International Journal of Information Studies & Libraries
4 (2) 2019, 11-18
http://publishingindia.com/ijisl/

Measuring the Coverage of Library and Information Science (LIS) Academia in Google Scholar and Microsoft Academic: A Scientometric Comparison

Sanjay Kumar Maurya*, Akhandanand Shukla, R. K. Ngurtinkhuma*****

Abstract

The purpose of study is to measure the coverage of LIS academia in Google Scholar and Microsoft Academic using scientometric indicators. The study is based on simple search queries in Google Scholar and Microsoft Academic for each LIS faculty whose information is available on their university website. Duplicate names for LIS faculties have been filtered by adding the extra search string with their names and non-LIS publications have been removed from the study. The study found more coverage for Google Scholar than Microsoft Academic. Google Scholar based "document types" are quite unclear and considered as its' limitations. The study is limited to 26 LIS faculties from five Central Universities of North-East India. The results are limited due to use of simple search in Google Scholar and Microsoft Academic rather than profile search. The study provides insight for the individual researcher and LIS departments to focus on indexing of their research in Google Scholar and Microsoft Academic.

Keywords: Research Performance, Google Scholar, Microsoft Academic, Research Coverage, LIS Research

Introduction

Google Scholar (GS) is a freely accessible web search engine that provides a simple way to broadly search for scholarly literature indexed in the form of full-text or metadata of articles, theses, books, conference proceedings, reports, etc., originated from academic publishers, professional societies, institutional repositories and other websites. GS launched a new service called "Google Scholar Citations" on 16 November 2011, which allows researchers to create an editable, verified (using an institutional email) profile including their personal details, a list of their papers, and citations to those papers (Bar-Ilan et al., 2012). GS has certain advantages over other citation databases; firstly, it covers research output from Social Sciences and Humanities apart from Science-related subjects; secondly, it covers both print and electronic journals, conference proceedings, books, theses, dissertations, preprints, abstracts and technical reports available from major academic publishers, distributors, aggregators, professional societies, government agencies and pre-print/reprint repositories at universities, as well as those available across the web and index citation accordingly (Adriaanse & Rensleigh, 2013; Levine-Clark & Gil, 2008; Schroeder, 2007). Microsoft Academic (MA) is a free search engine and a new service offered by Microsoft since 2012 and was introduced to the bibliometric research community by Harzing (Hug et al., 2017) MA is developing at a fast pace and is on the verge of becoming a bibliometric superpower (Hug & Brändle, 2017). It has broad coverage, structured and rich metadata and citation analysis features as well as the social network for academics. Microsoft Academic Search (MAS) counts citations and provides bibliographic records grouped by authors, journals, institutions or research disciplines, with limited quality control that is being considered for research evaluation and scientific benchmarking (Ortega & Aguillo, 2014). In the electronic age, the development of databases like PubMed, Web of Science (WoS), Scopus and GS on the

* Research Scholar, Department of Library & Information Science, Mizoram University, Aizawl, Mizoram, India. Email: sanjay2015maurya@gmail.com
** Assistant Professor, Department of Library & Information Science, Mizoram University, Aizawl, Mizoram, India. Email: akhandanandshukla@gmail.com Orcid id: http://orcid.org/0000-0001-7157-5600
*** Professor, Department of Library & Information Science, Mizoram University, Aizawl, India. Email: rkn05@rediffmail.com

World Wide Web, offering search facilities for a particular subject, author, institution and the ability to perform citation analysis (Falagas et al., 2008). "Google and Microsoft are rolling out free tools that will enable researchers to analyse citation statistics, visualise research networks and track the hottest research fields" (Butler, 2011). Many organisations have signed with Microsoft in aspects of a journal article from prominent academics and publishers; and this puts it in a direct competition with GS (Carlson, 2006).

Interestingly, MA coverage for single academic is an excellent alternative for citation analysis from each of the four main citation databases available in Publish or Perish (POP) software also. GS and MA show similarity in citation and publication as well as both are freely accessible (Harzing, 2016a; Harzing & Alakangas, 2017). Thus, it will be interesting to find the research performance of authors based on two freely accessible databases, i.e., MA and GS.

Literature Review

Falagas et al. (2008) compared PubMed, Scopus, WoS and GS for citation analysis and found Scopus has 20% more coverage than WoS while GS citation has inconsistent accuracy and is updated infrequently. GS and WoS show the similar result in a citation-based ranking of scholars and journals in the field of Computer Science, but show a weaker correlation in h-index (Barreto et al., 2013; Franceschet, 2010). Jacsó (2011) analysed the content and features of MAS software along with its shortcomings from a bibliometric perspective. Adriaanse & Rensleigh (2013) compared three citation resources ISI WoS, Scopus and GS and found that WoS retrieved more citation results followed by GS and Scopus; Wos and Scopus retrieved no duplicates and less inconsistent than GS. Ortega & Aguillo (2014) analysed 771 personal profiles of different disciplines and found that GS citation includes more documents and citations; also it has less duplicated profiles and more frequently updated rate than MAS. Orduña-Malea et al. (2014) criticised MAS because it has been not updated since 2013 and indexing of new records is going on with a minimal rate without following any apparent pattern. Harzing (2016a) analysed both publication and citation coverage of MA and found that it has less coverage than GS, similar coverage with Scopus but more coverage than WoS. Hug & Brändle (2017) and Hug et al. (2017) concluded that MA offers structured and rich metadata, which facilitates data retrieval, handling and processing easily than GS. Harzing & Alakangas (2017) compared 145 academicians across five disciplines and found similar coverage by both MA and GS in four disciplines. MA has quick search findings and provides more reliable author counts as well as much improved bibliographical details. Thelwall (2018) advocated it as a new data source for scientometrics that allows automated searching, unlike GS. MA shows a more conservative approach when citations are identified by including an only limited set of preprint sources or from academic repositories. The studies discussed above left the research gap to measure the academic performances of LIS academia through MA and GS databases.

Scope and Method

The study is focusing on the research performance of 26 LIS academia (faculties) from five central universities of north-east India. The study covers all types of research publications indexed in GS and MA without any temporal restriction. The comparison of research performances is based on MA and GS databases. The related data for LIS academia (faculties) were retrieved by searching their name, author's profile, affiliation, subject of specialisation for retrieving maximum data. The raw data were extracted from 30 September to 3 October 2018. Further, raw data were filtered by removing duplicate records and non-LIS-related publications. The objective of the study is to analyse the faculty's research performance in GS and MA through different scientometric indicators which will determine the coverage as well as strength and weaknesses of MA and GS also in the present context.

Results

Year-Wise Publication and Citation Coverage

Fig. 1 shows the publication and citation coverage of two databases (GS and MA) for LIS academia. From the observation of Fig. 1, the increasing trend of publications and citations has been observed in both the databases. The temporal coverage for publications and citations has been observed since 1991 in both databases (GS and MA). The coverage of GS publications has been observed more than MA during the period. Similarly, citations coverage for GS is much higher than MA. Statistically, the publication coverage of GS (721 publications) is much higher

than MA (220 publications). MA has 30% coverage of publications in comparison to GS publications, i.e., GS has more than 3.27 times higher publication coverage than MA. The citation coverage of MA (257 citations) is much less as compared to GS (955 citations) and found 26.9% of GS citations, i.e., GS has more than 3.71 times higher citation coverage than MA. The overall Average Citation Per Publication (ACPP) has been calculated for GS (1.32) and MA (1.16) which is higher for GS. The highest number of publication coverage for GS (312) and MA (80) has been observed during 2011-2015 while the highest citation coverage for GS (500) has been observed during 2011-2015 and for MA (118), it has been observed during 2006-2010. The inference has been drawn from the analysis that coverage of GS is much better and MA in terms of publication and citations.

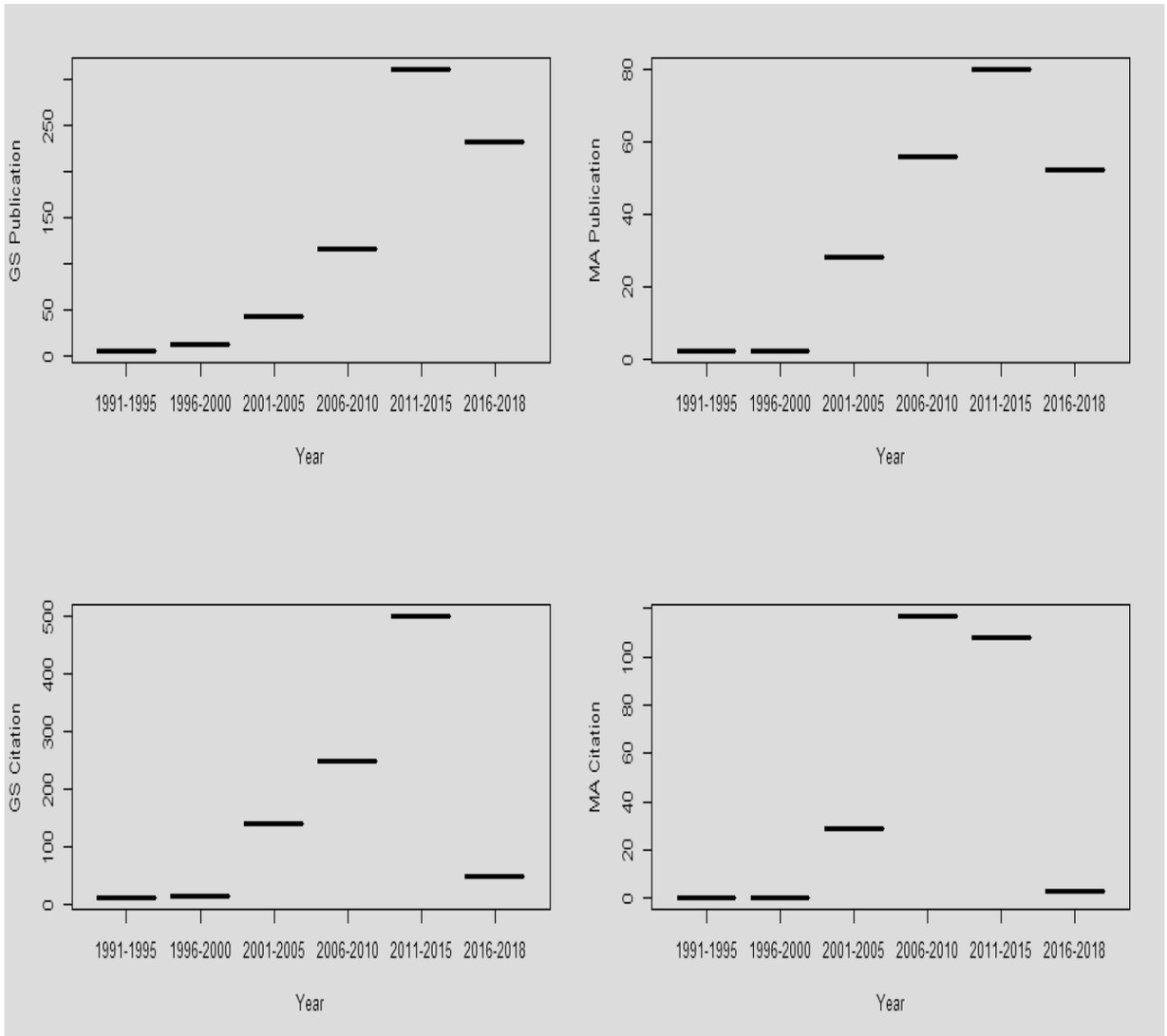

Fig. 1: Year-Wise Publication and Citation Coverage

Prolific Journal's Coverage

Fig. 2 shows the top 10 journal of publication by the LIS academia of north-east India through GS and MA databases. Out of the top 10 LIS journals, two journals, i.e., Asian Journal of Multidisciplinary Studies and Pearl: A Journal of Library and Information Science, have been commonly appeared in both GS and MA databases. Among the top 10 LIS journals, GS has a total of 57 publications with 65 citations while MA has 39 publications with 50 citations. Out of top 10 LIS journals, six LIS journals indexed in GS has received citations while MA has only two LIS journals with citations. Fig. 2 revealed that Library Herald has the highest number of publications (seven) without citations followed by Annals of Library and Information Studies with five publications and 48 citations in MA. In case of GS, Kelpro Bulletin and Library Waves have the highest number of publications (eight each) with 20 and zero citations, respectively; while the journal Library Progress (International) has only four publications with total 21 citations which is the highest citation amongst any top 10 journals of GS.

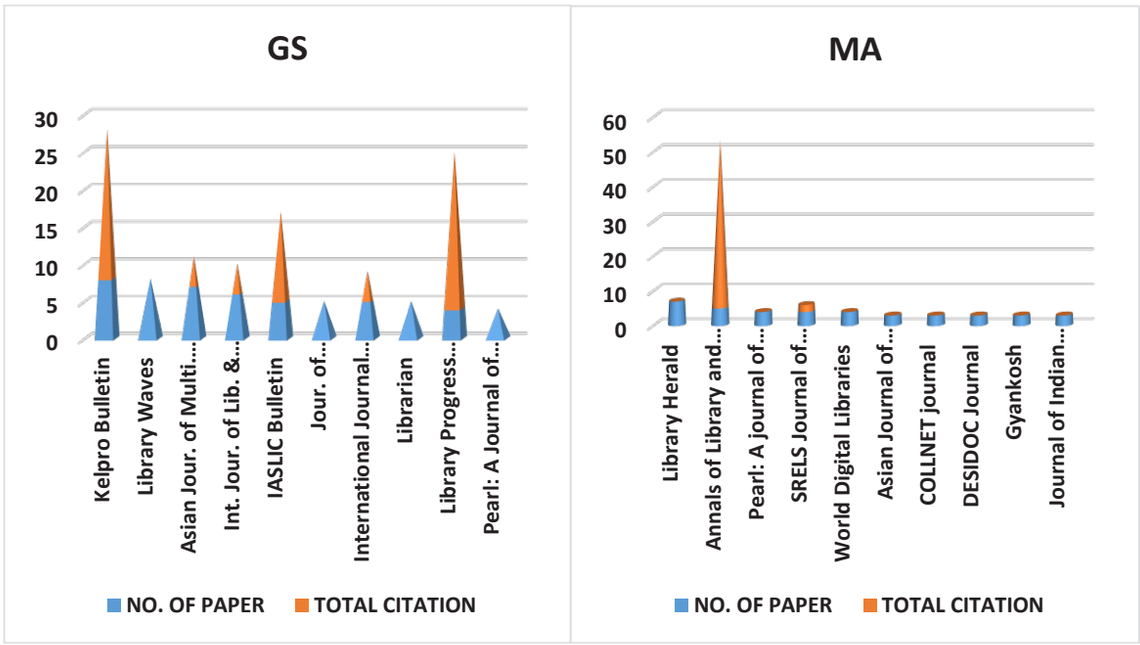

Fig. 2: Top 10 Prolific Journals in GS and MA

Coverage in Forms of Publication

Forms of publication are one of the major attributes to determine the coverage pattern of a database. Generally, forms of publication include book, book chapters, conference proceedings, journal articles, reports, technical reports, etc. From Fig. 3, it has been observed that GS has covered mainly three different forms of publication with the highest percentage of conference proceedings (13.45%) followed by book chapters (8.73%) and books (1.8%). MA covers mainly journal paper as the main forms of publication which covers 25% of total publications while 75% of publications were not specified by its forms and displayed as blank.

Similarly, 67.26% "blank" is observed in GS also, while 8.73% observed as Other (includes PDF, Citation, Doc.) which is not properly ascertained as to which category it belongs. The percent error is calculated on the basis of blank & Other publication forms and it is found that both GS and MA have approximately 75% error in coverage in forms of publication. GS has more forms of publication than MA. In case of GS, Mayr & Walter (2007) study found that the high ratio of journals found was reflected as Citations (28%) document types followed by full-text in PDF (19%) while other forms of documents were negligible in ratios.

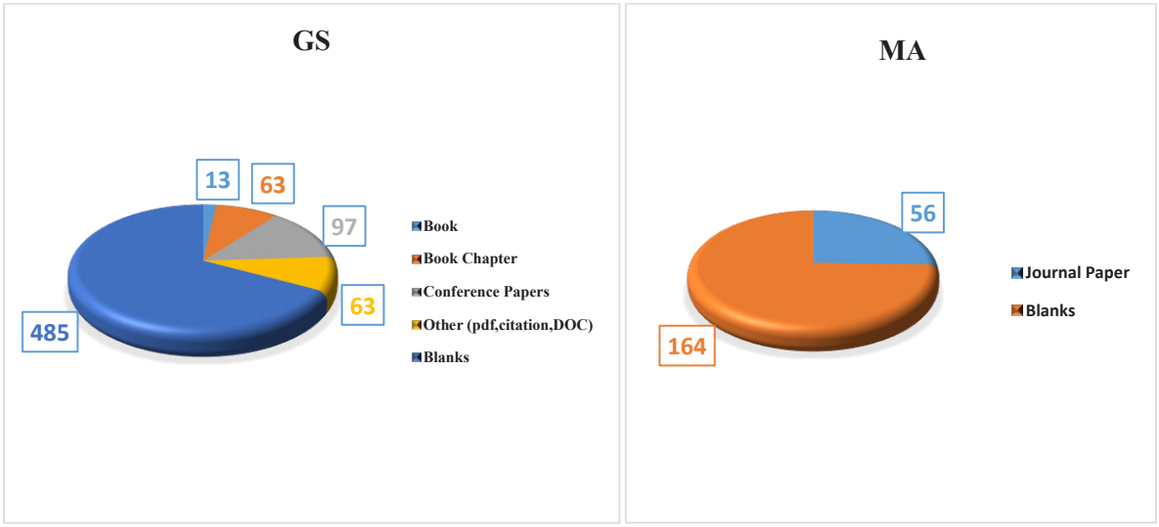

Fig. 3: Forms of Publication in GS and MA

The h-index and g-index Coverage

The *h*-index measures the quality as well as quantity of research publication and used as an indicator for scientometric study to evaluate the works of researcher's. Further, this indicator becomes more effective if applied to the researchers that belongs to the same disciplines (Baldock et al., 2009). The *g*-index highlights the shortcoming of *h*-index and rewards one or several cited publications of researchers' and thus helps to determine the impact of a researcher. The higher values of *g*-index and *h*-index make it easier to differentiate the performance of the researchers (Huang & Chi, 2010). In Fig. 4, the *h*-index and *g*-index of LIS academia (faculties) have been analysed through GS and MA. The *h*-index and *g*-index value for two faculties are found nil in GS while no values observed for six faculties in MA. GS shows the highest *h*-index for M. K. Sinha (10) followed by M. K. Verma (5) and A. Shukla (4) while the highest *g*-index for M. K. Sinha (13) followed by S. Ravikumar (8) and S. N. Singh (8). MA shows the highest *h*-index for M. K. Sinha (5) followed by T. P. Devi (2) and P. Hanging (2) while the highest *g*-index for M. K. Sinha (7) followed by T. P. Devi (6), R. N. Mishra (3), P. Rath (3) and M. Devi (3). It has been observed that *h*-index and *g*-index values vary for LIS faculties in GS and MA. From the analysis, it has been observed that GS is much better performed than MA in terms of *h*-index and *g*-index coverage.

Discussion and Conclusions

The bibliographic databases provide facilities to extract the academic research output of faculties and researchers at one place. There is the number of data extraction, and visualisation tools are available through which we can analyse and interpret the result of research. By the results of numbers of research, it is well understood that no bibliographic database is complete in terms of indexing of bibliographic records as well as retrieval of those indexed records. Data extraction tools have some limitations also, which limits the search result, but there is no other way to measure the research performance of academia of any field. The results based on such databases, data extraction tools and data visualisation tools are just a representation of data indexed with certain limitations.

GS and MA databases are freely available databases, indexing the academic research output of academic disciplines from various online sources at their own level. GS is inviting the researchers to make GS Profile for improved indexing of such records; and automatic indexing, as well as manual entry of bibliographic records, is facilitated by such databases. The complete coverage of such records is not possible due to the dependency of the publishing medium. If any research is published online, then chances for indexing are increased while there's a reduced possibility in case of the print medium. The coverage in terms of number of publications

and citations in such databases suffers a lot of technical, human and promotional issues. The "technical issues" relate to the document types covered in the database, algorithm for indexing of records in database, and manual entry support in the database. The "human issues" relate to the availability of number of such databases, quality issues of such databases, lack of personal interest of researcher, lack of time to maintain the online account, feeling of encroachment in personal and professional life, creates comparison among peers, authenticity and usefulness of such databases for career promotion, etc. The "promotional issues" relate to the public awareness programme run by the database creator or sponsor(s) among the researchers. What benefit the researcher will get by using such databases? How simple data entry form is available with the database? And from among many such databases, why to choose some particular [GS or MA or ResearchGate] database? After creating an online account with the database, how many numbers of unique facilities researcher will get? And finally, ease of use is also a concern for such databases.

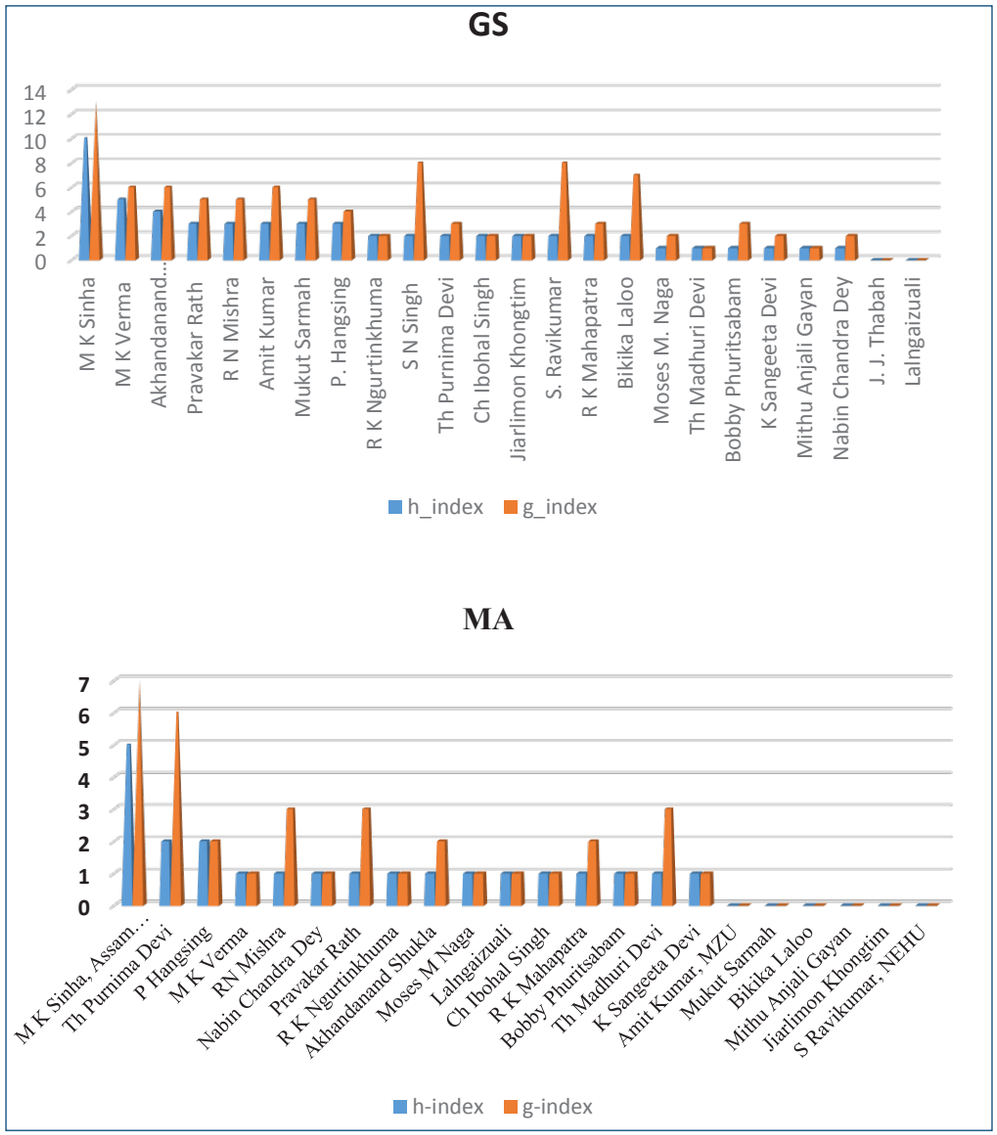

Fig. 4: The *h*-index and *g*-index Coverage of LIS Academia

Study observed that publication and citation coverage of GS is much higher than MA. In terms of journal's coverage, GS has indexed number of journals than MA; the number of publication and citations coverage of journals are also better than MA. Forms of publications covered by GS are more than MA as well as numbers of publications under forms of publications are much higher for GS than MA; though GS has a problem in identifying

the forms of publications and categorisation of them which MA has not. The *h*-index and *g*-index coverage of GS have been observed higher than MA. From the study, it is evidenced that coverage of GS is better than MA for LIS discipline based on certain scientometric indicators. Though a similar result cannot be obtained in other academic disciplines, i.e., coverage may be discipline-specific. MA is an emerging database started after GS and is continuously improving in terms of coverage (Harzing, 2016a; Harzing, 2016b; Harzing & Alakangas, 2017; Hug & Brändle, 2017). The study suggests researchers to be registered with GS since coverage is much better than MA, and if possible also join MA since coverage is improving.

References

Adriaanse, L. S., & Rensleigh, C. (2013). Web of Science, Scopus and Google Scholar: A content comprehensiveness comparison. *Electronic Library*, *31*(6), 727-744. Retrieved from https://doi.org/10.1108/EL-12-2011-0174

Baldock, C., Ma, R., & Orton, C. G. (2009). The h index is the best measure of a scientist's research productivity: Point/Counterpoint. *Medical Physics*, *36*(4), 1043-1045. Retrieved from https://doi.org/10.1118/1.3089421

Bar-Ilan, J., Haustein, S., Peters, I., Priem, J., Shema, H., & Terliesner, J. (2012). *Beyond citations: Scholars' visibility on the social web* (p. 14). Presented at the 17th International Conference on Science and Technology Indicators, Montreal, Canada. Retrieved October 10, 2018 from https://arxiv.org/abs/1205.5611

Barreto, M. L., Aragao, E., Sousa, L. E. P. F., Santana, T. M., Barata, R. B., Barreto, M. L., … Barata, R. B. (2013). Differences between h-index measures from different bibliographic sources and search engines. *Revista de Saúde Pública*, *47*(2), 231-238. Retrieved from https://doi.org/10.1590/S0034-89102013000100008

Butler, D. (2011). Computing giants launch free science metrics. *Nature*, *476*(7358), 18-18. Retrieved from https://doi.org/10.1038/476018a

Carlson, S. (2006, April). Challenging Google, Microsoft unveils a search tool for scholarly articles. *Chronicle of Higher Education*, *52*(33). Retrieved October 10, 2018 from https://www.chronicle.com/article/Challenging-Google-Microsoft/17846

Falagas, M. E., Pitsouni, E. I., Malietzis, G. A., & Pappas, G. (2008). Comparison of PubMed, Scopus, Web of Science, and Google Scholar: Strengths and weaknesses. *The FASEB Journal*, *22*(2), 338-342. Retrieved from https://doi.org/10.1096/fj.07-9492LSF

Franceschet, M. (2010). A comparison of bibliometric indicators for computer science scholars and journals on Web of Science and Google Scholar. *Scientometrics*, *83*(1), 243-258. Retrieved from https://doi.org/10.1007/s11192-009-0021-2

Harzing, A. W. (2016a). Microsoft Academic versus Google Scholar, Scopus, and Web of Science: Anne-Wil Harzing's case. Retrieved April 28, 2018 from http://googlescholardigest.blogspot.com/2016/06/microsoft-academic-versus-google.html

Harzing, A. W. (2016b). Microsoft Academic (Search): A Phoenix arisen from the ashes?. *Scientometrics*, *108*(3), 1637-1647. Retrieved from https://doi.org/10.1007/s11192-016-2026-y

Harzing, A. W., & Alakangas, S. (2017). Microsoft Academic is one year old: The phoenix is ready to leave the nest. *Scientometrics*, *112*(3), 1887-1894. Retrieved from https://doi.org/10.1007/s11192-017-2454-3

Huang, M., & Chi, P. (2010). A comparative analysis of the application of h-index, g-index, and a-index in institutional-level research evaluation. *Journal of Library and Information Studies*, *8*(2), 1-10.

Hug, S. E., & Brändle. (2017, June 19). Microsoft Academic is on the verge of becoming a bibliometric superpower. Retrieved May 1, 2018 from http://blogs.lse.ac.uk/impactofsocialsciences/2017/06/19/microsoft-academic-is-on-the-verge-of-becoming-a-bibliometric-superpower/

Hug, S. E., Ochsner, M., & Brändle, M. P. (2017). Citation analysis with Microsoft Academic. *Scientometrics*, *111*(1), 371-378. Retrieved from https://doi.org/10.1007/s11192-017-2247-8

Jacsó, P. (2011). The pros and cons of Microsoft Academic search from a bibliometric perspective. *Online Information Review*, *35*(6), 983-997. Retrieved from https://doi.org/10.1108/14684521111210788

Levine-Clark, M., & Gil, E. L. (2008). A comparative citation analysis of Web of Science, Scopus, and Google Scholar. *Journal of Business & Finance Librarianship*, *14*(1), 32-46. Retrieved from https://doi.org/10.1080/08963560802176348

Mayr, P., & Walter, A.-K. (2007). An exploratory study of Google Scholar. *Online Information Review, 31*(6), 814-830. doi: https://doi.org/10.1108/14684520710841784

Orduña-Malea, E., Martín-Martín, A., Ayllon, J. M., & Delgado López-Cózar, E. (2014). The silent fading of an academic search engine: The case of Microsoft academic search. *Online Information Review*, *38*(7), 936-953. Retrieved from https://doi.org/10.1108/OIR-07-2014-0169

Ortega, J. L., & Aguillo, I. F. (2014). Microsoft Academic search and Google Scholar Citations: Comparative analysis of author profiles. *Journal of the Association for Information Science and Technology*, *65*(6), 1149-1156. Retrieved from https://doi.org/10.1002/asi.23036

Schroeder, R. (2007). Pointing users toward citation searching: Using Google Scholar and Web of Science. *Portal: Libraries and the Academy*, *7*(2), 243-248. Retrieved from https://doi.org/10.1353/pla.2007.0022

Thelwall, M. (2018). Does Microsoft Academic find early citations?. *Scientometrics*, *114*(1), 325-334. Retrieved from https://doi.org/10.1007/s11192-017-2558-9