

Đánh giá định lượng kết quả nghiên cứu khoa học

(Theo bài viết của Hội đồng chức danh giáo sư nhà nước: http://hdgsnn.gov.vn/tin-tuc/danh-gia-dinh-luong-ket-qua-nghien-cuu-khoa-hoc_425)

12/03/2019

1. Giới thiệu

Các hoạt động liên quan đến khoa học và công nghệ gồm nghiên cứu, ứng dụng và sản xuất, trong đó hoạt động nghiên cứu thường được chia ra ba loại hình: nghiên cứu cơ bản, nghiên cứu ứng dụng và nghiên cứu phát triển. Nghiên cứu cơ bản nhằm tìm ra những tri thức khoa học nền tảng về thiên nhiên và xã hội, như việc chứng minh ước đoán Poincaré trong toán học hay việc xác định các gien gây ra bệnh tật. Nghiên cứu ứng dụng nhằm tìm ra tri thức khoa học mới và cần cho các nhu cầu thực tế như cách làm ra nước ngọt ở vùng nước mặn hay cách chẩn đoán cấp độ bệnh viêm gan dựa trên xét nghiệm máu nhưng không làm sinh thiết. Nghiên cứu phát triển nhằm tìm ra tri thức để làm ra các sản phẩm cụ thể, như việc làm hệ nhận dạng chữ Việt VnOCR hay việc hãng Microsoft nghiên cứu làm hệ điều hành máy tính Windows 7. Kết quả chủ yếu của nghiên cứu cơ bản và nghiên cứu ứng dụng là các ấn phẩm khoa học (chủ yếu là các bài báo, ngoài ra là sách chuyên khảo, báo cáo kỹ thuật, ...) hoặc bằng sáng chế phát minh, trong khi kết quả chủ yếu của nghiên cứu phát triển là các tri thức không công bố, tiềm ẩn trong các sản phẩm có giá trị sử dụng.

Ở các nước công nghiệp tiên tiến, nghiên cứu phát triển thường chiếm tỷ lệ khoảng hai phần ba toàn bộ hoạt động và kinh phí nghiên cứu, là loại hình nghiên cứu chủ yếu của các doanh nghiệp và thường có tỷ lệ cao nhất ở các viện nghiên cứu. Trong khi đó nghiên cứu cơ bản và nghiên cứu ứng dụng là các loại hình chủ yếu của đại học và có tỷ lệ cao ở các viện nghiên cứu [5]. Nghiên cứu phát triển ở ta có tỷ lệ còn thấp do phần lớn các doanh nghiệp chưa đầu tư cho nghiên cứu khoa học và công nghệ, và sâu xa hơn do ta hầu như chưa có công nghiệp chế tạo. Bài viết này trao đổi việc đánh giá định lượng kết quả nghiên cứu khoa học và công nghệ- một vấn đề gần đây được đề cập nhiều, cả trong và ngoài nước.

Kết quả nghiên cứu cơ bản và ứng dụng của một cá nhân, một đại học hay viện nghiên cứu, ... được đánh giá bởi *số lượng* ấn phẩm và *chất lượng* của chúng. Số lượng các ấn phẩm khoa học có thể đếm được khá dễ, nhưng đánh giá chất lượng của chúng lại không đơn giản. Có hai phương pháp làm việc này, một là *đánh giá chủ quan* qua một hệ thống bình duyệt bởi con người (peer review), và hai là *đánh giá khách quan* dựa trên các độ đo được tính toán tự động. Đánh giá chủ quan bởi hệ thống bình duyệt của con người, còn gọi *đánh giá định tính*, có thể cho phân tích sâu và xác đáng nhưng tốn kém tiền bạc, cần nhiều thời gian, và phụ thuộc nhiều vào chủ quan và hiểu biết của người đánh giá. Đánh giá khách quan, còn gọi *đánh giá định lượng*, dựa trên các độ đo (metrics) xác định từ các nguồn dữ liệu khoa học, được thực hiện tự động nên nhanh và rẻ, cung cấp những thông tin hữu ích, nhưng cũng dễ bị hiểu chưa chính xác và giải thích chưa thích hợp.

Ba độ đo đánh giá định lượng đang được dùng phổ biến gồm *chỉ số trích dẫn* (citation index) cho các ấn phẩm khoa học, *hệ số ảnh hưởng* (impact factor) cho các tạp chí, và gần đây là *chỉ số H* (h-index) cho các nhà khoa học. Cần chú ý là chỉ gần đây các độ đo này mới có thể tính được tự động do tiến bộ của công nghệ thông tin, đặc biệt là Web. Thêm nữa, cả hệ số ảnh hưởng và chỉ số H

đều được định nghĩa dựa trên chỉ số trích dẫn (citation-based metrics), do đó mang theo mọi hay dở của chỉ số trích dẫn. Việc hiểu rõ những điểm hay và hạn chế của các độ đo định lượng đang được bàn luận nhiều trên báo chí khoa học quốc tế, được quan tâm trong giới khoa học ở nhiều nước. Hiện nay nhiều tổ chức và quốc gia có xu hướng dùng các phương pháp đánh giá định lượng, khách quan để bổ sung hoặc là cách thông dụng thay cho đánh giá định tính, chủ quan.

Bài báo này giới thiệu ba độ đo tiêu biểu kể trên, cung cấp các thông tin chọn lọc quan trọng để hiểu chúng, nhấn mạnh những đặc điểm cần chú ý và nêu một số ý kiến bàn luận.

2. Về các độ đo

2.1. Chỉ số trích dẫn

Chỉ số trích dẫn (citation index) của một ấn phẩm, do Eugene Garfield đề xuất năm 1955, là số lần ấn phẩm này được trích dẫn, được tham khảo trong tất cả các ấn phẩm khác [7]. Từ đó đến nay, chỉ số trích dẫn đã được dùng làm một độ đo quan trọng để đánh giá các công trình nghiên cứu, là cơ sở để định nghĩa các độ đo khác cho các tạp chí và nhà khoa học. Câu hỏi có thể làm ta ngạc nhiên là tại sao một chỉ số đơn giản như vậy lại được dùng rộng rãi cho đến nay để đo chất lượng và giá trị các công trình khoa học? Có thể nói chỉ số trích dẫn được “tin dùng” do dựa trên một giả định được thừa nhận rộng rãi, là các nhà khoa học có ảnh hưởng hơn, các công trình quan trọng và có giá trị sử dụng hơn thường được trích dẫn nhiều hơn. Nói nôm na, chỉ số trích dẫn đo mức độ “hữu xạ tự nhiên hương” của các ấn phẩm.

Đặc điểm đáng chú ý nhất là *chỉ số trích dẫn chỉ có ý nghĩa so sánh trong từng ngành khoa học*.

Điều đã biết này được khảo sát định lượng qua thống kê số lần trích dẫn trung bình của các bài báo trong các ngành khoa học. Theo [1], các bài báo trong các ngành khoa học về sự sống (life sciences, như sinh học phân tử và tế bào, y sinh học) có trung bình khoảng 6 trích dẫn, trong vật lý và hóa học khoảng 3 trích dẫn, trong toán học, tin học và khoa học xã hội khoảng 1 trích dẫn (hình vẽ). Theo các số liệu thống kê trên – để có một định ý và giả sử số trích dẫn trong các ngành tăng một cách tuyến tính – trong một chừng mực nào đấy có thể xem bài báo ngành toán có trích dẫn mười lần là được trích dẫn nhiều trong ngành này như một bài trong ngành vật lý được trích dẫn khoảng ba chục lần hay một bài trong ngành khoa học sự sống được trích dẫn khoảng sáu chục lần.

Có nhiều lý do dẫn đến sự khác biệt lớn như vậy giữa các ngành về chỉ số trích dẫn của các ấn phẩm khoa học, mà chủ yếu do khác biệt về “văn hóa ngành”. Văn hóa này phụ thuộc vào bản chất khoa học, cách làm và cách công bố nghiên cứu. Trong khi cần những nghiên cứu nghiêm túc về khác biệt chỉ số trích dẫn trung bình giữa các ngành, người viết nêu ý kiến riêng của mình dưới đây và cho rằng mỗi người đọc nên thử tự lý giải điều này. Các nghiên cứu thực hiện trên các ý tưởng trừu tượng, lập luận và tính toán như trong toán học, vật lý lý thuyết hoặc tin học, thường ít liên quan, ít “dựa” hơn vào các nghiên cứu khác cùng lĩnh vực. Các nghiên cứu chủ yếu bằng thực nghiệm, thường cần nhiều liên hệ và so sánh với các nghiên cứu cùng lĩnh vực về phương pháp và kết quả, và do vậy khi khẳng định tính mới mẻ của kết quả và đề thuyết phục cần đưa ra nhiều trích dẫn (chứng cứ) hơn.

Đặc điểm quan trọng thứ hai cần biết rõ là *các chỉ số trích dẫn tính được từ các nguồn khác nhau thường khác nhau và có sai số*.

Sau khi nêu khái niệm chỉ số trích dẫn, Garfield xây dựng Viện Khoa học Thông tin ISI (Institute for Scientific Information) – gần đây sát nhập vào tập đoàn Thomson Reuters – và thiết lập các cơ sở dữ liệu ISI, trong đó tiêu biểu là:

- Cơ sở dữ liệu Chỉ số Trích dẫn Khoa học SCI (Science Citation Index), từ 1964, hiện có 3773 tạp chí của 100 ngành và cơ sở dữ liệu SCIE (SCI mở rộng, Science Citation Index Expanded) với 8207 tạp chí của 150 ngành;
- Cơ sở dữ liệu Chỉ số Trích dẫn Khoa học Xã hội SSCI (Social Sciences Citation Index), từ 1973, hiện có 2697 tạp chí và 3500 công trình của 50 ngành;
- Cơ sở dữ liệu Chỉ số Trích dẫn Nghệ thuật và Nhân văn A&HCI (Arts & Humanities Citation Index), từ 1978, hiện có 1470 tạp chí và 6000 công trình khác.

Ngoài ra, cũng cần kể đến cơ sở dữ liệu Chỉ số Trích dẫn Tuyển tập Hội nghị CPCI (Conference Proceedings Citation Index) chứa thông tin của hơn 110,000 tuyển tập hội nghị kể từ năm 1990 của 256 ngành thuộc về khoa học tự nhiên và khoa học xã hội nhân văn (chú ý là uy tín của các tạp chí ISI và hội nghị ISI khác nhau đáng kể). Các cơ sở dữ liệu ISI tuyển chọn các tạp chí ảnh hưởng nhất của các ngành. Từ 1997, bảy cơ sở dữ liệu ISI được chuyển lên mạng dưới tên *Web of Science* (<http://isiwebofknowledge.com>).

Quảng trên dưới mười năm trở lại đây, Web đã cho ra đời hơn 100 cơ sở dữ liệu và công cụ cho phép tìm kiếm chỉ số trích dẫn, như arXiv, CiteSeer, ScienceDirect, SciFinder Scholar, PubMed, ... Trong số này, *Scopus* của Elsevier (<http://info.scopus.com>, từ 2004) và *Google Scholar* của Google (<http://scholar.google.com>, từ 2005) cùng với Web of Science đang là ba hệ phổ biến nhất [3]. Scopus chứa thông tin của 16.500 tạp chí, 600 ấn phẩm nghề nghiệp, 350 loạt sách chuyên khảo, khoảng 3,6 triệu bài báo từ hội nghị. Google Scholar chứa thông tin của hầu hết các tạp chí có thẩm định của các nhà xuất bản lớn tại châu Mỹ và châu Âu, các báo cáo kỹ thuật, luận văn, sách và nhiều loại tài liệu khác (Google Scholar không công bố danh sách các tạp chí của mình).

Cần lưu ý là chỉ số trích dẫn của mỗi ấn phẩm khoa học tính từ các nguồn kể trên thường khác nhau vì chúng có số lượng tạp chí, kỷ yếu hội nghị, ... khác nhau. Một thí dụ là cuốn sách *Quantum Computation and Quantum Information* của M. Nielsen và I. Chuang (xuất bản năm 2000, Cambridge University Press). Tính đến năm 2007, từ Web of Science cuốn sách này được trích dẫn hơn 2800 lần, nhưng từ Scopus số trích dẫn là 3150, và từ Google Scholar có 4300 trích dẫn [14]. Một khảo sát khác phân tích chỉ số ảnh hưởng của 328 bài báo từ ba tạp chí y học hàng đầu trong thời gian 6 tháng của mười năm về trước [11]. Các tác giả chỉ ra số lượng trích dẫn của các bài báo này từ ba nguồn kể trên là rất khác nhau: từ Web of Science có 68.088 trích dẫn, từ Scopus có 82.076 trích dẫn và từ Google Scholar có 83.538 trích dẫn (gấp 1.226 lần so với Web of Science).

Các công cụ tìm kiếm (search engine) thường cho kết quả đúng về chỉ số trích dẫn của một bài báo trên một cơ sở dữ liệu cố định do tên của bài báo thường là duy nhất và xác định, nhưng không luôn cho kết quả đúng với các độ đo liên quan tới một tác giả cụ thể như tổng số trích dẫn hoặc số

ấn phẩm của một tác giả có trích dẫn nhiều hơn một ngưỡng nào đó (như chỉ số H), do nhiều tác giả có thể có tên trùng nhau hoặc giống nhau khi viết tắt.

Có hai độ đo cơ bản về chất lượng của các công cụ tìm kiếm là *độ chính xác* (precision) và *khả năng tìm hết* (recall). Độ chính xác là tỷ lệ của số tài liệu tìm được và tìm đúng trên số tài liệu tìm được, còn khả năng tìm hết là tỷ lệ của số tài liệu tìm được và tìm đúng trên toàn bộ số tài liệu cần tìm. Các công cụ tìm kiếm hiện nay cho kết quả của hai độ đo này chưa cao. Trong việc tính hệ số trích dẫn, độ chính xác thấp chủ yếu do các hệ tìm kiếm tự động hiện chưa phân biệt được các tác giả có tên trùng nhau hay giống nhau, và khả năng tìm hết của các hệ còn thấp chủ yếu do các cơ sở dữ liệu hiện nay không có được tất cả các ấn phẩm có trích dẫn đến bài báo đang xem xét.

Độ chính xác khi tìm chỉ số trích dẫn của các tác giả có tên phổ biến nói chung thấp hơn của các tác giả có tên ít gặp. Thí dụ khi tìm trên Google Scholar tổng số bài được trích dẫn của tác giả Nguyễn Anh Tuấn, ta gõ tên “Nguyen Anh Tuan” và hệ tìm được 100 bài báo có trích dẫn. Thường thì không phải tất cả 100 bài này đều của Nguyễn Anh Tuấn. Giả sử trong số 100 này chỉ đích thực có 60 trong số tất cả 80 bài có trích dẫn của Nguyễn Anh Tuấn, và trong số 40 bài còn lại có 5 bài của tác giả Nguyễn Ánh Tuấn, 35 bài của tác giả Nguyễn Anh Tuấn. Khi này, độ chính xác về trích dẫn của Nguyễn Anh Tuấn do hệ tìm được là $60/100 = 0.6$ và khả năng tìm hết là $60/80 = 0.75$. Khi tìm trên Google Scholar chẳng hạn cho tác giả Khuất Phương Trường (Khuat Phuong Truong), độ chính xác thường sẽ cao hơn.

2.2. Hệ số ảnh hưởng của tạp chí

Hệ số ảnh hưởng (impact factor, viết tắt IF) của một tạp chí được định nghĩa, thừa nhận và dùng rộng rãi lâu nay. Hệ số này của một tạp chí thay đổi theo từng năm, và hệ số ảnh hưởng của một tạp chí T trong năm N được tính bằng tỷ số A/B, trong đó A là tổng số lần trích dẫn, tính trong tất cả các ấn phẩm của năm N, đến các bài đăng trên T trong hai năm liên tiếp ngay trước N, và B là tổng số các bài đăng trên T trong hai năm này. Nếu trong hai năm 2007 và 2008 tạp chí T đăng tất cả 100 bài báo, và có 250 lần các bài trong số 100 bài này của T được trích dẫn trong tất cả các bài ở các tạp chí, hội nghị, ... của năm 2009, thì hệ số ảnh hưởng của T trong năm 2009 sẽ là $250/100 = 2.5$. Tạp chí Physical Review Letters có IF năm 2009 là 7,180 có nghĩa là về trung bình mỗi bài báo của tạp chí này công bố năm 2007 và 2008 được trích dẫn 7,180 lần trong năm 2009. Người ta thường chỉ nói hệ số ảnh hưởng của một tạp chí và không nêu cụ thể một năm nào đấy. Nhưng thực ra hệ số này có thể thay đổi rất nhiều theo thời gian, thí dụ tạp chí *Bioinformatics* có IF theo ISI là 4,328 vào năm 2008, 4,894 năm 2007, 5.742 năm 2004, 6.701 năm 2003, 4.615 năm 2002, và 3.421 năm 2001.

Hệ số ảnh hưởng của tạp chí được dùng cho nhiều mục đích, như cho biết uy tín và sự phát triển của tạp chí, nhà khoa học chọn tạp chí gửi bài, nhà quản lý dùng để đánh giá hiệu quả nghiên cứu của các nhà khoa học, như cơ sở để xét biên chế, giải thưởng, cấp duyệt kinh phí. Hệ số ảnh hưởng của tạp chí còn được dùng để đánh giá các khoa, trường và viện nghiên cứu, đo hiệu quả khoa học của các quốc gia.

Điều đầu tiên cần biết rõ là *hệ số ảnh hưởng của tạp chí cũng khác nhau giữa các ngành*. Chẳng hạn theo JCR (Journal Citation Reports) của Web of Science, vào năm 2008 tạp chí của ngành y có IF cao nhất là 74,575 (CA: A Cancer Journal for Clinicians của Hội Ung thư Mỹ), IF thứ nhì là

50,017 (The New England Journal of Medicine), ... và IF thứ 100 cũng là 3,733 (Epilepsia, xếp thứ 739 trong toàn bộ tạp chí của JCR). Trong ngành toán lý thuyết, tạp chí có IF cao nhất là 3,806 (Communications on Pure and Applied Mathematics, xếp thứ 711 trong JCR), thứ nhì là 3,5 (Bulletin of the American Mathematical Society, xếp thứ 851 trong JCR), ... và thứ 100 là 0,584 (Monatshefte Fur Mathematik, xếp thứ 5248 trong JCR). Có sự khác biệt này là điều dễ hiểu, vì hệ số ảnh hưởng của tạp chí được tính dựa trên chỉ số trích dẫn của các bài báo của tạp chí, và như đã phân tích ở phần trên, chính sự khác biệt của “văn hóa ngành” đã tạo ra số trích dẫn rất khác nhau này.

Trong [2], các tác giả khảo sát sự khác nhau của hệ số ảnh hưởng của tạp chí theo thời gian và giá trị IF trung bình của tạp chí trong các ngành. Hình bên trích từ [2] cho thấy hệ số ảnh hưởng trung bình của các tạp chí trong ngành sinh học phân tử và tế bào là 4,763, trong y học là 2,896, trong hóa học là 2,61, trong vật lý là 1,912, trong tin học và toán học tương ứng là 0,631 và 0,566. Một cách giải thích nôm na có thể chấp nhận trong một chừng mực nào đấy về những con số này là—giả sử IF tăng tuyến tính trong các ngành— một tạp chí có IF khoảng 9,5 trong ngành sinh học phân tử và tế bào có ảnh hưởng ở ngành này quăng như ảnh hưởng của một tạp chí có IF 6 trong ngành y (truyền thông), một tạp chí có IF 4 trong ngành vật lý hay một tạp chí có IF quăng 1,2 trong ngành toán và tin học.

Mặc dù được dùng rộng rãi lâu nay, cách tính hệ số ảnh hưởng của tạp chí có một số hạn chế [6], [14], tiêu biểu là:

- Hệ số ảnh hưởng của một tạp chí chỉ cho ta giá trị trung bình về ảnh hưởng của các bài báo trong tạp chí đó. Giá trị trung bình này thường bị ảnh hưởng rất nhiều bởi một số ít bài có trích dẫn cao (như các bài đánh giá tổng quan) hoặc bởi quá nhiều bài không có hoặc có trích dẫn thấp. Một nghiên cứu gần đây chỉ ra rằng trong công thức tính hệ số ảnh hưởng IF, nếu các bài được xếp theo thứ tự của số trích dẫn, thì 15% bài báo đầu đóng góp 50% số trích dẫn, 50% bài đầu đóng góp 90% số trích dẫn, và các bài ở nhóm 50% đứng đầu có trích dẫn khoảng 10 lần hơn các bài ở nhóm 50% đứng cuối. Rõ ràng, hệ số ảnh hưởng của một tạp chí không phản ánh chính xác được ảnh hưởng của từng bài cụ thể đăng trong tạp chí này.
- Việc dùng trích dẫn trong thời gian 2 năm sau khi công bố (citation window) để tính IF là ngắn và sớm đối với một số ngành, tức công thức này chưa tính được hệ số ảnh hưởng thật của nhiều tạp chí (gần đây một vài hệ thống như JCR có đưa thêm vào hệ số ảnh hưởng tính trong thời gian 5 năm, ở đó hệ số ảnh hưởng của tạp chí trong một số ngành tăng lên và trong một số ngành giảm đi rõ rệt).
- Hệ số ảnh hưởng này chưa tính được đến những bài của một tạp chí được dùng nhưng không được trích dẫn.
- Hệ số ảnh hưởng của tạp chí biến động đáng kể từ năm này qua năm khác, biến động nhiều hơn ở các tạp chí nhỏ hơn (tuy hệ số ảnh hưởng hay được nói đến như một giá trị không đổi) [2]. Thí dụ tạp chí Bioinformatics kể trên có IF năm 2003 lớn gần gấp đôi IF năm 2001.

- Chỉ một phần nhỏ các tạp chí tham gia vào việc tính hệ số ảnh hưởng. Có tất cả khoảng trên 100 nghìn tạp chí các loại trên đời và các ấn phẩm ở bất kỳ tạp chí nào trong chúng cũng đều phải trích dẫn, nhưng các hệ cơ sở dữ liệu của ISI, Scopus, ... chỉ chứa khoảng 10-15 nghìn tạp chí hàng đầu của các ngành. Thêm nữa, các tạp chí không xuất bản bằng tiếng Anh hoặc tạp chí của các ngành “thiếu số” cũng ít khả năng nằm trong số này, và do vậy việc tính chỉ số trích dẫn và hệ số ảnh hưởng rõ ràng chưa toàn vẹn.

Chính Campbell, trưởng ban biên tập của tạp chí danh giá Nature, cũng cho rằng nên quan tâm đến chính giá trị của bài báo hơn là việc bài báo được đăng ở đâu [6].

2.3. Chỉ số H

Nếu chỉ số trích dẫn được dùng để “đo” các bài báo và hệ số ảnh hưởng “đo” các tạp chí, người ta còn muốn có những độ đo cho chính người làm nghiên cứu. Một cách là dựa trên số lượng các ấn phẩm và chất lượng của chúng qua chỉ số trích dẫn hoặc ảnh hưởng của nơi chúng được công bố. Một trong các độ đo là *chỉ số H* (h-index) đề nghị bởi nhà vật lý J.E. Hirsch vào năm 2005 [9], định nghĩa như sau: Một người có chỉ số H là N nếu đã xuất bản N bài báo có chỉ số trích dẫn ít nhất là N , và các bài còn lại có chỉ số trích dẫn nhiều nhất là N . Nếu một người công bố 40 bài báo, trong đó 9 bài có nhiều hơn 10 trích dẫn, 5 bài có 10 trích dẫn và 26 bài còn lại có ít hơn 10 trích dẫn, thì chỉ số H của người này là 10.

Ưu điểm cơ bản của chỉ số H là việc tính đến sự cân bằng giữa số lượng và chất lượng các công trình của người làm nghiên cứu, và tính toán đơn giản (như hệ QuadSearch dựa trên Google Scholar khá dễ dùng <http://delab.csd.auth.gr/~lakritid/index.php?lan=1&s=2>). Với những ưu điểm này, chỉ số H nhanh chóng được dùng phổ biến. Tuy nhiên, có những đặc điểm của chỉ số H cần được nhận biết rõ.

- Chỉ số H dựa trên chỉ số trích dẫn, và do đó chịu ảnh hưởng của toàn bộ những hạn chế của chỉ số trích dẫn nêu ra ở phần trên. Trước hết do sự khác biệt về chỉ số trích dẫn ở các ngành, chỉ số H có ý nghĩa hơn với những người thuộc cùng một ngành và khi so sánh những người khác ngành cần chú ý đến sự khác biệt này. Thêm nữa cần chú ý rằng khi tính tự động chỉ số H , độ chính xác và khả năng tìm hết thường không cao. Do sự trùng tên họ của người châu Á khá phổ biến, những chỉ số H được tính ra (nếu không kiểm chứng kỹ) thường cao hơn giá trị thật.
- Chỉ số H đánh giá thành quả một nhà khoa học nhưng không phân biệt được sự đóng góp khác nhau của các tác giả trong một công trình. Đây là một chuyện không đơn giản. Trong [16], các tác giả phân ra bốn cách viết thứ tự tên tác giả trong các ấn phẩm khoa học: Một là theo thứ tự đóng góp do các tác giả xác định (quence- determined credit); Hai là theo thứ tự chữ cái của tên khi các tác giả xem đóng góp mọi người như nhau (equal contribution norm); Ba là theo thứ tự “đầu-cuối” với nhấn mạnh rằng tác giả đầu (thường là nghiên cứu sinh) và cuối (thường là thầy cô hướng dẫn hay người phụ trách đề tài) là quan trọng nhất, và những người ở giữa theo đóng góp; Bốn là theo thứ tự đóng góp do các tác giả xác định một cách định lượng, (percent contribution indicated). Cách tính chỉ số H không phân biệt được sự đóng góp của các tác giả, đặc biệt trong các ngành có nhiều tác giả trong mỗi ấn phẩm.

- Chỉ số H không đánh giá được những tác giả có một số ít công trình ảnh hưởng rất lớn, điều thường có với những nhà khoa học trẻ xuất sắc, chẳng hạn giáo sư Ngô Bảo Châu có chỉ số H chỉ là 6.
- Chỉ số H không phân biệt được các trích dẫn phê bình. Trong [13], Lawani tóm tắt sáu lý do trích dẫn một bài báo: (a) Ghi nhận công trạng của tác giả, (b) Thể hiện sự kính trọng tác giả, (c) Nói về phương pháp liên quan, (d) Cung cấp thông tin nền có ích, (e) Trích dẫn để phê phán, (f) Trích dẫn để làm cơ sở cho bài báo. Như vậy các trích dẫn bởi lý do (e) không tăng mà chỉ giảm giá trị của bài báo, nhưng vẫn luôn được tính thêm vào thành tích của tác giả khi tính chỉ số H.

Có những cố gắng để khắc phục các hạn chế kể trên của chỉ số H, như chỉ số G (g-index) đề xuất năm 2006 bởi Leo Egghe. Giả sử các ấn phẩm của một tác giả được xếp theo thứ tự giảm dần của chỉ số trích dẫn, khi này chỉ số G chính là số lớn nhất sao cho G bài báo đầu tiên có trích dẫn trung bình lớn hơn G. Chỉ số G này đề cao giá trị của những bài báo có rất nhiều trích dẫn trong đánh giá một tác giả.

3. Một vài nhận xét và ý kiến

Phần này nêu vài nhận xét về những độ đo đánh giá định lượng và vài ý kiến bàn luận.

(1) Cần hiểu rõ đặc điểm và dùng đúng các độ đo đánh giá định lượng: Các độ đo đánh giá định lượng kết quả nghiên cứu và việc có thể dùng chúng dễ dàng là một bước tiến rất lớn. Tiến bộ này cho phép người quản lý và nhà khoa học có thể có những nhìn nhận sâu hơn và chính xác hơn về hoạt động nghiên cứu. Khái niệm các độ đo về chỉ số trích dẫn, hệ số ảnh hưởng và chỉ số H đều đơn giản và dễ hiểu, nhưng để biết đặc điểm của chúng cũng cần một chút nỗ lực tìm hiểu. Chẳng hạn việc tính chỉ số trích dẫn của một ấn phẩm và hệ số ảnh hưởng của một tạp chí trên một cơ sở dữ liệu về cơ bản là chính xác, nhưng các chỉ số liên quan tới một cá nhân cụ thể như số lượng ấn phẩm, chỉ số H, tổng số trích dẫn, ... thường cao hơn giá trị thật do các hệ tìm kiếm chưa phân biệt rõ được tên người. Hiểu rõ đặc điểm các độ đo này ta sẽ dùng chúng đúng hơn, như lý giải kết quả tính toán có thể sai đúng ra sao, kiểm tra các chỉ số khi cần thiết, không so sánh chúng giữa các lĩnh vực khác nhau một cách thô sơ, ...

(2) Nên dùng kết hợp nhiều độ đo đánh giá định lượng: Không nên chỉ đánh giá nghiên cứu dựa trên riêng một độ đo nào, vì như đã chỉ ra mỗi độ đo đều có những hạn chế. Nên kết hợp nhiều độ đo, vì mỗi độ đo cho ta nhìn thấy một khía cạnh của kết quả nghiên cứu. Chẳng hạn đối với một người làm nghiên cứu, chỉ số H cho một ý niệm cơ bản về cân bằng số lượng và chất lượng công bố, các hệ số ảnh hưởng của tạp chí cho một ý niệm về uy tín nơi các kết quả được công bố, và chỉ số trích dẫn cho thấy kết quả nghiên cứu của người này có ảnh hưởng như thế nào.

(3) Có thêm thẩm định của chuyên gia khi cần: Vì các độ đo đánh giá không thể “đo” được hết mọi khía cạnh tinh tế của kết quả nghiên cứu, khi phải đánh giá kỹ một kết quả nghiên cứu rất cần có phân tích của các chuyên gia trong ngành. Chẳng hạn các công cụ hiện nay khi tính chỉ số trích dẫn, và do đó chỉ số H, vẫn chưa phân biệt các trích dẫn của chính tác giả (self citation) hay các trích dẫn để phê phán, hoặc phân biệt đóng góp của tác giả trong bài báo. Chẳng hạn khó có thể nói một người làm nghiên cứu cơ bản lâu năm là xuất sắc nếu giá trị các độ đo này đều thấp, nhưng

cũng chưa hẳn có thể nói ngay một người làm nghiên cứu là xuất sắc khi thấy giá trị các độ đo này cao, đặc biệt khi các công bố thường cùng nhiều tác giả. Khi này cần đến ý kiến chuyên gia trong ngành. Lưu ý là khi tham khảo các độ đo định lượng, các chuyên gia cũng dễ có ý kiến xác đáng hơn.

(4) *Cần chú ý đến ảnh hưởng của khác biệt văn hóa ngành lên các độ đo và số lượng ấn phẩm*: Sẽ rất dễ sai nếu so sánh một cách thô sơ hiệu quả của người làm nghiên cứu ở những ngành khác nhau khi dựa trên số bài báo hay các độ đo đánh giá định lượng kể trên. Có thể nêu thêm ít nhất ba điểm về khác biệt văn hóa ngành liên quan đến số lượng ấn phẩm và các độ đo đánh giá định lượng. Một là một số tạp chí trong một số ngành đòi hỏi mỗi bài báo phải giới hạn trong vòng ba bốn trang, viết theo một cấu trúc nhất định, để có thể công bố nhanh trong vòng vài tháng sau khi gửi bài. Ở một số ngành khác (lại như toán học chẳng hạn), mỗi bài báo thường dài ít nhất khoảng mười trang (bài báo của Ngô Bảo Châu về bổ đề cơ bản cho các đại số Lie dài 197 trang), và thời gian từ lúc gửi đến lúc được đăng thường là hai hoặc ba năm. Hai là các bài báo nghiên cứu lý thuyết như trong toán học thường có trung bình (và phần lớn) hai tác giả, nhưng trong nhiều ngành khoa học thực nghiệm các bài báo thường có nhiều tác giả. Giả sử số tác giả trung bình của các bài báo trong ngành A là sáu và ngành B là hai, và giả sử cần một năm để làm được một bài báo, thì có thể nói nôm na rằng việc người làm nghiên cứu trong ngành A về trung bình có số ấn phẩm gấp ba lần số ấn phẩm của một người làm nghiên cứu trong ngành B cũng là chuyện thường tình. Và do vậy số trích dẫn và chỉ số H của người làm nghiên cứu trong ngành A cũng thường cao hơn. Ba là một số ngành như công nghệ thông tin coi việc công bố kết quả nghiên cứu ở các hội nghị khoa học hàng đầu là quan trọng và có giá trị không kém việc công bố ở các tạp chí có ảnh hưởng cao. Việc hội nghị khoa học có vai trò khác nhau giữa các ngành có liên quan đến các độ đo đánh giá kết quả nghiên cứu khoa học, do hầu hết các cơ sở dữ liệu chưa có thống kê hoặc chưa phân biệt được rõ về thứ hạng của các hội nghị quốc tế.

(5) *Những “sân chơi” khác nhau và những “sân chơi” mới*: Trong [12] các tác giả chỉ ra khoảng 90% các bài báo công bố trên các tạp chí khoa học không bao giờ được trích dẫn, và khoảng 50% các bài báo không bao giờ được ai khác đọc ngoài chính tác giả và những người phản biện. Như ở trên đã nói, các cơ sở dữ liệu phổ biến chứa thông tin của khoảng 15 nghìn trong số trên 100 nghìn các loại tạp chí, và các chỉ số trích dẫn, hệ số ảnh hưởng, chỉ số H ta có cũng chỉ được tính từ các nguồn này. Đẳng cấp của các tạp chí được đánh giá bởi hệ số ảnh hưởng, và “sân chơi” của các tạp chí hàng đầu khá khép kín, nên chừng hơn 80 nghìn tạp chí luôn không được xếp hạng. Các hội nghị khoa học cũng rất thượng vàng hạ cám dù luôn có trong tên cụm từ “hội nghị quốc tế”. Dù xếp thứ hạng các hội nghị có phần khó hơn xếp hạng các tạp chí, đã có những cố gắng trong nhiều ngành để chia các hội nghị thành các nhóm có uy tín khác nhau, chẳng hạn cộng đồng nghiên cứu và giáo dục ngành khoa học máy tính Australia đã đưa ra bảng xếp hạng các hội nghị của ngành (<http://core.edu.au/index.php/categories/conference%20rankings>). Một hiện tượng có thể quan sát được là một số cộng đồng khoa học, vốn không dễ dàng thâm nhập được vào các “sân chơi” của các tạp chí hàng đầu, đã tạo ra những “sân chơi” mới của mình, chẳng hạn như WASET (World Academy of Science, Engineering and Technology, <http://www.waset.org>), WSEAS (World Scientific and Engineering Academy and Society, <http://www.worldses.org/wseas.htm>). Tuy nhiên, chất lượng hội nghị và tạp chí của các cộng đồng này nói chung chưa cao, và cách mời chào ráo riết của họ trên mạng cũng góp phần làm tăng sự băn khoăn về chính họ.

(6) *Không ngừng cải tiến các độ đo*: Bài này lưu ý nhiều về những đặc điểm của các độ đo định lượng, và cũng muốn thêm rằng đang có nhiều nghiên cứu để tạo ra những độ đo tốt hơn, như chỉ số G bổ sung cho chỉ số H kể ở trên. Có những đề nghị cùng với chỉ số trích dẫn nên dùng thêm số lần tải bài báo (download count) và từ đó cải thiện các độ đo liên quan. Một thí dụ khác là hệ số riêng. Xuất phát từ những hạn chế của chỉ số trích dẫn, như không phân biệt trích dẫn từ các nơi có uy tín khác nhau, Bergstrom [4] đã đề xuất khái niệm hệ số riêng (eigenfactor), dựa trên một ý tưởng thú vị. Khi xem các trích dẫn đến một bài báo hay bài báo này trích dẫn nơi khác cũng giống như các đường dẫn đến hay đường dẫn đi từ một trang Web, tác giả đã áp dụng thuật toán nổi tiếng PageRank, nền tảng tìm kiếm của Google, để tính hệ số riêng của mỗi ấn phẩm như cách tính tầm quan trọng cho mỗi trang Web của Google. Chỉ số riêng hiện được bổ sung trong Web of Science, và được nhiều người xem là hợp lý hơn hệ số ảnh hưởng vì có tính đến tầm quan trọng của các trích dẫn. Hy vọng chúng ta sẽ thấy những độ đo tốt hơn trong một tương lai gần.

(7) *Thông tin khoa học ngày càng phong phú và minh bạch*: Người làm nghiên cứu khoa học ngày nay thật may mắn vì có thể có được rất nhiều, thậm chí là hầu hết, tài liệu liên quan đến việc mình muốn làm. Tuy vậy, sự đầy đủ và bình đẳng thông tin không chỉ đem đến cơ hội mà cả những thách thức. Dễ thấy nhất là tốc độ nghiên cứu tăng lên rất nhanh, số công bố cần theo dõi, cần trích dẫn cũng tăng lên nhanh. Điều rất đáng chú ý là tính minh bạch của thông tin được cải thiện rất nhiều. Dường như mọi sai đúng về thông tin khoa học đều có thể kiểm chứng. Ngày nay các tạp chí hay hội nghị dễ dàng xác định những vi phạm như đạo văn hoặc nộp một bài nhiều nơi cùng một lúc (double submission). Hầu hết mọi công bố khoa học có giá trị của bất kỳ ai trong vòng mấy chục năm qua đều có thể tìm thấy. Khi tính minh bạch của thông tin tăng lên thì những gì chưa trung thực dễ lộ ra. Những gì ít minh bạch xưa kia có thể sẽ không sửa được vì đã lên “bia Web”. Người liên quan đến khoa học nhưng ít có tính tôn trọng sự thật hoặc hay nói quá hoặc có chút tính toán để tự đề cao chắc đều nên chú ý một điều là ngày nay những thông tin khoa học đều có thể dễ dàng kiểm chứng trên Web và qua các độ đo định lượng. Cũng cần nói là những người trẻ tuổi sớm quen và thạo hơn với các công nghệ mới của internet, biết nhiều điều hơn về các thể hệ cha chú mình, tuy có thể có phần chưa cảm nhận đủ về những hoàn cảnh khác nhau theo thời gian.

(8) *Về một số đánh giá tình hình nghiên cứu của ta*: Trong những năm qua đã có một số tác giả ở trong và ngoài nước dùng các độ đo đánh giá định lượng để khảo sát và đánh giá tình hình nghiên cứu của ta, tiêu biểu như các tác giả Phạm Duy Hiển và Nguyễn Văn Tuấn (Tuan’s blog). Các khảo sát này sử dụng dữ liệu từ các nguồn ISI, Google Scholar, Scopus, ... và đưa ra những so sánh kết quả và khả năng nghiên cứu giữa Việt Nam và các nước xung quanh. Một số người nghi ngại khi thấy số ấn phẩm của đại học Chulalongkorn của Thái Lan nhiều lần hơn số ấn phẩm của các cơ sở

nghiên cứu và đại học hàng đầu của ta, và cho rằng các cơ sở dữ liệu ISI không đáng tin cậy.

Như đã trình bày ở trên, các cơ sở dữ liệu phổ biến hiện nay chỉ chứa thông tin các tạp chí ảnh hưởng nhiều trong các ngành và các độ đo còn những hạn chế, nhưng chúng đều được xây dựng trên các tiêu chí rõ ràng, vô tư với mọi người, mọi quốc gia và nếu có các sai số cũng không làm lượng công bố của ta tụt xuống vài lần. Bảng số liệu mới dưới đây, dịch từ [15], so sánh số lượng bài báo có thẩm định quốc tế (TDQT) năm 2008 của bốn đại học hàng đầu của ta (hai Đại học Quốc gia, Đại học Bách Khoa và Sư Phạm Hà Nội), Viện Khoa học và Công nghệ Việt Nam và hai Đại học Chulalongkorn và Mahidol hàng đầu của Thái Lan. Mỗi người làm quản lý hoặc nghiên

cứu khoa học của ta chắc đều có những suy nghĩ về những con số này. Dựa trên cơ sở dữ liệu Scopus, cổng thông tin SCImago Journal & Country Rank (<http://www.scimagojr.com/index.php>) gần đây đã cung cấp và so sánh xếp hạng các tạp chí và hiệu quả nghiên cứu ở nhiều lĩnh vực, của mọi quốc gia dựa theo nhiều chỉ số. Có thể thấy ở đây những thông tin đáng lưu tâm về việc khoa học và công nghệ của ta được họ xếp ở thứ hạng nào so với các nước khác, nói chung hay nói riêng từng ngành.

(9) *Về Quỹ NAFOSTED, SCI và SCIE*: Quỹ Phát triển Khoa học và Công nghệ Quốc gia NAFOSTED đã tạo ra một bước tiến rất tích cực trong đánh giá và đầu tư cho các nghiên cứu khoa học cơ bản của ta, do đã dùng các công bố có trong cơ sở dữ liệu SCI và SCIE làm điều kiện cần để xét tuyển đề tài và tiêu chí giao nộp sản phẩm. Cần chú ý là ISI chọn các tạp chí vào SCI và SCIE không chỉ dựa theo giá trị của hệ số ảnh hưởng trong từng ngành mà còn một số yếu tố khác nữa. Như đã phân tích, bản thân SCI và SCIE cũng có khác biệt với các cơ sở dữ liệu lớn khác. Sau thời gian đầu dựa vào các cơ sở dữ liệu này do sự cần thiết, Quỹ NAFOSTED nên chăng nghiên cứu để bổ sung một vài cơ sở dữ liệu khác, dùng thêm độ đo khác để việc tuyển chọn và tiêu chí giao nộp sản phẩm mềm dẻo, thích hợp hơn nhưng vẫn giữ được chất lượng cao? Cũng cần chú ý là nếu xét theo hệ số ảnh hưởng, các tạp chí đứng cuối ở SCIE có giá trị thấp hơn nhiều so với các tạp chí SCI. Nếu không có những điều chỉnh thích hợp, rất có thể NAFOSTED sẽ thu được rất nhiều kết quả công bố ở phía cuối của SCIE, và sẽ khó đạt mục tiêu mong đợi của mình hơn.

(10) *Liên hệ với tình hình của ta*: Cuối cùng và quan trọng hơn cả, là sau khi tìm hiểu kỹ hơn về các độ đo đánh giá nghiên cứu cơ bản và nghiên cứu ứng dụng ta cần dùng chúng một cách thích hợp trong hoàn cảnh cụ thể của mình. Nghĩ đến việc này đầu tiên chắc là người làm quản lý khoa học và công nghệ. Một vài ý kiến sơ bộ có thể như sau. Một là nên từng bước định ra những cách thích hợp để dùng các độ đo định lượng và số lượng ấn phẩm có trong các cơ sở dữ liệu được chọn lọc kỹ như ISI hay Scopus để đánh giá hoạt động khoa học. Tuy nhiên, phải hết sức cẩn trọng khi dùng chúng để đánh giá các nhà khoa học và các tổ chức chuyên ngành. Hai là khi vẫn chỉ có một số ít kết quả nghiên cứu của ta vào được các tạp chí tốt như của SCI, SCIE, cần có đầu tư tăng chất lượng một số tạp chí hàng đầu trong nước, để tạo ra chỗ công bố kết quả khác nhau cho một số đông người làm nghiên cứu. Ba là về nguồn lực con người, rất nên xem xét dùng các độ đo này một cách thích hợp trong việc đánh giá để xét tuyển các vị trí giáo sư-phó giáo sư, bổ sung hoặc thay cho cách “tính điểm” xưa nay, để đề cao chất lượng nghiên cứu khoa học.

4. Kết luận

Bài này giới thiệu khái niệm và một số đặc điểm cơ bản của các độ đo tiêu biểu dùng để đánh giá các ấn phẩm nghiên cứu khoa học, gồm chỉ số trích dẫn, hệ số ảnh hưởng và chỉ số H. Hai đặc điểm của các độ đo này được nhấn mạnh: sự khác nhau của chúng giữa các ngành và một số lý do của việc chúng có thể được tính hoặc giải thích chưa chính xác.

Mặc dù các độ đo đánh giá định lượng nghiên cứu khoa học còn những hạn chế như đã phân tích, cần nhấn mạnh rằng khả năng tính được chúng tự động là một bước tiến nhiều ý nghĩa và việc sử dụng chúng đang là xu thế trong phát triển khoa học. Hiểu rõ và dùng đúng các độ đo định lượng

để đánh giá kết quả nghiên cứu chắc chắn có tác dụng thúc đẩy phát triển khoa học và công nghệ của chúng ta.

Hồ Tú Bảo

Trường Khoa học Tri thức

Viện Khoa học và Công nghệ Tiên tiến Nhật Bản

Tài liệu tham khảo chính

1. Adler, R., Ewing, J., Taylor, P., “Citation Analysis”, *Statistical Science*, 24(1), 1-14, 2009.
2. Althouse, B.M., West, J.D., Bergstrom, T.C., and Bergstrom, C.T., “Differences in Impact Factor Across Fields and Over Time”, *Journal of the American Society for Information Science and Technology*, 60(1), 27-34, 2009.
3. Bakkalbasi, N., Bauer, K., Glover, J., Wang, L., “Three Options for Citation Tracking: Google Scholar, Scopus and Web of Science”, *Biomedical Digital Libraries*, 3(7), 1-8, 2006.
4. Bergstrom, C.T., “Eigenfactor: Measuring the Value and Prestige of Scholarly Journals”, *College & Research Library News*, 68(5), 2007.
5. Hồ Tú Bảo, “Tổ chức và quản lý đề tài nghiên cứu khoa học ở Nhật”, *Tạp chí Tia Sáng*, 8.2008.
6. Campbell, P., “Escape from the Impact Factor”, *Ethics in Science & Environmental Politics*, 8, 5-7, 2008.
7. Eugene Garfield, “Citation Indexes for Science: A New Dimension in Documentation through Association of Ideas”, *Science*, 122(3159), 1955.
8. Harzing, A. K., Wal, R., “Google Scholar as a New Source for Citation Analysis”, *Ethics in Science & Environmental Politics*, Vol. 8, 61-73, 2008.
9. Hirsch, J.E., “An Index to Quantify an Individual’s Scientific Research Output”, *Proceedings of the National Academy of Sciences*, 102 (46), 16569–16572, 2005.
10. Iglesias, J.E., Pecharromen C., “Scaling the h-index for Different Scientific ISI Fields”, *Scientometrics*, 73(3), 303-320, 2007.
11. Kulkarni, A.V., Aziz, B., Shams, I., Busse, J.W., “Comparisons of Citations in Web of Science, Scopus, and Google Scholar for Articles Published in General Medical Journals”, *The Journal of the American Medical Association*, 302(10), 1092-1096, 2009.
12. Kumar, M.J., “Evaluating Scientists: Citations, Impact Factor, h-Index, Online Hits and What Else?”, *IETI Technical Review*, 26(3), 165-168, 2009.
13. Lawani S.M. “Citation Analysis and the Quality of Scientific Productivity”, *BioScience*, 27, 26-31, 1977.
14. Meho, L. I., “The Rise and Rise of Citation Analysis”, *Physics World*, 20(1), 32-36, 2007.

15. Pham Duy Hien, “A Comparative Study of Research Capacities of East Asian Countries and Implication for Vietnam”, *Higher Education*, Springer, Feb. 2010.
16. Tschardtke, T., Hochberg, M.E., Rand, T.A., Resh, V.H., Krauss, J., “Author Sequence and Credit for Contributions in Multiauthored Publications”, *PLoS Biology*, 5(1), 0013-0014, 2007.