhận được bài: 16-9-2014; Ngày phản biện đánh giá: 12-11-2014; Ngày chấp nhận đăng: 01-12-2014).