Scientometric Analysis of Literature on Big Data Research Output in India Based on Scopus Database

Mr. Akhilesh Kumar Varma
Junior Research Fellow
Department of Library & Information Science
Banaras Hindu University, Varanasi-221005

Dr. Kunwar Singh
Assistant Professor
Department of Library & Information Science
Banaras Hindu University, Varanasi-221005
e-mail: singhdlisbhu2015@gmail.com

Abstract *- The present study has been undertaken with the purpose of finding out the growth and characteristics of big data research output in India. Over 1265 articles for the period 2012-2016 were collected from the Scopus database and were analyzed to study year wise distribution, form wise distribution, subject wise distribution, most prolific authors, subject wise, top ranked Institute and ranking of periodicals etc.*

Keywords: *Big data, Literature, Scientometric, Data Analysis*

Introduction

Recent advancements in the field of instrumentation, adoption of some of the latest internet technologies and applications, and the declining cost of storing large volumes of data, have enabled researchers and organizations to gather increasingly large and heterogeneous datasets. Due to their enormous size, heterogeneity and high speed of collection such large datasets are often referred to as 'big data' (Arora, 2016, p.1).

Literature Review

Vijayalakshmi and Ambuja (2013) studied and explained about the growth of literature, the degree of collaboration, relative growth rate and doubling time of "Remote Sensing literature in Scopus database: A bibliometric analysis. Gupta, Kumbar and Tiwari (2014) revealed the research performance of 25 leading universities research output, citations and International collaborative papers on Ranking of Indian Universities in Social Sciences using bibliometric indicators during 2008-12. Raju (2014) carried a research study on ranking and scattering of journals in Physics: A quantitative study. Study has highlighted about the international journal of information dissemination and technology papers in various aspects bibliographic form-wise distribution, ranking and scattering of journals, the productivity, the obsolescence of literature and half-life of journals citations. Zell, H. et al., (2010) analyzed air pollution researches during the period from 1955 to 2006, 26,253 items were listed through web of science database. The study identified 124 countries in 24 different languages. Further they analyzed citation levels, most productive countries; highest number of author's publications and his citation range, most productive subject area and journals are analyzed. Dobrot, Marina et al., (2013) examined research Europe Union member countries air pollution per inhabitant, using a novel statistical approach I-distance method. The study measured the air pollution per inhabitant and evaluating the measurement by ranking that the worst situation occurs in Luxembourg, Bulgaria, Ireland, Estonia, and Greece, while situation in Sweden,

Portugal, Germany, Slovakia, and United Kingdom is much better with far less air pollution per inhabitant. The study also explains the results of ranking and abilities of specific countries to scope with the environmental problems such as air pollution.

Objectives of The Study

The present study aims at identify the promising importance of literature published in the field of 'Big Data' during 2012-2016 with a view to identify the main objectives of present study are:

- To identify the year wise distribution of articles;
- To identify the documents type of contributions;
- To identify the most prolific authors;
- To identify the top ranked institute and their contribution in this field;
- To identify the subjective wise distribution of contributions;
- To identify the top ranked periodicals.

Methodology

The present study has been undertaken with the purpose of finding out the literature published on 'Big data' in India, the Scopus database was found to be the most inclusive and suitable source of literature for the present study. Scopus is the largest bibliographic database containing abstracts and citations for academic journal articles, books and conference proceedings. It covers nearly 22,000 titles from over 5,000 publishers, of which 20,000 are peer-reviewed journals in the scientific, technical, medical, and social sciences. It is owned by Elsevier and is available online by subscription. The term 'Big Data' was used to search the articles indexed in Scopus for retrieving the results. There were a total of 1267 documents indexed in Scopus from India. The data were filtered and finally exported to MS-Excel where the tabulation and simple statistical methods was applied. All articles referring to 'Big Data' research output from India during 2012-2016 were assessed by analysing various aspect: year wise distribution, form wise distribution, subject wise distribution, most prolific authors, subject wise, top ranked Institute and ranking of periodicals etc. Finally the analyzed data were arranged and presented to fulfil the objectives of this study.

Data Analysis

Year Wise Distribution

It has been observed from the Table.1 that 1265 articles were published during 2012-2016. Further, it shows that 713(56.27%) maximum number of articles published in the year 2016 and minimum number of article 2(0.16%) in the year 2012. It revealed that from 2013 to 2016 articles publication rate have been increased.

Table 1: Year Wise Distribution

Year	No of Articles	%
2012	2	0.16
2013	26	2.05
2014	121	9.55
2015	405	31.97
2016	713	56.27
Total	1267	100.00

Form Wise Distribution

The table 2 revealed that the literature on the 'Big Data' is being published in different forms. It shows that 989(78.06%) articles were published in the conference followed by 255(20.13%) Journal articles, 15(1.18%) book chapter and 8(0.63%) review paper. It is evident from the analysis of data that the conference papers are the most widely used form of documents in which most of the literature on the subject 'big data' has been published.

Table 2: Form Wise Distribution

Types	No of items	%
Article	255	20.1
Conference Papers	989	78.1
Book Chapter	15	1.18
Review	8	0.63
Total	1267	100

Most Prolific Authors

Table 3 shows that the most prolific authors, who have contributed a large number of papers in the field of big data. The authors are ranked according to the number of publications. The below table 3 shows that the most productive authors in the field of big data is Johari, R and Vijayakumar, V, who have contributed 7 papers in the field and occupied first rank in the list. Dubey, S. K., Kumar, M., Pandey, R., and Singh, S. N. have contributed 6 paper each and occupied second rank in the list. Whereas, Agrawal, S., Arockiam, N. T., Rath, S. K., Singhal, R., Subrahmanyam, K., Syed Ibrahim, S. P. and Vijayakumar, V have contributed 5 papers each and scored third rank in the list. It is clear from the study that Johari, R and Vijayakumar, V, have contributed 7 papers and dominated the rank list of prolific authors.

Table 3: Most Prolific Authors

Author	No. of Contribution	Rank
Johari, R.	7	1
Vijayakumar, V.	7	1
Dubey, S.K.	6	2
Kumar, M.	6	2
Pandey, R.	6	2
Singh, S. N.	6	2
Agrawal, S.	5	3
Arockiam, N. T.	5	3
Rath, S. K.	5	3
Singhal, R.	5	3
Subrahmanyam, K.	5	3
Syed Ibrahim, S. P.	5	3
Vijayakumar, V.	5	3

Ranking of core periodicals

Table 4 shows the ranking of core periodical in the 'big data', first rank occupied by the journal titled Procedia Computer Science which published 133 articles of total records and next two positions are occupied by the journal ACM International Conference Proceeding Series with share of 60 and International Journal of Applied Engineering Research 52 articles respectively. The ranking list may be useful for the library and information professionals for taking the policy decisions regarding the subscription list of periodicals on the subject of 'big data'. It will also be helpful for documentation office while preparing a comprehensive documentation list of core journals carrying the highest percentage of items.

Table 4: Ranking of core periodicals

Source	No.	Rank
Procedia Computer Science	133	1
ACM International Conference Proceeding Series	60	2
International Journal of Applied Engineering Research	52	3
Smart Innovation Systems And Technologies	51	4
Lecture Notes In Computer Science Including Subseries; Lecture Notes In Artificial Intelligence; And Lecture Notes In Bioinformatics	50	5
Advances In Intelligent Systems And Computing	44	6
International Journal Of Control Theory And Applications	34	7
Indian Journal Of Science And Technology	24	8
International Journal Of Pharmacy And Technology	20	9
Arpn Journal Of Engineering And Applied Sciences	15	10
Communications In Computer And Information Science	6	11
Research Journal Of Pharmaceutical Biological And Chemical Sciences	6	11
Fujitsu Scientific And Technical Journal	5	12
International Journal Of Electrical And Computer Engineering	5	12
Journal Of Theoretical And Applied Information Technology	5	12
Physics Procedia	4	13
International Journal Of Communication Networks And Distributed Systems	3	14
Asian Journal Of Information Technology	2	15
Cluster Computing	2	15
Cybernetics And Information Technologies	2	15
IT Professional	2	15
Iioab Journal	2	15
Indonesian Journal Of Electrical Engineering And Computer Science	2	15
Information Sciences	2	15
International Journal Of Database Theory And Application	2	15
Journal Of Big Data	2	15
Journal Of Chemical And Pharmaceutical Sciences	2	15
Proceedings Of The ACM SIGMOD International Conference On Management Of Data	2	15
Proceedings Of The ACM Symposium On Applied Computing	2	15
Sadhana Academy Proceedings In Engineering Sciences	2	15

Subject Wise Distribution

Table 7 revealed that the highest percentage of documents i.e. 1075(84.85%) of the total 1265 records is published in Computer Science followed by 292(23.05%) in Engineering, 242(19.10%) in Mathematics and 229(18.07%) in Decision Sciences. It is clear from the study that Computer Science 1075(84.85%) dominated the subject wise distribution of big data research.

Table 5: Subject Wise Distribution

Subject	No. of Contribution	%
Computer Science	1075	84.9
Engineering	292	23.1
Mathematics	242	19.1
Decision Sciences	229	18.1
Social Sciences	65	5.13
Medicine	42	3.31
Energy	35	2.76
Multidisciplinary	29	2.29
Pharmacology	29	2.29
Business	28	2.21
Physics and Astronomy	27	2.13
Biochemistry	17	1.34
Materials Science	16	1.26
Chemical Engineering	8	0.63
Earth and Planetary Sciences	5	0.39
Economics	4	0.32
Environmental Science	4	0.32
Health Professions	3	0.24
Immunology and Microbiology	3	0.24
Neuroscience	3	0.24
Total	1267	100

Top Ranked Institute

It is observed from the Top ranked Institute that Amity University has contributed maximum number of 99(21.85%) articles followed by Thapar University has contributed minimum number of 12(2.65%) article on big data. It revealed that Amity University has dominated the Top ranked Institute list by contributing 99(21.85%) articles alone.

Table 8: Top Ranked Institute

Name of the Institute	No. of Contributions	%
Amity University	99	21.9
Vellore Institute of Technology	63	13.9
VIT University	45	9.93
Amity School of Engineering & Technology	33	7.28
Anna University	30	6.62
Sathyabama University	29	6.4
K L University	22	4.86
Amrita Vishwa Vidyapeetham University	19	4.19

Symbiosis International University	18	3.97
Bharathiar University	16	3.53
SRM University	14	3.09
Thapar University	12	2.65
Indian Institute of Science	12	2.65
Guru Gobind Singh Indraprastha University	11	2.43
National Institute of Technology Karnataka	10	2.21
Indian Institute of Technology Roorkee	10	2.21
Lovely Professional University	10	2.21
Total	453	100

Findings

The following findings of the study:

- It revealed that from 2013 to 2016 articles publication rate have been increased.
- It is evident from the analysis of data that the conference papers are the most widely used form of documents in which most of the literature on the subject 'big data' has been published.
- It is clear from the study that Johari, R and Vijayakumar, V, have contributed 7 papers and dominated the rank list of prolific authors.
- The first rank occupied by the journal 'Procedia Computer Science' which published 133 articles.
- It is clear from the study that Computer Science 1075(84.85%) dominated the subject wise distribution of big data research.
- It revealed that Amity University has dominated the Top ranked Institute list by contributing 99(21.85%) articles alone.

Conclusion

The scientometric study on the 'Big Data' research output in India over a period of five years 2012-2016 was selected to find out the publication patterns in the field. The study was conducted on the data collected from Scopus database during 2007-2013 on big data. This research study was conducted by applying scientometric tools and techniques. After the collection of data from Scopus database, it was analyzed and results were presented in the form of tables for better understanding.

References

1. Vijayalakshmi, S. and Ambuja, R. (2013). Remote Sensing Literature in Scopus Database: a Bibliometric Analysis. *International Journal of Library and information Studies*, 3 (1), p. 70-85.

2. Gupta, Ritu., Kumbar, B. D. and Tiwari, Rishi (2014). Ranking of Indian Universities in Social sciences using bibliometric indicators during 2008-2012. *DESIDOC Journal of Library and Information Technology*, 34 (3), p.197-205.

3. Raju, N. G. (2014). Ranking and Scattering of Journals in Physics: A Quantitative Study. *International Journal of Information Dissemination and Technology*, 4 (1), p. 8-20.

4. Zell, H. et al. (2010). Air Pollution Research: Visualization of Research Activity Using Density-Equalizing Mapping and Scientometric Benchmarking Procedures. *Journal of Occupational Medicine and Toxicology*, 5 (5), p. 1-9.

5. Dobrot, Marina et al. (2013). Measuring and Evaluating Air Pollution per Inhabitant: A Statistical Approach. *APCBEE Procedia*, (5) 33-37.

6. Arora, Ritu (2016). An Introduction to Big Data, High Performance Computing, High-Throughput Computing, and Hadoop. In Ritu Arora (ed.). *Conquering Big Data with High Performance Computing (p. 1-144)*, Switzerland: Springer International Publishing.