2010-05

Research by the Numbers: Assessing the Performance of Three Products for Citation Analysis

Conte, Marisa; Song, Jean; Townsend, Whitney https://hdl.handle.net/2027.42/94430

Downloaded from Deep Blue, University of Michigan's institutional repository



Objective

Citation analysis projects constitute a significant portion of the research requests submitted to the Taubman Health Sciences Library A random sample of 50 author names was generated from a prior citation analysis request submitted from an administrative unit at the University of Michigan (UM) Health System and conducted by a THL librarian. The 50 author names were run against each of the three (THL). Researchers and administrators are increasingly interested in citation analysis as a way to describe research performance designated databases: Collexis Research Profiles, ISI Web of Science, and Scopus using the specific search strategies detailed below. or impact, with particular interest on the cited-by and h-index metrics. Several commercial products provide access to citation Citation analysis metrics for total number of publications, h-index and total cited-by citations were recorded. For the graphs displayed analysis metrics, but users are often unaware of these features and generally do not understand what the numbers really mean or the context in which they can be appropriately applied. The purpose of this study is to compare the citation analysis functions of below, a further random sample of 10 names was selected. Collexis Research Profiles, ISI Web of Science, and Scopus in terms of accuracy, coverage, and overall functionality.

Background

Common Citation Metrics

Coverage and Data Sources

Metric	Definition	Advantage	Disadvantage	Resource	h-index	Total Citation	Cited By counts
Number of	The total number of	Measures	Does not measure			counts	
Papers	papers published by	productivity	importance or	Collexis Research	ISI Web of Science	Total PubMed/	ISI Web of Science
•	a researcher		impact	Profiles	MEDLINE subset	MEDLINE content	MEDLINE subset
Number of	The total number of	Measures total	May be inflated by	ISI Web of	Depth of	Total Web of	Total Web of
Citations	citations that cite a researcher's papers	impact	big hits	Science	subscription coverage	Science content	Science content
h-index	Distribution of citations to a researcher's publications	Addresses quality and quantity	Misinterpretation: low h-index does not necessarily reflect under-achievement	Scopus	Scopus database content 1996 - present	Total Scopus database content (100% MEDLINE coverage)	Total Scopus database content (100% MEDLINE coverage)



Diagrammatic View of h-index¹

- x-axis: number of publications
- y-axis: number of citations
- h: point at which number of citations = number of papers

¹http://en.wikipedia.oeg/wiki/H-index

Citation Analysis Functions Summary

citation / analysis	r arrectoris sammary		
Citation Analysis	Collexis Research Profiles	Scopus	ISI Web of Science
Author disambiguation	UM's instantiation involves time-consuming manual curation	Limited author record marking	Distinct author sets are incomplete; Ambiguity still exists within author sets
Metrics location	Expert Overview and Publications page	Citation Overview (citation tracker button)	Citation Report (create citation report link)
h-index	Limited to ISI Web of Science MEDLINE subset	Different values for the same author depending on database function	Limited to institutional database subscription coverage
Total publications	Time intensive manual calculation	Limited to Scopus database coverage	Limited to ISI Web of Science database coverage
Cited by	Counts must be manually totaled, introducing possibility of error	Limited to Scopus database coverage	Limited to ISI Web of Science database coverage

Research by the Numbers: Assessing the Performance of Three Products for Citation Analysis

Marisa L. Conte*, MLIS; Jean C. Song**, MSI; Whitney A. Townsend, MLIS Taubman Health Sciences Library; University of Michigan; Ann Arbor, MI

Methodology

Search Strategies:

Collexis Research Profiles:

- Using 'By Last Name' search feature, enter author's last name
- Select appropriate author from list
- Record total number of publications and h-index from Researcher Profile
- Manually calculate cited-by citation counts from Researcher Profile

- Using 'Author' search field, enter author's last name and first initial with an asterisk
- If further disambiguation is necessary, enter first and middle initial with an asterisk
- Select relevant author sets using the 'Distinct Author Sets' tool
- Use the 'Create Citation Report' tool to generate citation report
- Record total number of publications, h-index, cited-by citation counts

Scopus:

- Using 'Author Search' feature, enter author's last name and first initial
- If further disambiguation is necessary, enter full first name
- Select relevant records and select 'Citation Tracker' feature
- Record total number of publications, h-index, cited-by citation counts

Selected Results and Discussion Points

- In general, for the sample subset of 10 random researchers used to display the data for this poster:
- the cited by metric was higher in Scopus than the other two tools
- the total publications metric was higher in ISI Web of Science than the other two tools variability in one citation metric was an indicator of variability in the other citation metrics
- Given that ISI Web of Science is the data source for Collexis' citation metrics, the variability between the h-index and cited by citation metrics was greater than expected. It is presumed that this is due to the limitations of the ISI Web of Science author sets, which will be further investigated.

	Citation M	etrics for	Random S
Name		h-index	
	Collexis	Scopus	Web of Science
Chen, Y	4	26	10
Chugh, A	23	25	25
Cronin, P	6	6	5
agle, K	60	60	63
Good, E	20	22	21
lussain, H	13	13	11
Nallamothu, B	23	26	21
Neubig, R	35	30	38
Vang, M	4	7	13
ahuranec, D	5	5	5









Collexis Research Profiles Misek, David E
Lubman, David M
Nesvizhskii, Alexe... Coauthor Network Research Network

ISI Web of Science Distinct Author Sets

Veb of Knowledge™						
s	Select a Dat	abase	Web of Science	Additional Resources		
ite	d Reference Searc	h Advanc	ed Search Search History Ma	rked List (0)		
ci	ence® – with C	Conference	Proceedings			
su	mmary					
ct	Author Set	s: omer	nn g*			
e u elo	Author Set feature i sed as a tool to foc w are sets of pape earch for this autho	is a discover us your sear ers for omen or name	y tool showing sets of papers likely ch rather than as a definitive list of a n g*	written by the same person. C specific author's works. (Tell	itation data is ar me more.)	alyzed to create these sets. This feature
01	Sets for omen	in g*	Page	1 of 3 Go >>		
e o	r more sets and c	lick "View F	Records".			
liev	w Records Cl	lear All	How do I provide feedback?			
t	Author Names	# of Records	Source Titles (top 5 by record count)		Publication Years	Subject Areas (top 5 by record count)
	Omenn, GS Omenn, G	155	- CANCER EPIDEMIOLOGY BIOMARKERS & PREVENTION (12) - MOLECULAR & CELLULAR PROTEOMICS (10) - PROTEOMICS (10) - JOURNAL OF THE NATIONAL CANCER INSTITUTE (7) - ANNAL SO THE NEW YORK ACADEMY OF SCIENCES (4)		1970 - 2009	- Life Sciences & Biomedicine (143) - Multidisciplinary Science & Technology (10) - Social Sciences (7) - Physical Sciences (1)
	OMENN, GS	25	- NEW ENGLAND JOURNAL OF MEDIC	NE (25)	1965 - 1968	- Life Sciences & Biomedicine (25)
	Omenn, GS	19	- AMERICAN JOURIAL OF ROENTGENOLOGY (3) - JAMA-JOURIAL OF THE AMERICAN MEDICAL ASSOCIATION (3) - CLINICAL RESEARCH (2) - ACADEMIC MEDICINE (1) - AMERICAN JOURNAL OF MEDICAL GENETICS (1)		1977 - 2002	- Life Sciences & Biomedicine (18) - Multidisciplinary Science & Technology (2) - Social Sciences (1)
	Omenn, GS	8	- BUCHEMISTRY (3) - JOURNAL OF BACTERIOLOGY (1) - JOURNAL OF THE AMERICAN CHEMISTRY (1) - JOURNAL OF THE AMERICAN CHEMICAL SOCIETY (1) - NATURE (1)		1968 - 1971	- Life Sciences & Biomedicine (5) - Multidisciplinary Science & Technology (2) - Physical Sciences (1)
	OMENN, GS	8	AMERICAN JOURNAL OF HUMAN GE ACTA GENETICAE MEDICAE ET GEMI CLINICAL GENETICS (1) NEUROLOGY (1) NEW ENGLAND JOURNAL OF MEDIC	NETICS (3) ELLOLOGIAE (1) NE (1)	1972 - 1978	- Life Sciences & Biomedicine (8)
			- ADVANCES IN CHEMISTRY SERIES (1)	1965 - 1967	

Scopus Author Results

11 🔲 2008 A compatible exon-exon junction ... Cited by Comparison for Selected Researchers ample of Selected Researcher Result **Total Publications** Web of 110 29 78 1968 ----Collexis - Scor 8 28 39 66 432 39 13 13 17 65 78 78 2000 Chen, Y Chugh, A Eagle, K Good, E Wang, M Cronin, P Hussain, H Nallamothu, Neubig, R Zahuranec, **Total Publication** Comparison for Selected Researche Eagle, K Good, E Wang, M in, P Hussain, H Nallamothu, Neubig, R Zahuranec, D B Chen, Y Chugh, A ------Scopus

h-index Comparison for

Selected Researchers

ISI Web of Science Citation Report



Scopus Citation Overview



Limitations

Search strategies were not intended to be comprehensive or capture a definitive list of each author's research output. Rather, the intent was to utilize search features within each resource for author disambiguation and compare the utility and accuracy of the results.

The different data sources represented in this study each have their own limitations. While none of these limitations represent fatal flaws in the respective products, they are valid concerns or considerations that any end user needs to consider when evaluating the product and its results.

Collexis Research Profiles:

- comprehensive for all UM or UMMS faculty.

ISI Web of Science:

Scopus:

- publications not linked to an author name record.
- message persisted throughout the data collection period.

Conclusion

The citation analysis functions for each of the three tools have unique strengths and weaknesses. Clearly, author name disambiguation continues to pose a problem, particularly in Scopus and ISI Web of Science. UM's instantiation of the Collexis Research Profiles goes a long way to addressing this issue through manual curation, but this is a time- and resource-intensive solution that may not be appropriate for all institutions. Collexis Research Profiles, however, uses only a subset of citations from ISI Web of Science to calculate its h-index and cited by metrics. While all three products provide readily available citation metric data through their reporting features, end users need to be aware of the limitations of coverage and accuracy when using this information.

Acknowledgements

*Funding provided by the Michigan Institute for Clinical Health Research, NIH Grant UL1RR024986.



• UM's current instantiation includes only those authors with a primary appointment to clinical or research tracks at the University of Michigan Medical School (UMMS). While all of the sampled authors met this criteria, the resource is currently not

• UM's current instantiation is manually curated to resolve author name disambiguation issues and validated against faculty curriculum vitae to increase publication count accuracy. However, the h-index and cited-by counts are currently based on ISI Web of Science's MEDLINE subset, which does not represent full MEDLINE.

• There is no way to manually refine a name search within the Research Profiles.

• Cited-by counts are not totaled and need to be manually calculated.

• Citation lists using the 'Distinct Author Sets' tool are markedly smaller than citation lists using author last name and first initial. Many citations do not appear to have been grouped into the provided author sets.

• Some common name/initial sets are impossible to disambiguate using the 'Distinct Author Sets' tool. Full author first names are not searchable, making the disambiguation process for common names even more difficult.

• For name disambiguation issues, a user can select only up to 15 names. This resulted in the exclusion of many individual

• Unclear algorithm for calculating h-index using different database features: in several instances, a different h-index was obtained from the same set of articles due to an unexplained difference in the number of articles used to calculate h-index. A query to Scopus Helpdesk was still open at the time the poster was printed.

• During the period of data collection, the Scopus database presented the following error message: "Scopus is currently experiencing issues with the retrieval and display of some search results. We aim to correct this processing error as soon as possible. Please forgive the inconvenience." Communication with the Scopus Helpdesk indicated that the problem was fixed but the

**Funding provided by the National Center for Integrative Biomedical Informatics, NIH Grant U54DA021519.