Journal of Scientometric Res. 2022; 11(2):146-154
http://www.jscires.org

Research Article

A Comparative Analysis of Unified Informetrics with Scopus and Web of Science

Parul Khurana[1], Geetha Ganesan[2], Gulshan Kumar[3], Kiran Sharma[4,*]

[1]School of Computer Applications, Lovely Professional University, Phagwara, Punjab, INDIA.
[2]Advanced Computing Research Society, Porur, Chennai, INDIA.
[3]School of Computer Science and Engineering, Lovely Professional University, Jalandhar - Phagwara, Punjab, INDIA.
[4]School of Engineering and Technology, BML Munjal University, Sidhrawali, Gurugram, Haryana, INDIA.

ABSTRACT
Numerous bibliographic databases track the number of publications, citations, and *h*-index to maintain the progress of an individual. However, the choice of journals varies among these databases, hence they produce different numbers of publications, citations and *h*-index for the same author. Various literatures are available on the comparative analysis of such bibliographic databases stating the fact that there are the differences between bibliometrics generated by different bibliographic databases but none of the literature provides any comprehensive or complete solution to overcome such differences. At present, there is no common platform that can provide a single count of number of publications, citations and *h*-index across multiple bibliographic databases. To overcome this limitation, we propose a new method in academic research publication to calculate weighted unified (single) informetrics for an author. With the proposed solution, one can view a single article, citation, and *h*-index count computed from multiple indexing databases. In this study, the data from Scopus and Web of Science is used to generate a new database named “conflate”. Further, a comparative analysis of the proposed model is performed with Scopus and Web of Science at three levels: author, organization, and journal. The proposed model can be observed as a new single indicator to determine the research influence of the author, organization, and journal. Aim: We propose a new method in academic research publication to calculate weighted unified informetrics. With the proposed solution, one can view a single article, citation, and *h*-index count computed from multiple indexing databases. The proposed model can be observed as a new indicator to determine the research influence of author, organization, and journal.

Keywords: Unified informetrics model, Bibliometrics, Citation analysis, *h*-index.

Correspondence
Kiran Sharma
School of Engineering and Technology, BML Munjal University, Sidhrawali, Gurugram, Haryana, INDIA.
Email id: kiran.sharma@bmu.edu.in
ORCID ID: 0000-0002-3797-7363

Received: 24-06-2021
Revised: 07-01-2022
Accepted: 08-07-2022
DOI: 10.5530/jscires.11.2.16

INTRODUCTION

Scientific contributions work as a driving force for the continuous growth of science and society.[1] To measure the impact of such contributions, citations provide a quantitative evaluation which helps in describing publication patterns, research quantity, quality, and the influence of authors.[2,3] Citation data is often impacted by the coverage of bibliographic databases such as Scopus, Web of Science (WoS), Google Scholar etc., because they collect the citations received by the publications indexed by them only.[4,5] Hence, one can observe different bibliometrics such as number of publications, number of citations and *h*-index for same author in different bibliographic databases.

The prolific growth of bibliographic databases has created new opportunities.[6] Bibliographic databases are used worldwide to produce comparative statistics for bibliometricians[7,8] Different authors have shown the comparison between Scopus, WoS, PubMed, Google Scholar etc., in various literature based on availability of digital object identifiers (DOIs),[9] their bibliometric analysis,[10,11] their coverage,[12] their features and citation properties,[13-16] their strengths and weaknesses,[17] their content comprehensiveness and searching capabilities,[18-20] their longitudinal and cross-disciplinary comparison of coverage,[21,22] their language coverage,[23,24] their use in academic papers,[25,26] their systematic comparison of citations based on subject categories,[27] their journal coverage,[28-30] their retroactive growth comparison of universities,[31,32] *h*-index of authors,[33,34] and countries.[35] Comparative statistics provided by various authors are utilized by funding agencies, government bodies, promotion committees, ranking agencies, accreditation agencies, and other stakeholders to measure the quality and impact of authors. Hence, bibliometric analysis has emerged as a powerful tool and partial system for its stakeholders.

Table 1: List of studies included comparative analysis, based on Scopus, Web of Science, Google Scholar, etc. (sorted year wise).

Key idea or concept	Limitations identified by Authors	Any solution proposed?	Author Ref with year
To compare the major features of the Web of Science, Scopus, and Google Scholar as a citation database.	Traditional indexing databases lacks proper subject indexing, hence there is a need to introduce an idea which serves as a solution to citation-based searching for quantitative evaluations.	×	[1] 2005
To compare strength and weaknesses of PubMed, Scopus, Web of Science, and Google Scholar databases.	Databases are compared in context of their content and various practical aspects. No such limitations are mentioned.	×	[17] 2008
To compare h-indices of highly cited researchers of Israel based on their citation count in Web of Science, Scopus and Google Scholar.	Disciplinary differences in coverage and differences in citation counts are observed across databases.	×	[33] 2008
To gauge the comparability in determining the *h*-index from Scopus and Web of Science for 10 universities.	Significant differences are observed in the content and cited references from both databases.	×	[35] 2009
To compare two instruments like Scopus and Web of Science for a typical university in Portugal.	Different abstracting policies, and apparent errors in constructing the databases are identified.	×	[31] 2009
Three citation databases are compared with reference to book - introduction to informetrics.	Findings clearly reveal those citations across databases are clearly comparable. But there is no single citation database that can supplement other.	×	[14] 2010
Three citation resources are compared to find the one with most representative citation coverage.	Results show that there is a variation in the retrieved data in context of citation counts.	×	[20] 2013
Journal coverage across Scopus and Web of Science is described.	Results indicate that use of either of these databases for research evaluation may be biased. Hence both should be used with caution.	×	[28] 2016
Systematic and comprehensive comparison of coverage across Scopus, Web of Science and Google Scholar is provided.	All three databases provide sufficient coverage for cross disciplinary comparisons. But results show that specific metrics change the conclusions across databases.	×	[21] 2016
A light has been shed on the availability of DOIs in Scopus and Web of Science in publication items.	Both databases lack the 100% availability of DOIs, hence authors are encouraged for DOI based establishments.	×	[9] 2016
Research publication data from Web of Science is taken for Indian central universities from 1990-2014 for the study of their ranking and policy purposes.	Study introduces the idea of quality-quantity composite index for central universities in India. At the end, a generalized model using variables and their weights is also proposed as an optimal solution of ranking.	Partial	[45] 2016
Google Scholar, Web of Science, and Scopus are compared based on 252 subject categories.	Study provides evidence that Google Scholar has more citations as compared to Scopus and Web of Science. Google Scholar may be seen as a super set of Scopus and Web of Science.	Partial	[27] 2018
Bibliometrics based on highly cited documents in Scopus, Web of Science and Google Scholar is explored.	Study demonstrates that these databases miss a significant amount of information (if compared) based on counts of highly cited publications.	×	[12] 2018
Publications for Jordanian authors are studied based on literature databases such as Scopus, Web of Science, PubMed etc.	Results show that Scopus, Web of Science, PubMed etc., have differences in terms of their coverage, focus and the tools.	×	[11] 2019
Web of Science and Scopus are compared based on their language coverage of publications.	Results obtained at document level, languages and key areas are different from journal level analysis for both Scopus and Web of Science.	×	[24] 2019
Retroactive growth, correlation, and coverage of universities is validated based on Scopus, Web of Science and Google Scholar.	Institutional productivity varies across Google Scholar, Scopus and Web of Science in terms of total number of publications.	×	[32] 2019
Comparative, dynamic, and empirical study is presented based on academic papers available in Scopus and Web of Science.	A deeper analysis based on the content of Scopus and Web of Science requires further investigation.	×	[26] 2020
Citation coverage is presented based on Google Scholar, Microsoft Academic, Scopus, Dimensions, Web of Science, and Open Citations COCI.	Results reveal that no single database is adequate as a bibliographic database. Future studies may reveal that which data source is most suitable for the needs of stakeholders.	Partial	[22] 2021
Comparative analysis of journal coverage is aimed for Scopus, Web of Science and Dimensions.	Results indicate that databases have significantly different journal coverage.	×	[30] 2021
Impact of author ranking based on Scopus and Web of Science is introduced with an improvement to *h*-index.	Results reveal that there is significant difference between *h*-index calculated in Scopus and Web of Science.	Partial	[34] 2022

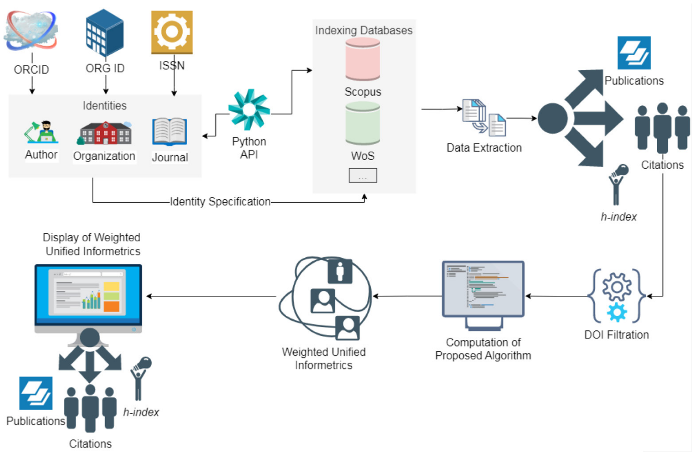

Figure 1: Schematic representation of the proposed weighted unified informetrics.

Motivation

Bibliometrics enable the global knowledge for categorization and practical analysis of research contributions through bibliometric data sources such as Scopus, Web of Science, etc. The generated dataset is accessed for many purposes, analytic functions, fundings and scientific extensions. Recognized scientific characteristics and investigations reveal that bibliometric data sources such as Scopus, and Web of Science presents the different performative characteristics (see Table 1) such as, different number of papers, citations and h-index of same scientific inputs.

Hence, this study provides a one-stop solution for various stakeholders to provide single paper, citation and h-index count for scientific individuals or groups across multiple bibliometric data sources. Here, we are using the DOI filtration to check the availability of publication in multiple bibliographic databases and aggregation to provide the actual and authentic count of informetrics.

Research Gap

Table 1 presents the key areas and limitations listed in the studies performed on comparative analysis of bibliographic databases. Results reveals that few studies have tried to provide such as[22,27,34,45] a partial solution but none of the literature provides any comprehensive or complete solution to overcome the limitations of bibliographic databases.

Due to such limitations, universities, accreditation agencies, ranking agencies, and hiring agencies ask authors to provide publications, citations, and *h*-index count of all bibliographic databases separately during their job applications as well as in their assessments. There is no common platform that can record or calculate single informetrics across multiple bibliographic databases. This situation has raised a requirement of bibliometrics where an author can provide a single count of number of publications, number of citations and *h*-index across different bibliographic databases. Author should not provide different set of publications, citations and *h*-index representing bibliographic databases like Scopus, WoS, etc. Hence, we propose an algorithm (Figure 1) to calculate bibliometrics for the single count of publications, citations, and *h*-index for authors to various stakeholders.

The objectives of our study are:

1. To propose a common platform that can provide a single article count, citation count, and *h*-index in the education field.
2. To check the statistical validity of the proposed platform in terms of the number of articles, the number of citations, and *h*-index at author, organization, and journal level.

The Weighted Unified Informetrics (WUI) Algorithm: In the proposed algorithm, we have used bibliographic databases such as Scopus and WoS due to their indexing age, availability of data, and authenticity. A weighted unified informetrics system named "conflate" has been discussed and proposed (Figure 1).

METHODS

Generation of doi based citation database: For data extraction from both Scopus and WoS, we require inputs at three levels. An ORCID ID is required for authors' information, organization name for university/institute access, and ISSN for journal information. Based on the author's ORCID ID, we retrieved the number of publications from both Scopus

Algorithm 1: Generation of doi based citation database	
	Input: Orcid ID
	Output: doi based citation database
1	**for** each author, **do**
	[A*i*] ∈ DB*i*, where i =1, ..., *N*, *N* >0
	/* [A*i*] is list of articles, doi numbers in database DB*i* */
2	**for** each *doi* in [A*i*], **do**
	[C_N] := list of citations
	for each citation in C_N, **do**
	if *doi* exists then
	CD*i* := *doi*
	/* CDi is *doi* based citation database, computed on [A*i*] */
3	Repeat step 1 and 2 to get A_1, …., A_N and CD_1, …, CD_N from DB_1, …, DB_N
	/* merge all citation databases for a given author */
4	$CD_{all} := CD_1 \cup CD_2 \cup \ldots \cup CD_N$
	/*where CD_{all} contains only those citations for a given author whose *doi* exists
	(including duplicates) */

Algorithm 2: Computation of weighted unified informetrics	
	Input: CD_{all} : Conflate citation database
	Output: Weighted unified informetrics and research indicators
1	$CDcommon := CD_1 \cap CD_2 \cap \ldots \cap CD_{all}$
2	CD*unique* := CD*all* – CD*common*
3	**for** each *doi* in CD*unique*, **do**
	P := Count (*doi*) in CD*all*
	W_{doi}=P/N
	/* *N* is number of citation databases, *N*>0 */
4	Compute *h*-index
5	Display number of publications, citations and *h*-index for a given author

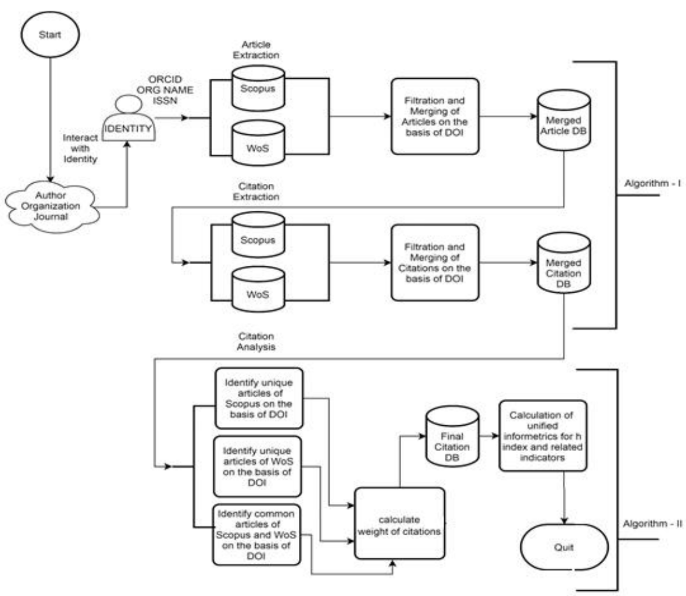

Figure 2: Flowchart demonstrates the computation of weighted unified informetrics.

and WoS. Further, we filtered those publications based on the DOI and got the required data set for the given author for both Scopus and WoS. The next step is to look after the number of citations received by the given author for each publication. We examined every citation received on all publications for a given author and filtered only those citations that have DOI associated with them. The necessary condition of DOI is used to match the given record among multiple databases. The collection of similar records and respective citations based on DOI for a given author among Scopus and WoS will lead to a new filtered database named "conflate". Algorithm 1 describes the steps for extracting article and citation details for the given ORCID ID. Similarly, the process is repeated at the organization and the journal levels.

Computation of weighted unified informetrics

The conflate database generated in Algorithm 1 is used for further computation of weighted unified informetrics. First, for a given author, common and unique citations will be filtered. Then, a weight is assigned for the final calculation of unified informetrics. Algorithm 2 describes the process in sequence and Figure 2 summarizes the computation of weighted unified informetrics for different entities.

Algorithm 2: Describes the process in a sequence which will be repeated for organization and journals as well.

Data extraction

To perform this study, data from Scopus and WoS was extracted with Python-based APIs. For data extraction from Scopus, we have used Python based API wrapper named as "pybliometrics". It is an easy to use library to pull, cache and extract data from Scopus database [36]. Scopus database access is based on API keys available at (https://dev.elsevier.com/apikey/manage). After successful creation of user account on Elsevier Developer Portal, anyone can obtain API key for programmatic access to citation data and abstracts, journals, research metrics and related metadata indexed by Scopus citation database. For data extraction from WoS, we have used Python client named as "wos python client". It is a SOAP (Simple Object Access Protocol) based client for querying WoS database to retrieve results in the format of XML [37]. Web Services Premium access, which is a paid service of WoS is required to extract data from WoS citation database (https://developer.clarivate.com/).

Data Description

Data selection is performed at three levels:

1. *Author level:* 400 faculty profiles out of 6316 profiles from various disciplines are accessed from "Monash University", a public university in Melbourne, Australia (https://research.monash.edu/en/persons/). The choice of

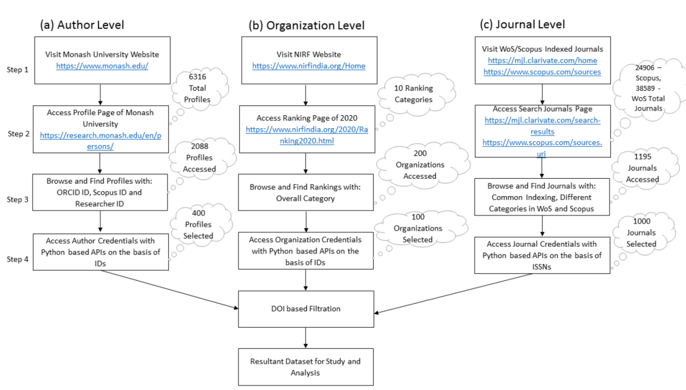

Figure 3: Flowchart demonstrates the process of visiting the author's, organization's, and journal's profile.

the data is mainly due to the openly available information of faculties especially ORCID ID, Scopus ID, and WoS ID. All 400 selected faculty profiles have ORCID ID, Scopus ID, and WoS ID. 91.5% documents in Scopus and 82.3% in WoS are with DOI.

2. *Organization level:* Top 100 Indian institutes out of 200 ranked institutions based on *National Institutional Ranking Framework* (NIRF) (https://www.nirfindia.org/) are accessed. 83.6% documents in Scopus and 77.2% in WoS are with DOI.

3. *Journal level:* A random selection of 1000 journals listed in both Scopus and WoS is used. 92.5% documents in Scopus and 84.2% in WoS are with DOI.

Data filtration is performed on the basis of DOI (Digital Object Identifier). It provides a unique authentication to the publications.[24] It's desirable that articles carrying DOI numbers must be considered only for any kind of evaluation and calculation to establish scientific assignments.[25] Hence, while combining articles across multiple databases, we have considered the articles with DOI numbers only (flowchart in Figure 3). To perform this study, data from Scopus and WoS was obtained with Python-based APIs.[36,37] The primary reason for data selection from Scopus and WoS is arbitrary and the availability of data.

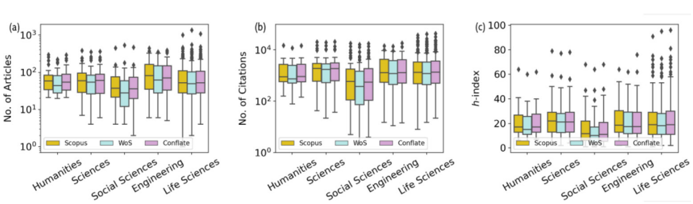

Figure 4: A comparative analysis between Scopus (coloured in golden), WoS (coloured in cyan), and conflate (coloured in pink) for 400 authors as (a) the number of articles, (b) the number of citations, and (c) *h*-index. The analysis is performed for five disciplines: Humanities, Sciences, Social Sciences, Engineering, and Life Sciences. The standard deviation recorded for Scopus: articles (94.13), citations (4324.97) and *h*-index (14.82); for WoS: articles (105.43), citations (4263.24) and *h*-index (14.56); and for conflate: articles (94.06), citations (4599.34) and *h*-index (15.16).

RESULTS

Here we have presented the comparative analysis of Scopus, WoS and conflate at author's, organization, and journal level.

Author level bibliometrics

Figure 4 shows the comparison of conflate with Scopus and WoS on the basis of the number of articles, the number of citations, and *h*-index of 400 authors. Authors have been categorized into five disciplines (number of authors) based on their work domain: *Social Sciences* (66), *Sciences* (43), *Humanities* (20), *Life Sciences* (211), and *Engineering* (60). Scopus contains the large number of articles for *Social Sciences, Sciences, Humanities, and Engineering* whereas WoS shows for *Life Sciences*. The number of articles in conflate ranges in between Scopus and WoS count, except *Life Sciences*. The large number of citations are reported in *Life Sciences* in Scopus and the smaller number of citations are reported in *Social Sciences* in WoS. For *Sciences, Engineering, and Life Sciences*, conflate has reported the highest number of citations as compared to Scopus and WoS. For the rest of the disciplines, the number of citations reported by conflate is in between the range of Scopus and WoS. For the *h*-index of 400 authors, we found that conflate has reported the same *h*-index in *Social Sciences* and *Sciences* as reported by Scopus. For *Humanities*

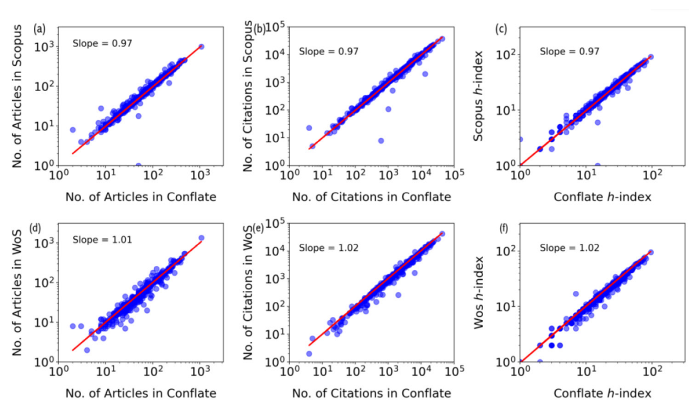

Figure 5: Comparative analysis of 400 authors between Scopus and conflate (a-c) and WoS and conflate (d-f). The number of articles and citations in Scopus (33182, 1097446), WoS (31732, 1024808), and conflate (32376, 1130306). The maximum *h*-index in Scopus (91), WoS (95), and conflate (96). The red line represents the best fit line.

Table 2: Comparative analysis of Scopus, WoS, and conflate for 400 author's articles, citations and *h*- index for five disciplines.

Authors – 400 (Disciplines)	Articles			Citations			Average *h*-index		
	Scopus	WoS	Conflate	Scopus	WoS	Conflate	Scopus	WoS	Conflate
Life Sciences	17793	18257	17951	647698	631244	680537	22	22	23
Social Sciences	3531	3113	3457	115904	94195	114158	15	13	15
Engineering	6741	5658	5940	161092	138631	162218	23	21	22
Sciences	3397	3187	3340	127009	121752	128423	24	23	24
Humanities	1720	1517	1688	45743	38986	44970	22	20	22

and *Engineering*, conflate has reported *h*–index in the range of Scopus and WoS. For *Life Sciences*, Scopus and WoS have reported the same *h*–index whereas conflate has reported one point higher of both.

Figure 5 shows the comparative analysis of the number of articles, citations and *h*–index between Scopus, WoS, and conflate. The best fit line (coloured in red) shows the less variation among the Scopus and conflate, and WoS and conflate. The overall slope is higher in WoS. In the comparative analysis of Scopus and conflate (Figure 5 (a–c)), it is observed that the average number of articles published by an author is 83, whereas in conflate it is 81. In Scopus, the average number of citations an author received is 2744 whereas in conflate it is 2826. The average *h*–index of an author in Scopus is 21 and 22 in conflate. Although the average number of articles calculated in Scopus is less than conflate; however, an average number of citations and average *h*–index are higher in conflate. Similarly, in the comparative analysis of WoS and conflate (Figure 5 (d–f)), the average number of articles published is 79 as compared to the average of 81 articles per author in conflate. The average number of citations published in WoS is 2562 as compared to 2826 in conflate. Average *h*–index per author in WoS is 20 whereas in conflate it is 22. Table 2 represents the comparative analysis of Scopus, WoS, and conflate for 400 author's articles, citations and *h*– index among different disciplines.

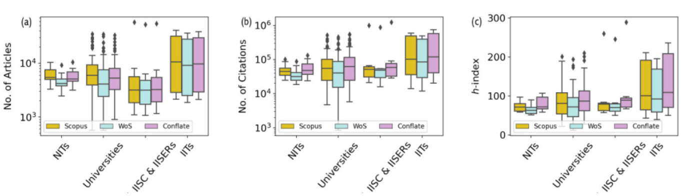

Figure 6: A comparative analysis between Scopus (coloured in golden), WoS (coloured in cyan), and conflate (coloured in pink) for top 100 Indian institutes as (a) the number of articles, (b) the number of citations, and (c) *h*-index. The analysis is performed for four categories of institutes: NITs, Universities, IISC & IISER, and IITs. The standard deviation recorded for Scopus: articles (10459.32), citations (161121.19), and *h*-index (47.36); for WoS: articles (9637.33), citations (138670.18), and *h*-index (44.24); and for conflate: articles (9774.53), citations (197156.68), and *h*-index (51.75).

Organization level bibliometrics

Here we analyzed the top 100 organizations in India and the categorization is done on the basis of their entity specification (count): Universities (69), IITs (16), NITs (8), and IISC & IISER (7). It is observed that the highest number of articles published among different databases are from IITs and the lowest number of articles are from NITs. Conflate reported that the number of articles published among different databases is varying between Scopus and WoS across all entities. In all entities, conflate reported the highest number of articles as compared to WoS and the lowest number of articles as compared to Scopus. Conflate reported the highest number of citations as compared to Scopus and WoS for all entities. In comparison with Scopus and WoS, Scopus has always reported a greater number of citations as compared to citations reported by WoS. *h*–index reported by conflate is also highest among both databases. IITs have received the highest *h*–index and NITs have received the lowest *h*–index among other entities. Conflate also reported that the results generated are always in between the range of Scopus and WoS. Among four entities, it can be observed that IITs have the highest *h*–index across multiple databases as shown in Figure 6. Further, one can analyze the different disciplines of these organizations to keep track of the most popular discipline in terms of research publications.

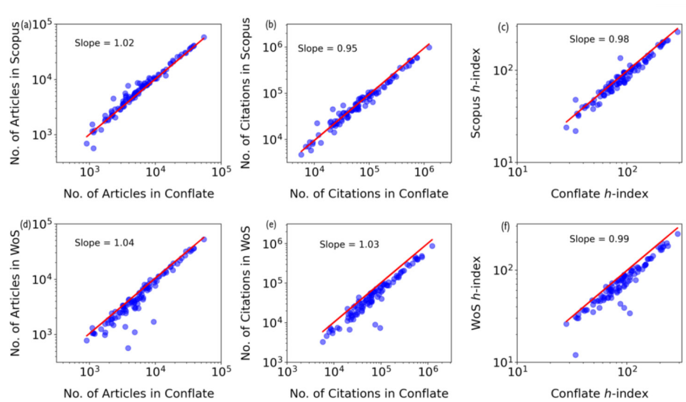

Figure 7: Comparative analysis of top 100 organizations between Scopus and conflate (a-c) and WoS and conflate (d-f). The number of articles and citations in Scopus (964093, 11399909), WoS (797158, 9337059), and conflate (873719, 13483112). The maximum *h*-index in Scopus (260), WoS (246), and conflate (289). The red line represents the best fit line.

Table 3: Comparative analysis of Scopus, WoS, and conflate for articles, citations and *h*-index for 100 organizations categorized in to 4 main head organizations.

Organizations – 100 (Type Wise)	Articles			Citations			Average *h*-index		
	Scopus	WoS	Conflate	Scopus	WoS	Conflate	Scopus	WoS	Conflate
NITs	50362	39059	47683	416302	318676	492760	74	66	80
Universities	565887	451489	499159	6051897	4917831	6997951	85	76	93
IISC & IISER	77349	70063	73835	1252987	1107018	1542242	98	92	111
IITs	270495	236547	253042	3678723	2993534	4450159	122	110	134

Figure 7 Shows the comparative analysis between Scopus, WoS, and conflate for top 100 Indian organizations. In the comparative analysis of Scopus and conflate (Figure 7 (a–c)), it is observed that the average number of articles in Scopus is 9641 as compared to 8737 in conflate. The difference in the average number of articles states that all articles published in Scopus are not considered in conflate. The average number of citations recorded in Scopus is 113999 as compared to 134831 in conflate. Average *h*-index calculated in Scopus for these organizations is 91 which is quite lesser than the average *h*-index (100) calculated in conflate. Similarly, in the comparative analysis of WoS and conflate (see Figure 7 (d–f)), it is observed that the average number of articles in WoS is 7971 whereas in conflate it is 8737. Conflate also reported a significantly higher number of citations with an average score of 134831 as compared to 93371 in WoS. Average *h*-index in conflate is also 100 which is quite higher than average *h*-index 82 reported by WoS. Table 3 represents the comparative analysis of Scopus, WoS, and conflate for articles, citations and *h*- index for 100 organizations categorized into 4 main head organizations.

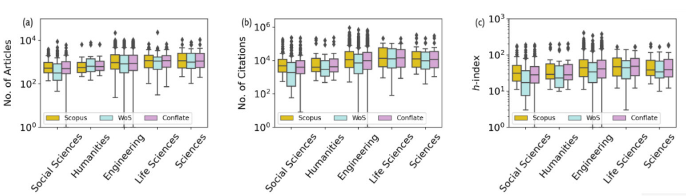

Figure 8: A comparative analysis between Scopus (coloured in golden), WoS (coloured in cyan), and conflate (coloured in pink) of 1000 journals as (a) the number of articles, (b) the number of citations, and (c) *h*-index. The analysis is performed for five disciplines: Humanities, Sciences, Social Sciences, Engineering, and Life Sciences. The standard deviation recorded for Scopus: articles (1794.63), citations (59090.78), and *h*-index (44.40); for WoS: articles (1819.10), citations (45753.64), and *h*-index (40.06); and for conflate: articles (1677.79), citations (55723.79), and *h*-index (45.13).

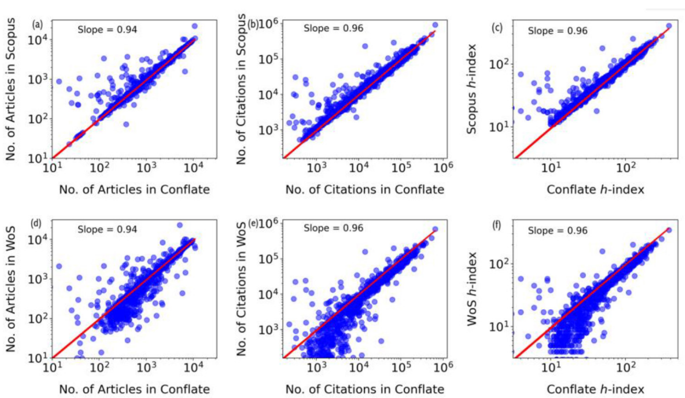

Figure 9: Comparative analysis of 1000 journals between Scopus and conflate (a-c) and WoS and conflate (d-f). The number of articles and citations in Scopus (1528904, 30964292), WoS (1415093, 22570461), and conflate (1481823, 29276118). The maximum *h*-index in Scopus (408), WoS (344), and conflate (381). The red line represents the best fit line.

Journal level bibliometrics

Here we analyzed 1000 journals and broadly divided into 5 disciplines (journal count), *Engineering* (800), *Social Sciences* (119), *Life Sciences* (35), *Sciences* (27), and *Humanities* (19). The number of articles observed in *Sciences* is highest in Scopus with lowest in *Social Sciences*. For *Social Sciences*, conflate reported the highest number of articles among Scopus and WoS. For *Humanities, Engineering* and *Sciences*, conflate has reported a number of articles in between Scopus and WoS. For *Life Sciences*, conflate has reported almost the same number of articles as compared to Scopus which is quite lesser than the WoS database. The number of citations reported by conflate is in between the range of Scopus and WoS for all disciplines where sciences is on top and social sciences is at the bottom *h*-index reported by conflate for 1000 journals is the same as reported by Scopus for *Humanities, Sciences*, and *Life Sciences*. For *Social Sciences* and *Engineering*, it is in between the range of Scopus and WoS. Lowest *h*-index is reported by WoS for *Social Sciences* and highest by Scopus for *Life Sciences* as shown in Figure 8.

Figure 9 Shows the comparative analysis between Scopus, WoS, and conflate for 1000 journals. In the comparative analysis of Scopus and conflate (Figure 9 (a–c)), it is observed that the average number of articles in Scopus is 1529 and in conflate is 1482. There is a slight hike in the average number of articles in Scopus. Similarly, the average number of citations in Scopus is 30964 and in conflate is 29276. Average *h*-index calculated in Scopus is 56 and in conflate is 53. Similarly, in the comparative analysis of WoS and conflate (Figure 9 (d–f)), it is observed that the average number of articles in WoS (1415) as compared to conflate (1482) shows significantly close values. The average number of citations in WoS is

Table 4: Comparative analysis of Scopus, WoS, and conflate for 1000 journals articles, citations and *h*-index for different disciplines.

Journals – 1000 (Disciplines)	Articles			Citations			Average *h*-index		
	Scopus	WoS	Conflate	Scopus	WoS	Conflate	Scopus	WoS	Conflate
Life Sciences	60061	68142	60049	1144780	858220	1130871	59	50	59
Social Sciences	94755	86719	98316	1712570	1268836	1569405	42	30	38
Engineering	1298929	1179771	1244492	26468041	19176940	24982597	58	46	55
Sciences	56390	53458	54009	1061160	783108	1030698	56	49	56
Humanities	18769	27003	24957	577741	483357	562547	54	46	53

22570 as compared to 29276 of conflate. Conflate clearly states that there is more scope of consideration of citations as compared to citations considered by WoS. The average *h*-index in WoS is 44 as compared to 53 in conflate. Conflate is clearly moving ahead in terms of the average number of citations and *h*-index calculation of WoS. Table 4 represents the comparative analysis of Scopus, WoS, and conflate for 1000 journals with their articles, citations and *h*- index among different disciplines.

DISCUSSION AND CONCLUSION

The key findings of the work can be summarized as follows: (i) It presents a unified method to maintain records associated with entities of author, organization, and journal. This method determines an absolute number of articles and citations for different entities. (ii) The mapping of multiple bibliographic databases for the calculation of *h*-index, and related informetrics. (iii) The proposed system facilitates its stakeholders for the establishment of a system providing a clear, authentic, and simulated environment for the research measurement of entities. (iv) Presents in-depth analyses of the core components like publications, citations, *h*-index, etc.

The presented work has some advantages as (i) The DOI-based data filtration helps us to identify the authenticity of received citations and publications (ii) Different stakeholders like government agencies, accreditation agencies, ranking organizations, and funding agencies can use the proposed system for the evaluation of the research contribution of individuals, organizations as well as journals. (iii) The proposed system is a novel system introduced with the conflate of two traditional bibliographic databases like Scopus and WoS.

The proposed informetrics provides a transparent and distributed view of the research contributors to its stakeholders. Calculated results also signify the efficiency of "conflate". Scopus and WoS have been used for the implementation due to the availability of the data. At the author level, the performance of the proposed informetrics is mainly equivalent to Scopus for the number of publications, citations, and *h*-index. On the other hand, a significant difference is visible at the organizational level. The proposed informetrics shows the gain in the number of citations and *h*-index in all organization categories; however, average performance is observed for the number of publications. At the journal level, WoS has a higher count in Humanities for the number of publications but has a lower count for citations and *h*-index, whereas the proposed informetrics gives an average performance in the number of publications and best in citations and *h*-index. For other disciplines, both Scopus and the proposed model have almost similar results. In general, the proposed informetrics will always result in the best from multiple databases.

The major limitation of the study is the fact that we have considered the publications where DOI exists. In case WoS and Scopus do not have DOI numbers for the particular publications, we will not be able to consider the publication as authentic and the author will lose publications count and their citations count as well. Moreover, it could be a citation loss for low profile authors who have their work indexed only in Scopus or in WoS. For such journals which are indexed only in Scopus or in WoS, also shows their limitations to other bibliographic databases. If an author publishes his work in a journal that is indexed in multiple bibliographic databases, there is a good chance of higher visibility of a scientific work to be read and cited worldwide. As new bibliographic databases may populate in the near future, the proposed system should support the integration of those databases into the existing system.

To conclude further, there are still several possible areas for further exploration and extension. Here are some interesting areas for possible future developments and research:

1. Different bibliographic databases: We have studied the features of two bibliographic databases such as Scopus and WoS. Hence, the performed study is limited to two bibliographic databases. One can extend the study further with the use of bibliographic databases like Google Scholar, Dimensions, Crossref, OpenAIRE, DataCite, Mendeley, Zenodo etc.[38-40] All these bibliographic databases may create conflate as per the model to calculate unified informetrics.
2. Different technological aspects: One can extend the study further with the use of "Distributed Ledger Technology" and its core elements in the research publishing industry. Distributed ledger technology has found its applications in the field of education for verification of academic records,[41] adoption of smart learning environments,[42] and in implementation of mobile-based higher education systems.[43] Features like decentralization, persistency, anonymity, and auditability of records give more confidence to its stakeholders in a system presenting a scientific work of authors, organizations, and journals.[44] Hence, using Distributed Ledger Technology in the research publication industry can be considered as a viable choice to systematically achieve a sustained system in the interest of its stakeholders.

ACKNOWLEDGEMENT

Both Scopus and WoS data has been downloaded from the Northwestern University, USA.

CONFLICT OF INTEREST

The authors declare that there is no conflict of interest.

REFERENCES

1. Sharma K, Khurana P. Growth and dynamics of Econophysics: A bibliometric and network analysis. Scientometrics. 2021;126(5):4417-36.
2. Meho LI, Sugimoto CR. Assessing the scholarly impact of information studies: A tale of two citation databases—Scopus and Web of science. Journal of the American Society for information science and technology. 2009;60(12):2499-508.
3. Archambault É, Campbell D, Gingras Y, Larivière V. Comparing bibliometric statistics obtained from the Web of science and Scopus. Journal of the American society for information science and technology. 2009;60(7):13206.
4. Martin B. The use of multiple indicators in the assessment of basic research. Scientometrics. 1996;36(3):343-62.
5. Bar-Ilan J, Levene M, Lin A. Some measures for comparing citation databases. Journal of Informetrics. 2007;1(1):26-34.
6. Abramo G, D'Angelo CA. Evaluating research: From informed peer review to bibliometrics. Scientometrics. 2011;87(3):499-514.
7. Bakkalbasi N, Bauer K, Glover J, Wang L. Three options for citation tracking: Google scholar, Scopus and Web of science. Biomedical digital libraries. 2006;3(1):1-8.
8. Neuhaus C, Daniel HD. Data sources for performing citation analysis: An overview. Journal of Documentation. 2008.
9. Gorraiz J, Melero-Fuentes D, Gumpenberger C, Valderrama-Zurián JC. Availability of digital object identifiers (dois) in Web of science and Scopus. Journal of Informetrics. 2016;10(1):98-109.
10. De Groote SL, Raszewski R. Coverage of Google scholar, Scopus, and Web of science: A case study of the *h*-index in nursing. Nursing Outlook. 2012;60(6):391-400.
11. AlRyalat SA, Malkawi LW, Momani SM. Comparing bibliometric analysis using PubMed, Scopus, and Web of Science databases. JoVE Journal of Visualized Experiments. 2019;(152):e58494.
12. Martín-Martín A, Orduna-Malea E, Delgado López-Cózar E. Coverage of highly-cited documents in Google Scholar, Web of Science, and Scopus: A multidisciplinary comparison. Scientometrics. 2018;116(3):2175-88.
13. Jacso P. As we may search: Comparison of major features of the Web of Science, Scopus, and Google Scholar citation-based and citation-enhanced databases. Current Science. 2005;89(9):1537-47.
14. Bar-Ilan J. Citations to the "Introduction to informetrics" indexed by WOS, Scopus and Google Scholar. Scientometrics. 2010;82(3):495-506.
15. Thelwall M. Dimensions: A competitor to Scopus and the Web of Science?. Journal of Informetrics. 2018;12(2):430-5.
16. Alberto Martín-Martín, Enrique Orduna-Malea, Mike Thelwall and Emilio Delgado-López-Cózar. Google scholar, Web of science, and Scopus: Which is best for me? Impact of Social Sciences Blog. 2019.
17. Falagas ME, Pitsouni EI, Malietzis GA, Pappas G. Comparison of PubMed, Scopus, web of science, and Google scholar: strengths and weaknesses. The FASEB Journal. 2008;22(2):338-42.
18. Burnham JF. Scopus database: A review. Biomedical Digital Libraries. 2006;3(1):1-8.
19. Chadegani AA, Salehi H, Yunus MM, Farhadi H, Fooladi M, Farhadi M, Ebrahim NA. A comparison between two main academic literature collections: Web of science and scopus databases. Asian Social Science. 20139(5):18-26.
20. Adriaanse LS, Rensleigh C. Web of science, Scopus and Google scholar: A content comprehensiveness comparison. The Electronic Library. 2013.
21. Harzing AW, Alakangas S. Google Scholar, Scopus and the Web of Science: A longitudinal and cross-disciplinary comparison. Scientometrics, 2016;106(2):787-804.
22. Martín-Martín A, Thelwall M, Orduna-Malea E, Delgado López-Cózar E. Google Scholar, Microsoft Academic, Scopus, Dimensions, Web of Science, and Open Citations' COCI: a multidisciplinary comparison of coverage via citations. Scientometrics, 2021;126(1):871-906.
23. Bartol T, Budimir G, Dekleva-Smrekar D, Pusnik M, Juznic P. Assessment of research fields in scopus and web of science in the view of national research evaluation in slovenia. Scientometrics, 2014;98(2):1491-504.
24. Vera-Baceta MA, Thelwall M, Kousha K. Web of Science and Scopus language coverage. Scientometrics, 2019;121(3):1803-13.
25. Li J, Burnham JF, Lemley T, Robert M. Britton. Citation analysis: Comparison of web of science®, scopus™, SciFinder®, and google scholar. Journal of electronic resources in medical libraries, 2010;7(3):196-217.
26. Zhu J and Liu W. A tale of two databases: The use of Web of Science and Scopus in academic papers. Scientometrics, 2020;123(1):321-35.
27. Martín-Martín A, Orduna-Malea E, Thelwall M, López-Cózar ED. Google scholar, Web of science, and Scopus: A systematic comparison of citations in 252 subject categories. Journal of informetrics, 2018;12(4):1160-77.
28. Mongeon P, Paul-Hus A. The journal coverage of Web of Science and Scopus: a comparative analysis. Scientometrics, 2016;106(1):213-28.
29. Liu W, Huang M, Wang H. Same journal but different numbers of published records indexed in Scopus and web of science core collection: causes, consequences, and solutions. Scientometrics, 2021;126(5):4541-50.
30. Singh VK, Singh P, Karmakar M, Leta J, Mayr P. The journal coverage of Web of Science, Scopus and Dimensions: A comparative analysis. Scientometrics, 2021;126(6):5113-42.
31. Aswathy S, Gopikuttan A. Bibliometric observation of publication output of university teachers: A study with special reference to physics. Journal of Scientometric Research, 2015;4(1):14-19.
32. Bilir S, Gogu E, Ta OO, Yontan T. A new ranking scheme for the institutional scientific performance. Journal of Scientometric Research, 2015;4(2):70-6.
33. Bar-Ilan J. Which *h*-index? A comparison of WoS, Scopus and Google Scholar. Scientometrics, 2008;74(2):257-71.
34. Khurana P, Sharma K. Impact of h-index on author's rankings: an improvement to the h-index for lower-ranked authors. Scientometrics, 127(8), 4483-4498, 2022.
35. Jacsó P. The *h*-index for countries in Web of Science and Scopus. Online information review, 2009.
36. Rose ME, Kitchin JR. Pybliometrics: Scriptable bibliometrics using a python interface to Scopus. SoftwareX. 2019;10:100263.
37. Bacis E. Enricobacis/Wos, 2019. URL https://github.com/enricobacis/wos.
38. Visser M, Eck NJ, Waltman L. Large-scale comparison of bibliographic data sources: Scopus, Web of science, Dimensions, Crossref, and Microsoft academic. Quantitative Science Studies. 2021;2(1):20-41.
39. Peters I, Kraker P, Lex E, Gumpenberger C, Gorraiz JI. Zenodo in the spotlight of traditional and new metrics. Frontiers in Research Metrics and Analytics. 2(2017):13,2017.
40. Sangwal K. Recent growth of scientific journals published in India: Some publishing and citation-related characteristics. Journal of Scientometric Research. 2013;2(1):59-69.
41. Aamir M, Qureshi R, Khan FA, Huzaifa M. Blockchain based academic records verification in smart cities. Wireless Personal Communications. 2020;1-10.
42. Ullah N, Al-Rahmi WM, Alzahrani AI, Alfarraj O, Alblehai FM. Blockchain technology adoption in smart learning environments. Sustainability. 2021;13(4):1801.
43. Arndt T, Guercio A. Blockchain-based transcripts for mobile higher-education. International Journal of Information and Education Technology. 2020;10(2):84.
44. Deshpande A, Stewart K, Lepetit L, Gunashekar S. Distributed ledger technologies/blockchain: Challenges, opportunities and the prospects for standards. Overview report The British Standards Institution (BSI). 2017;40:40.
45. Basu A, Banshal SK, Singhal K, Singh VK. Designing a Composite Index for research performance evaluation at the national or regional level: ranking Central Universities in India. Scientometrics 107. 2016;3:1171-1193.