

Bibliometric Analysis of Research Publications in Domain of Computer Science : Case Study of Pakistani Authors

FIRST A. HAMZA MUNIR¹, SECOND B. AGHA MUHAMMAD ANAS², AND THIRD C. MEHMOOD ALI.³

¹Information Technology University, Lahore 54000 (e-mail: msds20061@itu.edu.pk)

²Information Technology University, Lahore 54000 (e-mail: msds20068@itu.edu.pk)

³Information Technology University, Lahore 54000 (e-mail: msds20090@itu.edu.pk)

Corresponding author: First Hamza Munir (e-mail: msds20061@itu.edu.pk).

⋮ **ABSTRACT** Abstract will be written later

⋮ **INDEX TERMS** Bibliometric study, Computer Science, Pakistan, Author Patterns, Citations

I. LITERATURE REVIEW

IN this section we take a look at the previous work done for bibliometric analysis in different fields of science. The purpose is to take a general look at the different approaches adopted by different researchers to present their work in an efficient way possible.

Iqbal, W., Qadir, J., Tyson, G. et al [1] analyses the computer networking domain and gives a study which uses the article content and metadata of four important computer networking periodicals —IEEE Communications Surveys and Tutorials (COMST), IEEE/ACM Transactions on Networking (TON), ACM Special Interest Group on Data Communications (SIGCOMM), and IEEE International Conference on Computer Communications (INFOCOM)—obtained using ACM, IEEE Xplore, Scopus and CrossRef, for an 18-year period (2000–2017) to address important bibliometric questions. In this study, they aim to track the co-evolution of trends in the COMST and TON journals and compare them to the publication trends in INFOCOM and SIGCOMM. Their analyses of the computer networking literature include: (a) metadata analysis; (b) content-based analysis; and (c) citation analysis. In addition, they identified the significant trends and the most influential authors, institutes and countries, based on the publication count as well as article citations.

Ali, Muhammad Richardson, Joanna [2] performed the bibliometric analysis of the research published by the LIS (Library and Information Sciences) scholars in Pakistan through a survey. A questionnaire was designed by the authors and distributed to the LIS scholars through Email, Yahoo groups and Facebook to the representatives from all

the provinces of Pakistan. The results obtained from the total of 104 respondents were then analyzed in the tool SPSS version 21. They identified the relevant demographic information, e.g. gender and geographical, the extent of collaborative author-ship, the extent of publishing based on geographical regions, the strength of association between job title (seniority) and number of publications, the strength of citation metrics for national outputs, identify factors which may impact negatively on LIS scholar's ability to undertake research and/or to publish it.

R. S. Bajwa and K. Yaldrum [3] presents an analysis of the research trends in Pakistan in the field of biotechnology for the period 1980–2011. They analyzed the increase and decrease of Annual growth rate, comparison of organization who actively participate in research in biotechnology using Publication rate, citation rate average citation per paper and multiple indexing methods.

Garousi and Fernandes [4] computed and classified the top 100 papers in the field of software engineering. In their study they found the top papers based on the citation count and the average annual number of citations. They devised a GQM (Goal, Question, and Metric) methodology to find their goal by devising some questions in order to achieve it. They have also compared the top papers with the top papers of all areas of science. They also identified the fields to which the top cited papers are related. Further they also identified the venues in which the top papers are presented.

Muhammad Kamran, Hikmat Ullah Khan [5] in research presents a bibliometric analysis of articles in Blockchain in Internet of Things (IoT) domain, covering papers published

in top journals and conferences, and finds research trends. It also explores diverse research areas, the most influential publications, top publication venues, top funding agencies and future research direction. The research we are going to conduct is very similar to this but our area of interest is computer science including all domains.

Missen, M.M.S., Qureshi, S., Salamat [6] and other presents Pakistan case study to analyze the impact on International research from the timespan 2009 to 2018. The study examines 2000 published articles belonging to 50 research scholars of different disciplines. This analysis is conducted on three different levels: researcher level, field level and domain level. In this paper they discuss readability scores, title formats, single and multiple authorships of articles, citation rates, publication rates over time, the research contribution of both genders and the impact of PhD institution of authors in research publications.

Fiala, Dalibor Tutoky, Gabriel [7] presents a bibliometric study of 1.9 million computer science papers published from 1945 to 2014 and indexed in Web of Science. They analyze both the quantity and the impact of these publications according to document types, languages, disciplines, countries, institutions, and publication sources. The most frequent author keywords, cited references, and cited papers as well as the distribution of the number of references and citations per paper and of the age of cited references are also explored. Since conference proceedings play a tremendous role in this scientific field, they investigate the time and place of computer science conferences in terms of the most prolific months and locations. And, last but not least, the production of journal articles and conference papers over the whole time period and the level of collaboration in different computer science disciplines are inspected. One of the main results is the finding that “Artificial Intelligence” is the most productive subfield of computer science, “Neural Networks” [10] seems to be losing its charm but “Interdisciplinary Applications” has the highest relative impact.

In another research Bakri, A., Willett, P. [8] analyses the computer science publications in Malaysia. The purpose of this paper is to analyze the publications of, and the citations to, the current staff of 19 departments of computer science in Malaysian universities, and to compare these bibliometric data with expert peer reviews of Malaysian research performance. This study searches citation of the Scopus and Web of Science databases. The conclusion draw from research is both publication and citation rates are low, although this is at least in part due to some Malaysian universities having only a teaching function. More of the departments’ publications were identified in Scopus than in Web of Science, but both databases were needed for comprehensive coverage. Statistically significant relationships were observed between the departments’ publication and citation counts and the rankings of the departments’ parent universities in two evaluations of the research performance of Malaysian universities.

Ding et al. [9] performed the syntactic and semantic analysis of the citations in publications related to computer

science. The syntactic part involves the identification of the location where the citations are found (i.e. in which section of the article). The semantic analysis finds the motivation of the citations through manual approach of predefined categorizations or semiautomatic approach of NLP. In predefined categorization the citations are divided into categories defined by words or phrases in a decision tree.

II. DATA COLLECTION

To perform the analysis, data is collected from Web of Knowledge also known as Web of Science, Clarivate Analytics[11]. Web of Science is a website that provides subscription-based access to multiple databases that provide comprehensive citation data for many different academic disciplines. It was originally produced by the Institute for Scientific Information and is currently maintained by Clarivate Analytics.

Using Web of Knowledge source, from advanced search, data is queried using command in TABLE 1.

Query
CU=(PAKISTAN) AND SU=(Computer Science) Indexes = SCI-EXPANDED, SSCI, A&HCI, ESCI Timespan = (All years)

TABLE 1. Query to get data from Web of Knowledge

The the abbreviated terms in table 1 represents CU as Country/Region, SU as Research Area, SCI_EXPANDED as Science Citation Index Expanded, SSCI as Social Science Citation Index, AHCI as Arts Humanities Citation Index, ESCI as Emerging Sources Citation Index.

The query mentioned above returned a total of 8,807 publications with majority publications in English with a few exceptions in other languages. The results were extracted and stored in the form of Text(.txt) format file. The data was downloaded on 18th December 2020. The data-set contains 8807 rows 1 for each publication and 50 columns representing different information about publications.

REFERENCES

- [1] Iqbal, W., Qadir, J., Tyson, G. et al. A bibliometric analysis of publications in computer networking research. *Scientometrics* 119, 1121–1155 (2019).
- [2] Ali, Muhammad Richardson, Joanna. (2016). Research Publishing by Library and Information Science Scholars in Pakistan: A Bibliometric Analysis. *Journal of Information Science Theory and Practice*. 4. 6-20. 10.1633/JISTaP.2016.4.1.1.
- [3] Bajwa, R.S., Yaldrum, K. Bibliometric analysis of biotechnology research in Pakistan. *Scientometrics* 95, 529–540 (2013). <https://doi.org/10.1007/s11192-012-0839-x>.
- [4] V. Garousi and J. M. Fernandes, "Highly-cited papers in software engineering: The top100," *Information and Software Technology*, vol. 71, pp. 108-128, 2016.
- [5] Muhammad Kamran, Hikmat Ullah Khan, Wasif Nisar, Muhammad Farooq, Saeed-Ur Rehman, Blockchain and Internet of Things: A bibliometric study, *Computers Electrical Engineering*, Volume 81, 2020, 106525, ISSN 0045-7906, <https://doi.org/10.1016/j.compeleceng.2019.106525>.
- [6] Missen, M.M.S., Qureshi, S., Salamat, N. et al. Scientometric analysis of social science and science disciplines in a developing nation: a case study of Pakistan in the last decade. *Scientometrics* 123, 113–142 (2020). <https://doi.org/10.1007/s11192-020-03379-8>.

- [7] Fiala, Dalibor Tutoky, Gabriel. (2017). Computer Science Papers in Web of Science: A Bibliometric Analysis. Publications. 5. 10.3390/publications5040023.
- [8] Bakri, A., Willett, P. (2011). Computer science research in Malaysia: a bibliometric analysis. Aslib Proceedings, 63(2/3), 321–335. doi:10.1108/00012531111135727.
- [9] Y. Ding, G. Zhang, T. Chambers, M. Song, X. Wang, and C. Zhai, "Content-based citation analysis: The next generation of citation analysis," Journal of the Association for Information Science and Technology, vol. 65, pp. 1820-1833, 2014.
- [10] J. E. Dayhoff, Neural network architectures: an introduction: Van Nostrand Reinhold New York, 1990.
- [11] www.webofknowledge.com

...