



Original Article

Research Productivity and Quality of Higher Education Institutions in Vietnam: An Analysis Based on the Integrated Database of Web of Science and Scopus

Nguyen Huu Thanh Chung^{1,*}, Vo Dinh Hieu², Ngo Manh Dung³

¹*Faculty of Management Science, VNU University of Social Sciences and Humanities,
336 Nguyen Trai, Thanh Xuan, Hanoi, Vietnam*

²*Faculty of Information Technology, VNU University of Engineering and Technology,
144 Xuan Thuy, Cau Giay, Hanoi, Vietnam,*

³*VNU Centre for Applied Information Technology, 144 Xuan Thuy, Cau Giay, Hanoi, Vietnam*

Received 19 November 2019

Revised 02 December 2019; Accepted 04 December 2019

Abstract: This study integrates the WoS and Scopus articles by Vietnamese authors into the Vietnam Citation Gateway system of Vietnam National University, Hanoi (VNU), sets up a merging tool to filter the duplication for these two databases and establishes their bibliometric and citation index. Based on this integrated database, the reality of international publications of Vietnam and Vietnamese universities in the period of 2014-2018 was analyzed, benchmarked and discussed. Accordingly, in 2018, the annual productivity of international publications in Vietnam reached 10,000 articles, nearly 70% of those were contributed by higher education institutions. This shows a strong yearly increase in Vietnam's WoS & Scopus articles (of 34.7% for Vietnam in general and 41.6% for higher education institutions in particular). In terms of research quality, Vietnam has an average citation index of 9.2, which is comparable to that of Asian higher education institutions. Notably, the study clearly shows the correlation between the university autonomy and the improvement of research productivity and quality.

Keywords: Data merging techniques, bibliometrics, research productivity, research quality, citation index.

* Corresponding author.

E-mail address: chungnht@vnu.edu.vn

<https://doi.org/10.25073/2588-1116/vnupam.4201>



Phân tích năng suất và chất lượng nghiên cứu khoa học của các cơ sở giáo dục đại học Việt Nam dựa trên cơ sở dữ liệu tích hợp Web of Science và Scopus

Nguyễn Hữu Thành Chung^{1,*}, Võ Đình Hiếu², Ngô Mạnh Dũng³

¹*Khoa Khoa học Quản lý, Trường Đại học Khoa học Xã hội và Nhân văn, ĐHQGHN, 336 Nguyễn Trãi, Thanh Xuân, Hà Nội, Việt Nam*

²*Khoa Công nghệ Thông tin, Trường Đại học Công nghệ, ĐHQGHN, 144 Xuân Thủy, Cầu Giấy, Hà Nội, Việt Nam*

³*Trung tâm Công nghệ Thông tin, ĐHQGHN, 144 Xuân Thủy, Cầu Giấy, Hà Nội, Việt Nam*

Nhận ngày 19 tháng 11 năm 2019

Chỉnh sửa ngày 02 tháng 12 năm 2019; Chấp nhận đăng ngày 04 tháng 12 năm 2019

Tóm tắt: Công trình này nghiên cứu phát triển cơ sở dữ liệu (CSDL) Web of Science và Scopus trên hệ thống Vcgate của ĐHQGHN đối với các bài báo của các tác giả Việt Nam; xây dựng công cụ tích hợp CSDL của 2 hệ thống đó để lọc sự trùng lặp; xây dựng chỉ mục và chỉ số trích dẫn. Dựa trên CSDL tích hợp thu được, thực trạng về công bố quốc tế của Việt Nam và các cơ sở giáo dục đại học (CSGDĐH) Việt Nam trong giai đoạn 2014-2018 được phân tích, đối sánh và thảo luận. Lần đầu tiên số liệu đầy đủ về tình hình công bố quốc tế WoS & Scopus được xác lập. Theo đó, năm 2018, số lượng công bố quốc tế hàng năm của Việt Nam đã cán mốc gần 10.000 bài, trong đó các CSGDĐH đóng góp tới 70%. Tốc độ gia tăng các bài báo WoS & Scopus hàng năm của Việt Nam tăng mạnh (34,7% đối với cả nước và 41,6% đối với riêng các CSGDĐH). Bên cạnh việc phân tích đóng góp của hệ thống GDĐH, sự gia tăng của các trường đại học trẻ như Trường ĐH Tôn Đức Thắng và Trường ĐH Duy Tân cũng đã được đề cập. Về chất lượng nghiên cứu, Việt Nam có chỉ số trích dẫn bằng 9,2 - ngang mức trung bình của các CSGDĐH châu Á. Phân tích này đã định vị được và cho phép đề xuất giải pháp đầu tư phát triển nhóm các CSGDĐH có năng lực nghiên cứu khoa học và các lĩnh vực khoa học có tầm ảnh hưởng lớn. Đặc biệt, công trình này cũng chỉ ra mối liên hệ giữa tự chủ đại học và việc nâng cao năng suất và chất lượng nghiên cứu khoa học.

Từ khóa: Công nghệ tích hợp dữ liệu khoa học, phân tích thư mục, năng suất nghiên cứu, chất lượng nghiên cứu, chỉ số trích dẫn.

*Tác giả liên hệ.

Địa chỉ email: chungnht@vnu.edu.vn

<https://doi.org/10.25073/2588-1116/vnupam.4201>

1. Mở đầu

Việc xây dựng thành công các hệ thống cơ sở dữ liệu (CSDL) khoa học (bao gồm thông tin siêu dữ liệu (metadata) hoặc thông tin thư mục (bibliometrics) và chỉ số trích dẫn của các bài báo xuất bản trên các tạp chí, báo cáo khoa học xuất bản trên kỷ yếu hội nghị, sách chuyên khảo, thậm chí cả các bằng sáng chế) đã mở ra khả năng phân tích, đánh giá và đối sánh năng suất và chất lượng nghiên cứu khoa học của các quốc gia, cơ sở nghiên cứu, các nhà khoa học góp phần hình thành một khoa học mới về đo lường và đánh giá khoa học (Sciencetometrics) [1-3]. Các kết quả nghiên cứu, phân tích của ngành khoa học mới này đã hỗ trợ không những cho các nhà khoa học tra cứu phục vụ nghiên cứu khoa học mà còn cho các nhà nghiên cứu chính sách, các tổ chức quản lý khoa học và cả Chính phủ của rất nhiều nước sử dụng để hoạch định chính sách phát triển và kế hoạch đầu tư [4-6].

Nhiều quốc gia và doanh nghiệp số đã tập trung đầu tư xây dựng các hệ thống CSDL khoa học. Ở mức độ quốc gia, có thể tìm thấy Chinese Science Citation Database (Trung Quốc) [7], Russian Citation Science Index (Nga) [8], Korea Citation Index (Hàn Quốc) [9], Malaysia Citation Index (Malaysia) [10]... Ở mức độ doanh nghiệp số toàn cầu, các hệ thống CSDL khoa học lớn, có khả năng truy cập, tìm kiếm và phân tích cao thuộc về hệ thống cơ sở dữ liệu Web of Science (WoS) [11], hệ thống cơ sở dữ liệu Scopus [12] và hệ thống cơ sở dữ liệu Google Scholar [13]. Bên cạnh chiến lược phát triển độc lập, các hệ thống CSDL toàn cầu này còn có kế hoạch tích hợp và thu tóm CSDL của các quốc gia. Thực tế, CSDL của Web of Science (trước đây được biết là ISI) đã tích hợp cả CSDL khoa học của Trung Quốc, Nga, Hàn Quốc và chỉ số trích dẫn SciELO [11].

Mặc dù WoS và Scopus là hai hệ thống CSDL khoa học lớn, mỗi hệ thống có tương ứng khoảng 22.200 và 24.600 tạp chí; 104.500 và 194.000 sách chuyên khảo [14, 15], nhưng chưa thể bao phủ toàn bộ hệ thống CSDL toàn cầu. Mỗi hệ thống chỉ chứa được một phần CSDL của thế giới, trong đó có bao gồm một phần của

CSDL hệ thống kia (tùy thuộc vào trạng thái đăng ký của từng tạp chí trong hai hệ thống). Mức độ chồng phủ của các tạp chí trong WoS và Scopus được đánh giá vào khoảng 70%. Điều này dẫn đến một bất cập là thông tin nhận được về năng suất và chất lượng nghiên cứu khoa học từ mỗi hệ thống CSDL khoa học độc lập này chưa phản ánh hết thực tế công bố quốc tế của các quốc gia, cơ sở nghiên cứu và nhà khoa học. Nếu chỉ thống kê từ CSDL WoS thì một số thành tích nghiên cứu công bố trên Scopus không được tính. Theo thói quen này, lĩnh vực Khoa học Xã hội và Nhân văn (nhất là đối với các nước đang phát triển) thường hay bị thiệt thòi vì họ thường công bố nhiều trong các tạp chí thuộc hệ thống Scopus. Ngược lại, các bảng xếp hạng đại học QS, THE thường sử dụng CSDL của Scopus sẽ tính thiếu các bài báo trong CSDL WoS. Ví dụ, khi tìm kiếm kết quả về năng suất công bố quốc tế của ĐHQGHN (ĐHQGHN) năm 2018 theo WoS, chỉ tìm thấy 500 bài, nhưng theo Scopus, lại có 669 bài, trong đó, có nhiều bài báo xuất hiện tại cả 2 CSDL trên. Điều này dẫn đến sai lệch kết quả thống kê về năng suất công bố quốc tế của ĐHQGHN. Về mặt tích hợp, Google Scholar có tính toàn cầu cao hơn. Đặc biệt, hệ thống này rất mạnh về việc thu thập chỉ số trích dẫn cho các tài liệu khoa học từ nhiều nguồn CSDL khác nhau. Tuy nhiên Google Scholar không có chức năng thống kê thư mục nên khả năng phân tích, đánh giá có hạn chế. Do vậy, việc xây dựng CSDL khoa học tích hợp là rất cần thiết.

Ở Việt Nam, nhóm nghiên cứu của Nguyễn Hữu Đức và Võ Đình Hiếu [16] đã triển khai xây dựng thành công Phần mềm hệ thống công thông tin chỉ số nghiên cứu Việt Nam (Vietnam Citation Gateway – Vcgate.vnu.edu.vn). Hiện nay, hệ thống này đã chỉ mục cho gần 70 tạp chí xuất bản trực tuyến của Việt Nam với gần 50.000 bài báo, bước đầu phục vụ việc xếp hạng chất lượng các tạp chí và cung cấp khả năng phân tích, đánh giá năng lực công bố kết quả nghiên cứu ở trong nước. Ngoài việc thu thập và chỉ mục cho các bài báo trên các tạp chí của Việt Nam, Vcgate còn có khả năng thu thập chỉ số trích dẫn. Do đó, Vcgate có chức năng tương tự như các hệ

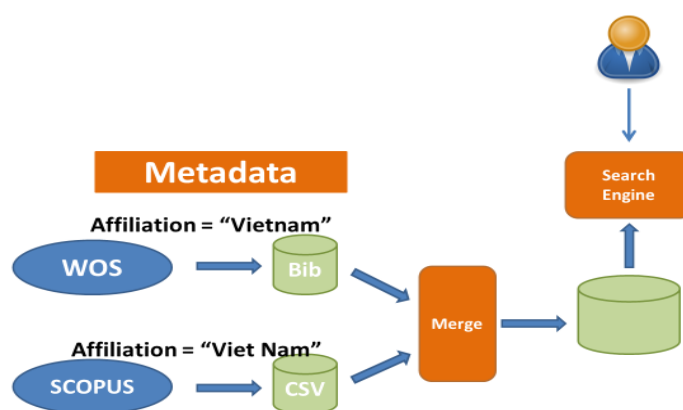
thống CSDL khoa học của các quốc gia khác, đặc biệt rất tương đồng với WoS và Scopus.

Để tiếp cận CSDL khoa học đầy đủ hơn, góp phần đánh giá, phân tích đúng năng lực nghiên cứu khoa học của quốc gia, các tổ chức khoa học và công nghệ, các nhà khoa học ở Việt Nam, công trình này nghiên cứu phát triển CSDL WoS và Scopus vào hệ thống Vcgate của ĐHQGHN đối với các bài báo của các tác giả Việt Nam; xây dựng công cụ tích hợp CSDL của 2 hệ thống này để lọc sự trùng lặp; xác định chỉ mục và chỉ số trích dẫn cho các bài báo nói trên. Dựa trên CSDL tích hợp thu được, thực trạng về công bố quốc tế của các trường đại học Việt Nam trong giai đoạn 2014-2018 được đánh giá, phân tích, thảo luận và đối sánh.

2. Hệ thống tích hợp cơ sở dữ liệu WoS và Scopus

Hệ thống CSDL tích hợp WoS và Scopus được nghiên cứu và xây dựng tại ĐHQGHN qua cổng thông tin điện tử <https://vcgate.vnu.edu.vn:3000>. Hệ thống này sử dụng công nghệ thu thập dữ liệu (Data Acquisition), công nghệ tìm kiếm (Search Engine), công nghệ tích hợp CSDL (Merge), công nghệ quản lý, phân tích dữ liệu và báo cáo thống kê (Database). Trong đó, công nghệ thu

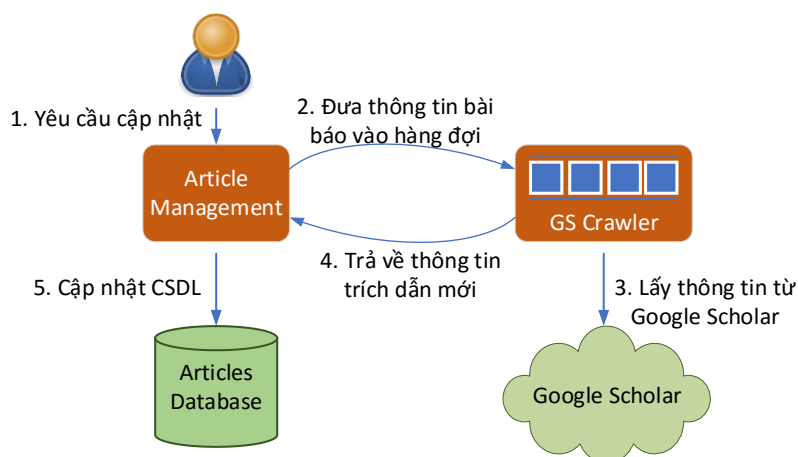
thập dữ liệu sử dụng các thư viện lấy dữ liệu từ Internet được viết trên ngôn ngữ Python như Scrapy (<https://scrapy.org>). Công nghệ tìm kiếm được áp dụng một trong hai máy tìm kiếm Apache Solr (<http://lucene.apache.org/solr/>) hoặc Elasticsearch (<https://www.elastic.co/>). Công nghệ quản lý, phân tích và báo cáo thống kê sử dụng hệ quản trị cơ sở dữ liệu mã nguồn mở MySQL. Công nghệ tích hợp (merge) CSDL WoS và Scopus thực chất là công nghệ phân loại văn bản và kiểm tra sự trùng lặp các bài báo xuất hiện trong cả hai CSDL để lọc và tính chung một lần [17]. WoS và Scopus hỗ trợ người dùng với nhiều định dạng siêu dữ liệu bao gồm CSV, Bibtext, RIS,... Tuy nhiên, BibTex (trong WoS) và CSV (trong Scopus) được chọn vì hai loại định dạng này cung cấp đầy đủ thông tin, thuận tiện cho việc xử lý (hình 1). Công nghệ này được thực hiện đối với các trường siêu dữ liệu (metadata), áp dụng đối với thông tin tạp chí (tên, ISSN, tập, số của tạp chí) và thông tin của bài báo khoa học (tên bài báo, tác giả, địa chỉ và DOI của bài báo). Nếu hai bài báo cùng DOI (sau khi được chuẩn hóa) thì được xác định là cùng một bài. Trong trường hợp một trong hai bài (hoặc cả hai) không có thông tin về DOI, thì tiêu đề bài báo, tên tạp chí và thông tin xuất bản (số tập, năm và trang tạp chí,...) sẽ được đối sánh, tính điểm để xác định có sự trùng lặp hay không.



Hình 1. Công nghệ tích hợp cơ sở dữ liệu từ nguồn WoS và Scopus.

Công nghệ thu thập số lần trích dẫn từ Google Scholar được thực hiện theo quy trình chỉ ra trên hình 2. Quá trình lấy trích dẫn được bắt đầu từ người dùng (là quản trị nội dung hệ thống). Người dùng có thể chỉ định lấy trích dẫn của một hay nhiều bài báo. Mô đun quản lý bài báo sẽ chuyển thông tin các bài báo cần thu thập

hoặc cập nhật trích dẫn vào hàng đợi của mô đun GS Crawler, là mô đun sẽ trực tiếp kết nối đến Google Scholar và tìm trích dẫn của các bài báo. Hàng đợi được sử dụng vì tại một thời điểm có thể có nhiều yêu cầu cập nhật trích dẫn. Thông tin trích dẫn mới sẽ được chuyển về mô đun quản lý bài báo và cập nhật vào CSDL để sử dụng về sau.

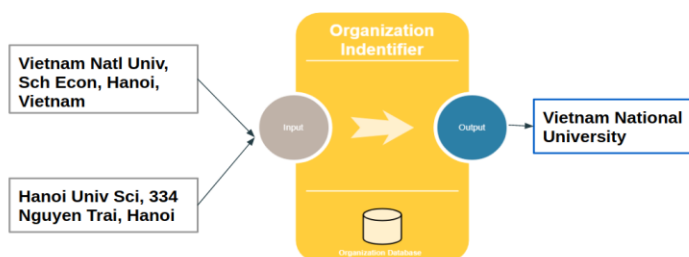


Hình 2. Sơ đồ thu thập dữ liệu về số lần trích dẫn từ nguồn Google Scholar.

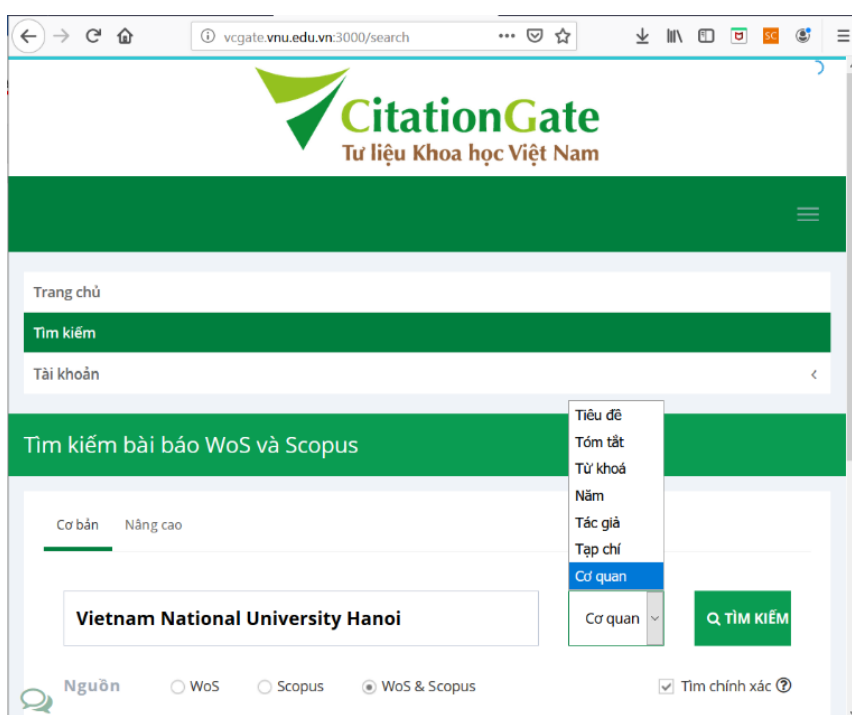
Sau khi tích hợp xong CSDL và thu thập chỉ số trích dẫn có thể áp dụng công nghệ tìm kiếm và công nghệ quản lý, phân tích, thống kê để chiết xuất các báo cáo đối sánh. Tuy nhiên, nhiều bài báo trong CSDL WoS và Scopus bị xác định nhầm cơ quan của tác giả, nhất là đối với các Viện hàn lâm, Đại học Quốc gia và Đại học vùng - ở đó việc ghi địa chỉ của các đơn vị thành viên và trực thuộc rất khác nhau. Hơn thế nữa, việc sử dụng tên cơ quan bằng tiếng Anh cũng được sử dụng rất tùy hứng. Điều này dẫn đến việc thống kê số bài xuất bản của từng cơ quan trong nước không chính xác. Đây là vấn đề quá chi tiết và đặc thù của Việt Nam, không một hệ thống CSDL nào có thể nhận biết và xử lý cụ thể được. Do đó, việc xây dựng mô hình nhận diện cơ quan của bài báo nhằm đưa ra các số liệu thống kê chính xác hơn cho Việt Nam cũng rất cần thiết (Hình 3). Có thể nêu ra ví dụ, đối với ĐHQGHN, dữ liệu đầu vào là các chuỗi cơ quan viết bằng tiếng Anh có cấu trúc với gần 20 địa chỉ và tên tiếng Anh khác nhau. Khi đó, dữ liệu đầu ra phải

được chuẩn hóa và mỗi cơ quan chỉ được biểu diễn bằng một chuỗi duy nhất để phục vụ việc thống kê các bài báo theo cơ quan. Trong trường hợp này, mô hình học máy kết hợp TF-IDF (Term Frequency – Inverse Document Frequency) và SVM (Support Vector Machine) đã được xây dựng.

Trang điện tử của hệ thống CSDL tích hợp WoS & Scopus Vcgate <https://vcgate.vnu.edu.vn:3000> được giới thiệu trên hình 4. Nhận thấy rằng hệ thống này có thể tìm kiếm, phân tích các bài báo thuộc hệ thống WoS hoặc Scopus một cách độc lập, hoặc tích hợp tùy theo mục đích của người dùng. Các tìm kiếm có thể thực hiện theo các phân loại: tên bài báo, tên tác giả, tóm tắt, từ khóa, năm xuất bản, địa chỉ cơ quan để chiết xuất các báo cáo phân tích, đánh giá phục vụ các đối tượng khác nhau. Các đánh giá, phân tích năng suất và chất lượng nghiên cứu khoa học của các cơ sở giáo dục đại học Việt Nam trình bày dưới đây được thống kê trên hệ thống Vcgate này.



Hình 3. Sơ đồ nguyên lý về công cụ nhận diện địa chỉ cơ quan của các bài báo.



Hình 4. Trang điện tử của hệ thống CSDL tích hợp WoS và Scopus Vcgate <https://vcgate.vnu.edu.vn:3000>.

3. Phân tích cơ sở dữ liệu WoS và Scopus giai đoạn 2014-2018 của Việt Nam

Số liệu công bố quốc tế trong CSDL WoS, Scopus và WoS & Scopus của Việt Nam nói chung và các cơ sở giáo dục đại học (CSGDĐH) nói riêng trong giai đoạn 2014-2018 được thống kê và trình bày trên bảng 1 và biểu diễn trên hình 5. Trong 5 năm đó, cả nước đã công bố 22.438 bài báo WoS, 29.932 bài báo Scopus và tổng cộng 32.732 bài trong CSDL tích hợp WoS & Scopus. Đồng thời, số liệu tương ứng của các

CSGDĐH là 13.728 (WoS), 21.702 (Scopus) và 23.144 bài (WoS & Scopus), chiếm trung bình khoảng 70% so với năng suất của cả nước. Nhận xét đầu tiên qua quan sát số liệu này là: tổng số bài báo Scopus cao hơn số bài báo WoS 1,33 lần; tổng số bài báo WoS & Scopus cao hơn số bài báo WoS 1,46 lần và số bài báo Scopus 1,09 lần. Lưu ý rằng, tổng số bài báo WoS & Scopus trong giai đoạn 2014-2018 của Việt Nam chỉ tương đương với năng suất công bố bài báo Scopus của Indonesia trong năm 2017 (21.300 bài) hoặc năm 2018 (33.988 bài).

Bảng 1. Số bài báo của Việt Nam và CSGDDH Việt Nam giai đoạn 2014-2018.

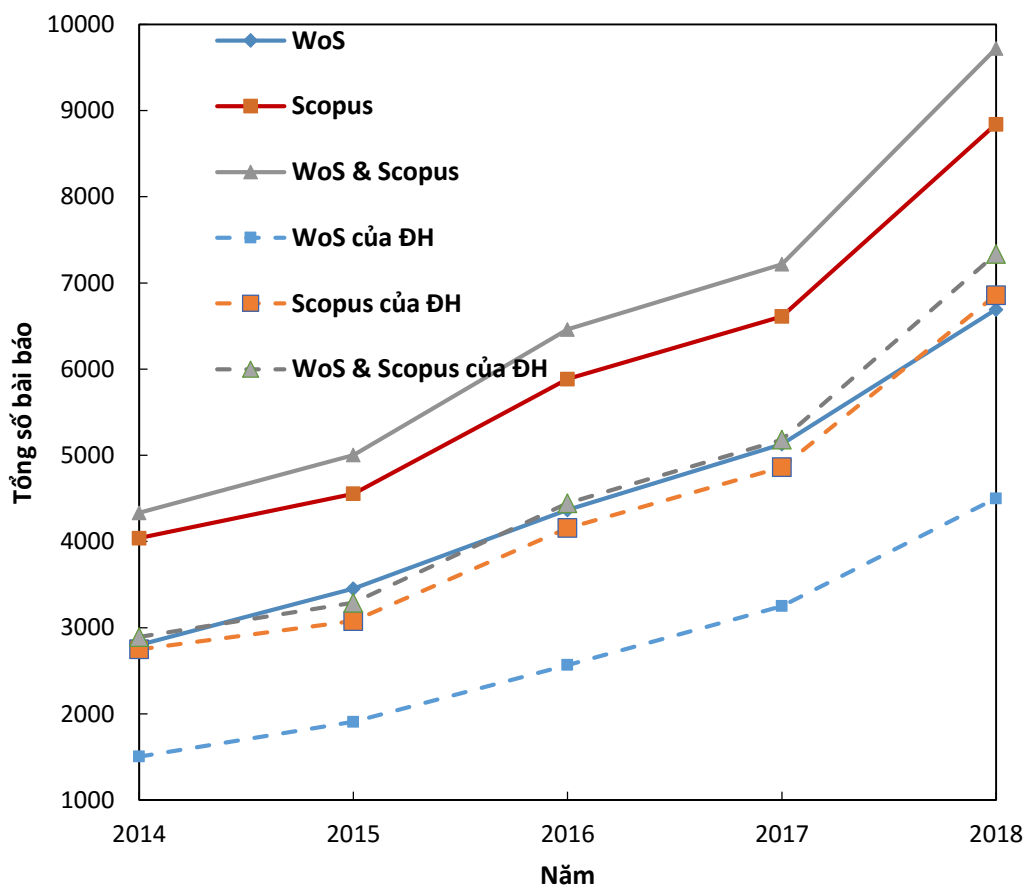
Năm	WoS			Scopus			WoS & Scopus		
	VN	ĐH	Tỉ lệ	VN	ĐH	Tỉ lệ	VN	ĐH	Tỉ lệ
2014	2799	1504	53,7%	4040	2749	68,0%	4332	2892	66,8%
2015	3453	1908	55,3%	4553	3077	67,5%	5003	3289	65,7%
2016	4366	2566	58,8%	5885	4156	70,6%	6461	4445	68,8%
2017	5129	3250	63,4%	6612	4862	73,5%	7217	5182	71,8%
2018	6691	4500	67,3%	8842	6858	77,6%	9719	7336	75,5%
Tổng	22.438	13.728	61,3%	29.932	21.702	72,5%	32.732	23.144	70,7%

Năng suất khoa học - số lượng bài báo công bố trên Việt Nam đang tăng qua từng năm và có dấu hiệu tăng mạnh hơn sau năm 2017. Trong giai đoạn 2014-2017, trung bình mỗi năm số lượng bài báo WoS & Scopus tăng khoảng 18,8% (từ 4.332 bài đến 7.217 bài trong 3 năm). Tuy nhiên, chỉ trong một năm từ 2017-2018, số lượng bài báo đã tăng lên 34,7% (từ 7.217 đến 9.719 bài). Như vậy, năm 2018, Việt Nam đã gần như đã cán mốc 10.000 công bố quốc tế một năm. Kết quả này cao hơn một chút so với năng suất công bố bài báo Scopus hàng năm (ví dụ: 8.800 bài năm 2018) của Trường ĐHQG Singapore. Trong số đó, tốc độ gia tăng số công bố trong hai giai đoạn phân kỳ ấy của WoS là 22,4% và 30,5%; của Scopus là 18,1% và 33,7%. Tương tự, số lượng bài báo WoS & Scopus của các CSGDDH Việt Nam, từ 2014-2017, trung bình mỗi năm tăng 21,8%, nhưng chỉ tính riêng năm từ 2017-2018, con số này tăng lên đến 41,6%. Trong đó, tốc độ gia tăng số công bố trong hai giai đoạn phân kỳ ấy của WoS là 29,4% và 38,5%; của Scopus là 21,3% và 41,1%. Kết

quả gia tăng công bố quốc tế vừa mô tả được tổng kết lại trên bảng 2. Nhận thấy rằng, tỉ lệ gia tăng các bài báo Scopus ngày càng chiếm trọng số quan trọng và đặc biệt là sự đóng góp vào sự gia tăng đến từ các CSGDDH. Kết quả này cho thấy các trường đại học tại Việt Nam đang dần quan tâm đến công bố quốc tế. Như có thể nhìn thấy từ độ dốc của các đồ thị trên hình 5, trước năm 2016 tốc độ gia tăng về số lượng bài báo trên 2 hệ thống CSDL WoS và Scopus là như nhau. Tuy nhiên, từ 2016 trở đi, tốc độ gia tăng số lượng bài báo đăng trên hệ thống CSDL Scopus bắt đầu vượt lên so với WoS. Điều này còn có thể liên quan đến xu thế gia tăng các bài báo trong lĩnh vực Khoa học Xã hội và Nhân văn. Đối với lĩnh vực này, công bố bài trên hệ thống CSDL Scopus đang là một lựa chọn phù hợp. Chính sự gia tăng năng suất công bố quốc tế của các CSGDDH đã quyết định mức độ gia tăng chung về công bố quốc tế của cả nước. Sự gia tăng có tính chất đột biến này xảy ra trong thời kỳ các CSGDDH Việt Nam được trao quyền tự chủ.

Bảng 2. Tốc độ gia tăng về năng suất công bố quốc tế của Việt Nam và các CSGDDH giai đoạn 2014 - 2018

Thời gian	Tốc độ gia tăng					
	WoS		Scopus		WoS & Scopus	
	VN	ĐH	VN	ĐH	VN	ĐH
2014 – 2017	22,4%	29,4%	18,1%	21,3%	18,8%	21,8%
2017 – 2018	30,5%	38,5%	33,7%	41,1%	34,7%	41,6%



Hình 5. Số bài báo của Việt Nam và các CSGDĐH (ký hiệu chung là ĐH) tại Việt Nam giai đoạn 2014-2018.

Số liệu về tổng số bài báo WoS & Scopus và số lần trích dẫn của chúng trong giai đoạn 2014-2018 được liệt kê trên bảng 3 và biểu diễn trên hình 6 cho top 30 CSGDĐH có số lượng bài báo nhiều nhất trong giai đoạn 2014-2018. Có thể thấy rằng, đến hết năm 2018, ĐHQGTpHCM và ĐHQGHN vẫn là hai cơ sở giáo dục có tổng số bài báo cao nhất. Điều thú vị là Trường Đại học Tôn Đức Thắng đã vươn lên vị trí top 3 và Trường đại học Duy Tân nằm ở vị trí top 5. Khi so sánh năng suất công bố quốc tế của các CSGDĐH có thể thấy số lượng bài báo của trường top 1 (ĐHQGTpHCM) chỉ nhiều hơn xấp xỉ 2 lần so với trường top 4 (Trường ĐHBK Hà Nội). Thực tế, đây là 4 CSGDĐH hàng đầu của Việt Nam đã có tên trong các bảng xếp hạng đại

học quốc tế như QS, THE và AWRU. Tuy nhiên, nếu so sánh với các trường top 15 thì sự khác nhau về số lượng công bố quốc tế đã là 10 lần. Đáng nói hơn, khi so sánh trường top 1 với top 30, sự khác biệt lên đến gần 50 lần. Điều này cho thấy ở Việt Nam mới chỉ có khoảng 15 CSGDĐH có định hướng và năng lực nghiên cứu nổi trội. Đây là thông tin có ý nghĩa, có thể sử dụng để hoạch định chính sách phát triển các CSGDĐH.

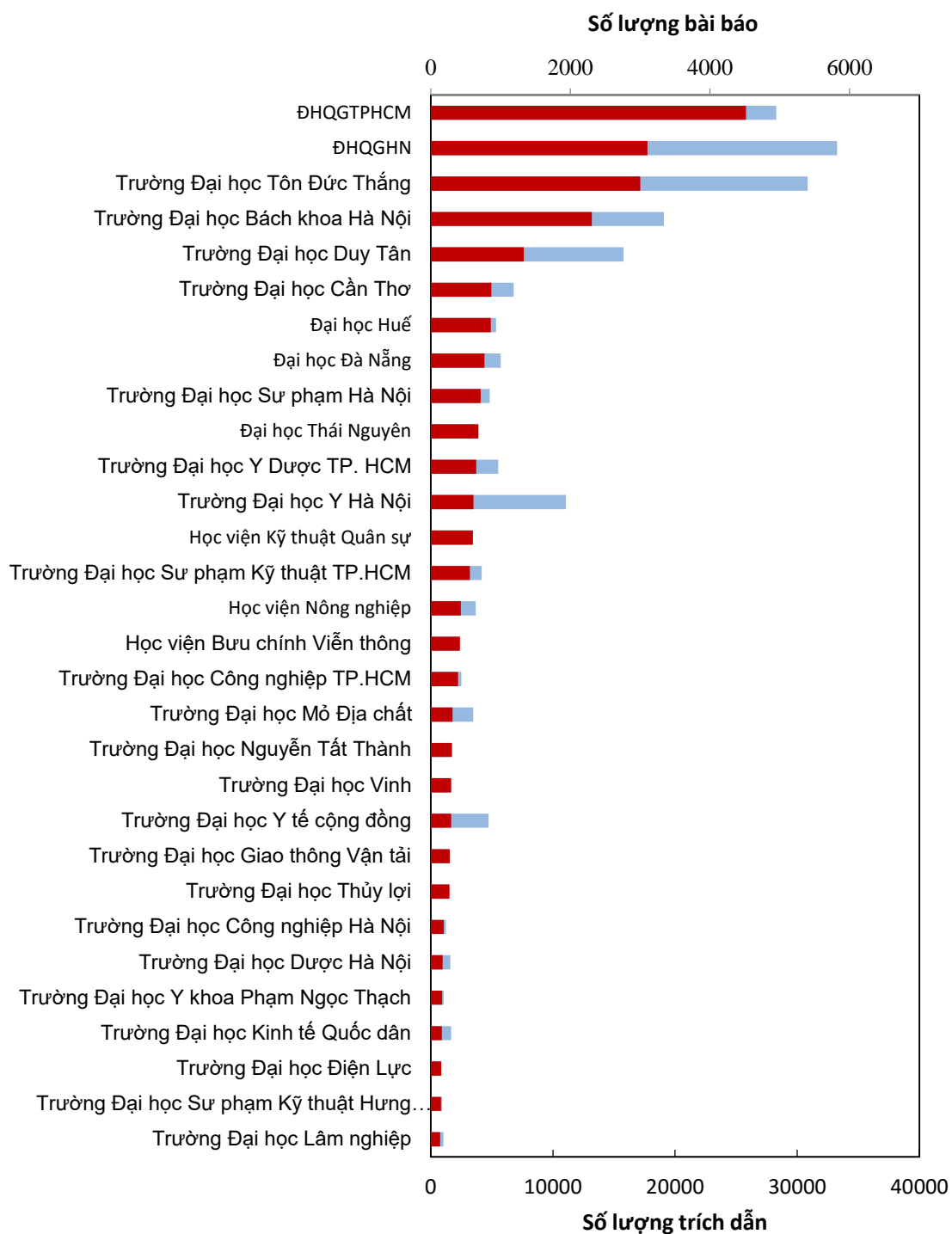
Khi khảo sát cụ thể mức độ gia tăng trong thời gian gần đây của một số CSGDĐH (hình 7) còn có thể thấy rằng, trong giai đoạn 2014-2018, tốc độ gia tăng về công bố quốc tế hàng năm của ĐHQGHN, ĐHQGTpHCM và các CSGDĐH truyền thống hầu như không có sự đột biến, trong

lúc đó Trường Đại học Tôn Đức Thắng và Trường ĐH Duy Tân có tốc độ gia tăng vượt trội. Năm 2017, số lượng bài báo WoS & Scopus của Trường ĐH Tôn Đức Thắng đã vượt qua ĐHQGHN và năm 2018 lại tiếp tục vượt qua ĐHQGTpHCM. Năm 2018, số lượng bài báo WoS & Scopus của Trường ĐH Duy Tân cũng đã tiến kịp Trường ĐHBKHN. Trước đó, năm 2014, hai trường này thậm chí còn không nằm trong top 10 của Việt nam. Đóng góp của Trường ĐH Tôn Đức Thắng đã làm thay đổi cán cân công bố quốc tế của các CSGDDH tại khu vực TP.HCM so với khu vực Hà Nội (hình 8). Cụ thể,

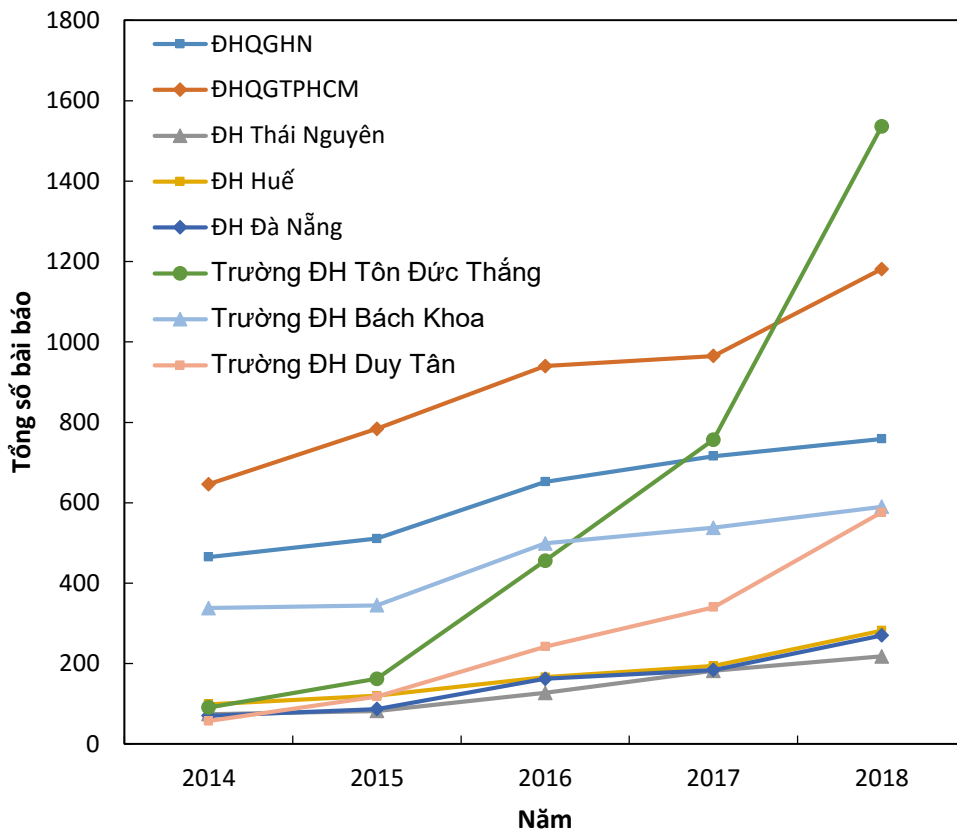
vào năm 2014, số lượng bài báo WoS & Scopus của các CSGDDH Hà Nội cao hơn, sự cân bằng giữa hai khu vực đã đạt được vào năm 2017. Vào năm 2018, số lượng bài báo của các CSGDDH tại TP.HCM đã tăng mạnh, vượt xa so với số lượng bài báo của các CSGDDH tại Hà Nội. Xu thế này đang tiếp tục được khẳng định trong năm 2019. Tính đến 11/2019, Trường ĐH Tôn Đức Thắng và Trường ĐH Duy Tân lần lượt đã công bố được 2300 và 980 bài báo WoS & Scopus, dẫn đầu các CSGDDH. Đồng thời, Trường ĐH Tôn Đức Thắng đã đứng đầu cả nước, vượt qua cả Viện Hàn lâm KH&CN Việt Nam.

Bảng 3. Tổng số bài báo WoS & Scopus và chỉ số trích dẫn của top 30 CSGDDH Việt Nam có nhiều công bố quốc tế nhất giai đoạn 2014-2018.

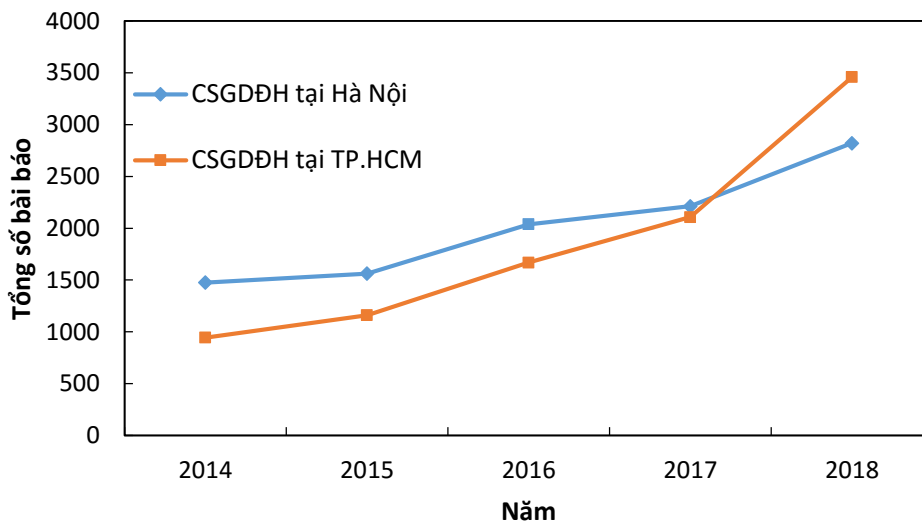
STT	Tên trường	Số bài báo WoS & Scopus	Tổng số trích dẫn	Số trích dẫn trung bình
1	ĐHQGTp.HCM	4516	28277	6,26
2	ĐHQGHN	3103	33269	10,72
3	ĐH Tôn Đức Thắng	3001	30863	10,28
4	Trường ĐH Bách Khoa Hà Nội	2307	19098	8,28
5	Trường đại học Duy Tân	1333	15777	11,84
6	Trường đại học Cần Thơ	867	6785	7,83
7	Đại học Huế	860	5346	6,22
8	Đại học Đà Nẵng	770	5736	7,45
9	Trường ĐH Sư phạm Hà Nội	713	4823	6,76
10	Đại học Thái Nguyên	683	3397	4,97
11	Trường ĐH Y dược TP.HCM	652	5533	8,49
12	Trường ĐH Y Hà Nội	613	11061	18,04
13	Học viện Kỹ thuật Quân sự	603	3334	5,53
14	Trường ĐH Sư phạm Kỹ thuật TP.HCM	562	4164	7,41
15	Học viện Nông nghiệp	431	3680	8,54
16	Học viện Bưu chính Viễn thông	419	2300	5,49
17	Trường ĐH Công nghiệp TP.HCM	392	2496	6,37
18	Trường ĐH Mỏ Địa chất	313	3466	11,07
19	Trường ĐH Nguyễn Tất Thành	305	1626	5,33
20	Trường ĐH Vinh	293	1648	5,62
21	Trường ĐH Y tế cộng đồng	293	4734	16,16
22	Trường ĐH Giao thông Vận tải	274	1464	5,34
23	Trường ĐH Thủy lợi	270	1410	5,22
24	Trường ĐH công nghiệp Hà Nội	186	1249	6,72
25	Trường ĐH Dược Hà Nội	170	1600	9,41
26	Trường ĐH Y khoa Phạm Ngọc Thạch	162	1066	6,58
27	Trường ĐH Kinh tế Quốc dân	158	1677	10,61
28	Trường ĐH Điện Lực	152	524	3,45
29	Trường ĐH Sư phạm Kỹ thuật Hưng Yên	145	928	6,4
30	Trường ĐH Lâm nghiệp	137	1038	7,58



Hình 6. Tổng số bài báo WoS & Scopus và số trích dẫn giai đoạn 2014-2018 của top 30 CSGDĐH Việt Nam.



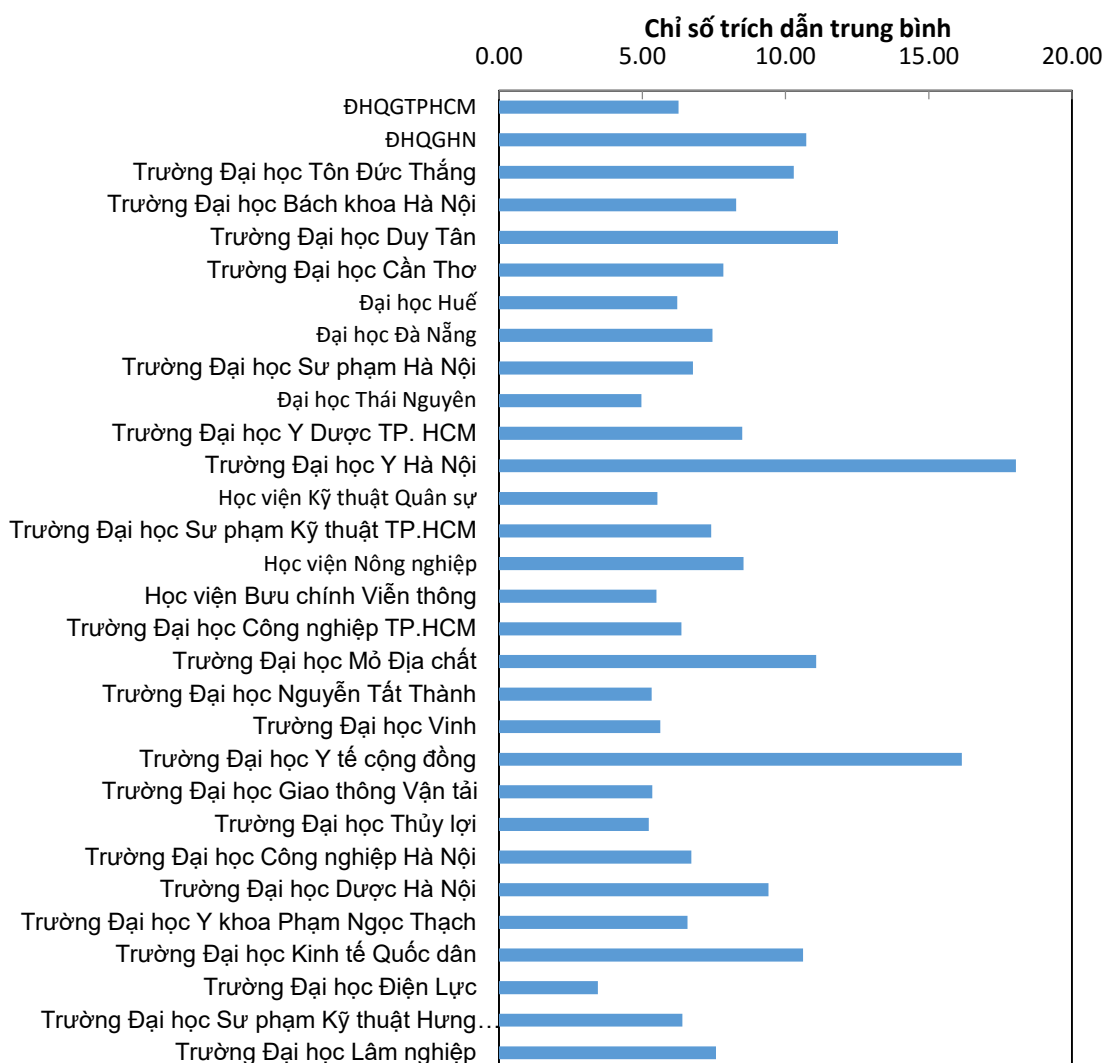
Hình 7. So sánh số lượng bài báo theo một số trường top đầu trong giai đoạn 2014 – 2018.



Hình 8. So sánh mức độ gia tăng về số lượng bài báo giữa các CSGDĐH tại khu vực Hà Nội và TP. HCM.

Chất lượng nghiên cứu thường được đánh giá bằng chỉ số trích dẫn của các bài báo khoa học. Tổng số trích dẫn tính và số trích dẫn trung bình thống kê tại thời điểm 11/2019 cho các bài báo WoS & Scopus của các CSGDĐH xuất bản trong giai đoạn 2014-2019 được liệt kê trong bảng 1 và biểu diễn trên hình 6. Các thống kê trong khoảng thời gian như vậy phù hợp với cách xác định của các bảng xếp hạng đại học QS, THE, đảm bảo được sự ổn định thống kê. Chỉ số trích dẫn trung bình của các bài báo được biểu diễn trên hình 9. Nói chung, chỉ số trích dẫn thu thập được theo phương pháp mà công trình này

áp dụng cao hơn gần 1,5 lần so với chỉ số thu thập được từ nguồn WoS hoặc Scopus. Điều này hoàn toàn phù hợp với các kết quả nghiên cứu của Martín-Martín và đồng nghiệp [18]. Nhóm này cho rằng Google Scholar có khả năng tìm thấy tới 93% - 96% số trích dẫn có trên tất cả các lĩnh vực, vượt xa Scopus (35% - 77%) và WoS (27% - 73%). Ngoài khả năng thu thập được gần như tất cả các trích dẫn của WoS (95%) và Scopus (92%), Google Scholar còn bao phủ được rộng hơn cả các trích dẫn ngoài nguồn tạp chí (chiếm 48% - 65%), bao gồm luận án, sách, tài liệu hội nghị và tài liệu chưa được công bố.



Hình 9. Chỉ số trích dẫn trung bình của top 30 CSGDĐH Việt Nam giai đoạn 2014-2018.

Kết quả phân tích cho thấy rằng, chỉ số trích dẫn trung bình của các CSGDĐH Việt Nam đạt giá trị 9,2. Sau khi đối sánh và chuẩn hóa với CSDL do bảng xếp hạng QS công bố cho ĐHQGHN, có thể thấy rằng chỉ số trích dẫn trung bình của Việt Nam xấp xỉ giá trị trung bình của khu vực châu Á. Theo đó, Việt Nam có 7 CSGDĐH có chỉ số trích dẫn cao hơn chỉ số trung bình của cả nước và của châu Á là: Trường ĐH Y Hà Nội (18,1); Trường ĐH Y tế cộng đồng (16); Trường ĐH Duy Tân (11,8); Trường ĐH Mở Địa chất (11,0); ĐHQGHN (10,7); Trường ĐH Kinh tế Quốc dân (10,6) và Trường ĐH Tôn Đức Thắng (10,3). Ngoài trường hợp các trường đại học thuộc lĩnh vực khoa học sức khỏe, chỉ số trích dẫn trung bình cao của Trường ĐH Mở Địa chất và Trường ĐH Kinh tế quốc dân là điều đáng ghi nhận. Các nghiên cứu của hai trường vừa nêu có thể vừa có tính cập nhật, đồng thời liên quan đến các vấn đề đặc thù của Việt Nam nên được các nhà khoa học trên thế giới quan tâm nhiều.

4. Kết luận

Công nghệ tích hợp, phân loại sự trùng lặp các bài báo trong CSDL WoS và Scopus đã được phát triển thành công, cho phép sử dụng để thiết lập CSDL tích hợp WoS & Scopus về các bài báo quốc tế của các tác giả Việt Nam, tạo điều kiện triển khai việc thống kê, phân tích đầy đủ hơn năng suất và chất lượng công bố quốc tế của Việt Nam trong tất cả các lĩnh vực. Với CSDL này, lần đầu tiên số liệu đầy đủ về tình hình công bố quốc tế WoS & Scopus và chỉ số trích dẫn được xác lập. Theo đó, năm 2018, số lượng công bố quốc tế hàng năm của Việt Nam đã cán mốc 10.000 bài, trong đó các CSGDĐH đã đóng góp 70%. Gần đây (năm 2017-2018), tốc độ gia tăng các bài báo WoS & Scopus của Việt Nam tăng mạnh (34,7% đối với cả nước và 41,6% đối với riêng các CSGDĐH). Đây là thời kỳ mà GDĐH được giao quyền tự chủ cao. Bên cạnh việc phân tích đóng góp của hệ thống GDĐH, sự gia tăng của các trường đại học trẻ như Trường ĐH Tôn Đức Thắng và Trường ĐH Duy Tân cũng đã được đề cập. Đóng góp của các trường đại học

này cũng đã làm thay đổi tình hình nghiên cứu và công bố quốc tế của các trường đại học khu vực Hà Nội và Tp Hồ Chí Minh. Về chất lượng nghiên cứu, Việt Nam có chỉ số trích dẫn bằng 9,2 - ngang mức trung bình của các CSGDĐH châu Á. Phân tích này cũng đã định vị được nhóm các CSGDĐH học định hướng và có năng lực nghiên cứu khoa học. Trong chiến lược phát triển, Việt Nam có thể tập trung đầu tư vào nhóm các CSGDĐH này cùng với các lĩnh vực, các hướng nghiên cứu có tầm ảnh hưởng lớn. Đặc biệt, công trình này cũng chỉ ra mối liên hệ giữa tự chủ đại học và việc nâng cao năng suất và chất lượng nghiên cứu khoa học.

Lời cảm ơn

Công trình này được tài trợ chính bởi Đề tài QG.18.62 của ĐHQGHN và có sự phối hợp với đề tài KHGD/16-20/07 thuộc Chương trình KH&CN cấp quốc gia giai đoạn 2016-2020 “Nghiên cứu phát triển KHGD đáp ứng yêu cầu đổi mới căn bản, toàn diện nền giáo dục Việt Nam”. Các tác giả cảm ơn Giáo sư Nguyễn Hữu Đức đã đặt hàng và trực tiếp chỉ đạo triển khai công trình nghiên cứu này.

Tài liệu tham khảo

- [1] J. Mingers, L. Leydesdorff, A review of theory and practice in scientometrics, *European journal of operational research* 246 (2015) 1-19.
- [2] K. Frenken, S. Hardeman, J. Hoekman, Spatial scientometrics: Towards a cumulative research program. *Journal of informetrics* 3 (2009) 222-232.
- [3] J. Priem, B.H. Hemminger, *Scientometrics 2.0: New metrics of scholarly impact on the social Web*. First Monday 15, 2010.
- [4] W.C.M. Mattens, N.H.T. Chung, N.H. Duc, *Bibliometric Analysis and Research Management*. *VNU Journal of Science: Policy and Management Studies* 31 (2) (2015) 1-13.
- [5] Nguyen Huu Duc, Nghiem Xuan Huy, Nguyen Huu Thanh Chung, *Bibliometric Analysis of World Innovation Trends in the Fourth Industrial Revolution (in Vietnamese)*, *VNU Journal of Science, Policy and Management Studies* 33 (3) (2017) 1-7.

- [6] Trong Hien Ton Nguyen, Industry 4.0 from technology adoption perspective: Evaluation based on bibliometric (in Vietnamese), DTU Journal of Science and Technology 34 (2019) 29-35
- [7] Chinese Science Citation Database, https://support.clarivate.com/ScientificandAcademicResearch/s/article/Chinese-Science-Citation-Database-Journal-List?language=en_US (accessed 11th June, 2019).
- [8] Russian Citation Science Index, <https://clarivate.libguides.com/webofscienceplatform/rsci> (accessed 11th June, 2019).
- [9] Korea Citation Index, <https://www.kci.go.kr/kciportal/main.kci?locale=en> (accessed 11th June, 2019).
- [10] Malaysia Citation Index, <http://www.mycite.my/> (accessed 11th June, 2019).
- [11] Web of Science, <https://clarivate.com/webofsciencegroup/> (accessed 11th June, 2019).
- [12] Scopus, <https://www.scopus.com/search/form.uri?display=basic> (accessed 11th June, 2019).
- [13] Google Scholar, <https://scholar.google.com/> (accessed 11th June, 2019).
- [14] Web of Science, <https://clarivate.com/webofsciencegroup/solutions/web-of-science/> (accessed 11th June, 2019).
- [15] Elsevier, <https://www.elsevier.com/solutions/scopus> (accessed 11th June, 2019)
- [16] Nguyen Huu Duc, Vo Dinh Hieu, Vietnam Citation Gateway, Copyrights, No 2179/2018/QTG, Ministry of Culture, Sports and Tourism of Vietnam (in Vietnamese)
- [17] Y. Jiang, C. Lin, W. Meng, C. Yu, A.M. Cohen, and N.R. Smalheiser). Rule-based deduplication of article records from bibliographic databases. Database: The journal of biological databases and curation, 2014, bat086. doi:10.1093/database/bat086
- [18] A. Martín-Martín, E. Orduna-Malea, M. Thelwall, E.D. López-Cózar, Google Scholar, Web of Science, and Scopus: A systematic comparison of citations in 252 subject categories, Journal of Informetrics 12(4) (2018) 1160-1177.