BIG DATA Output in Scopus during 2012 to 2016: A Bibliometric Analysis

Chaitanya Veer * Dr. D. K. Veer Gajanan P.Khiste *****

*** Librarian**
St. Gonsalo Garcia College of Arts & Commerce,
Vasai, Dist. Palghar, Mumbai, Maharashtra, India

**** Director**
Knowledge Resource Centre
Dr. Babasaheb Ambedkar Marathwada University, Aurangabad, Maharashtra, India

***** Information Scientist**
Knowledge Resource Center
Dr. Babasaheb Ambedkar Marathwada University, Aurangabad, Maharashtra, India

QR Code

Abstract: - *The present study discusses the "Big Data" as reflected in Scopus for the period from 2012–2016 and investigates the highly productive authors, document types and h-index. The study also aims to find out the top contributing Indian institutions, the preferred sources for publications by Geographical distribution by country, Subject area, Source Type, Affiliation, and Language etc. The result indicates that there were total 9191 documents with 54129 citations on Big Data during 2012 to 2016. The result shows that China and USA are the most active countries in the area of Big Data Analytics. Study shows publication trends in the subjects of computer science and engineering.*

Keywords: Scopus, Big Data, Data Analytics, Data Analytics

1. Introduction:

There are billions of sources of information available in different types of formats at different locations on the web and handheld devices, social media and other applications. Huge amount of information is available in various sectors like banking, corporate companies and in the online form, etc. It is difficult to manage information in such large amounts therefore new techniques came in existence. In the 21st century, many researchers and scientists conducted research on big data analysis. A variety of research was published with different methodology and techniques on 'big data'. Therefore, the present

paper has analysed research done on the concept of "Big Data". It is very a new and emerging concept for collecting, managing and analyzing huge amount of information with cloud based technology.

2. Conceptual Analysis

2.1 Big data

Big data is a concept about the use of prediction of user usage analysis and behaviour through various strategies. Big data as a broad term means to accomplish a goal, to collect, manage, store and analyze with proper methods and strategy. Big data main benefits is for decision making, time and cost reducing technology and it help to invent new cloud based product development. Present day big data touches various different industry like education, banking, health care, retail and manufacturing and government etc.

2.2 Scopus

Scopus was launched in November 2004. It is the largest abstract and citation database of peer-reviewed literature, featuring smart tools to track, analyze and visualize research. With over 21,500 titles from more than 5,000 international publishers, Scopus delivers the most comprehensive overview of the world's research output in the fields of science, technology, medicine, social science and arts and humanities.

3. Objectives of the Study

To analysethe term"Big Data" title in SCOPUS database by various parameters such as Author wise, Document Type wise, Country wise, Publication year wise , Research area wise , Source wise. The specific objectives of the present study are.

1. To Study the Document types and number of documents in which "Big Data" title has been used.
2. To find out highly prolific authors on Big Data.
3. Toanalyse the data geographically by types of documents.
4. To Identify Publication productivity on Big Data chronologically.
5. To know highly preferred journals by the authors for writing research papers onBig Data.

4. Scope & Limitation of Study

Present Study is limited to search results in the title of 'Big Data' in SCOPUS database during 2012 to 2016 with document types and number of documents in which Big Data was used. The data analyzed with the help of Microsoft office excel and VOSviewer, Publish or Perish software.

5. Methods and Materials

The growth of publications in the Big Data research was derived from the SCOPUS published by Elsevier. During the period 2012–2016, a total of 9191 records were found for the title 'Big Data'. Necessary data was tabulated into separate sheets in terms of authorship pattern, geographical distribution of contributors, ranking list of Sources and collaborative measures.

6. Review of Related Literature

Kalantari, A., et al. (2017) discussed about bibliometric analysis study on Big data from 1980 to 2015. He has analyzed 6572 records from Thomson Reuters Web of Science. He also found trends on the concept of Big Data was Computer Science, Engineering, and Telecommunications. Similar basis study Nobre, G.C. & Tavares, E. (2017) explained about bibliometric study on the application of big data and internet of things on the concept of circular economy. He analysed 32,550 documents from Scopus. Veer &Khiste (2017) discussed the published documents and its citation from Agricultural Universities in Maharashtra during the period from 2004 to 2016.Nagarkar, S. (2015) studied analysis of faculty of Life ScienceDepartments of SavitribaiPhule Pune University during 1999-2013 with the help of Web of Science. Xian, H. &Madhavan, K. (2014) studied big data analysis in engineering research 24,172 publications with the period of 2000 to 2011

7. Chronological Analysis

It is observed from the collected data that very huge documents were written on Big Data which is reflected in Table No.1

Table No.1 Year wise documents published in Scopus on Big Data

Sr. No.	Year	Documents	Percentage
1	2016	3489	37.96
2	2015	2727	29.67
3	2014	1766	19.21
4	2013	962	10.47
5	2012	247	2.69
	Total	**9191**	**100**

Table No.1 shows that year-wise distribution of documents. The highest number of documents were published in the year 2006 i.e., 3489 (37.96%) and lowest number of documents 247 (2.69%) were published in the year 2012. However on an average 1838.2 documents were published per year during last 5 years i.e. 2012-2016.

8. Most Prolific Authors

Table No. 2 depicts highly prolific authors. It is observed that Cuzzocrea, A. and Ranjan, R. ranks first who has contributed a maximum number of 27 documents, followed by Chen, J. with 22 documents while it is also observed from the compiled data that on 5th rank 17 documents published by Cheng, X. and Paul, A.

Table No.2

Top 5 Authors which write highest documents on the 'Big Data'

Sr. No.	Author Name	Documents	Rank
1	Cuzzocrea, A.	27	1
2	Ranjan, R.	27	1
3	Chen, J.	22	2
4	Herrera, F.	19	3
5	Zhang, X.	19	3
6	Ahmad, A.	18	4
7	Cheng, X.	17	5
8	Paul, A.	17	5

The below figure indicates about co-author network. The minimum number of five documents author selected for the below graph.

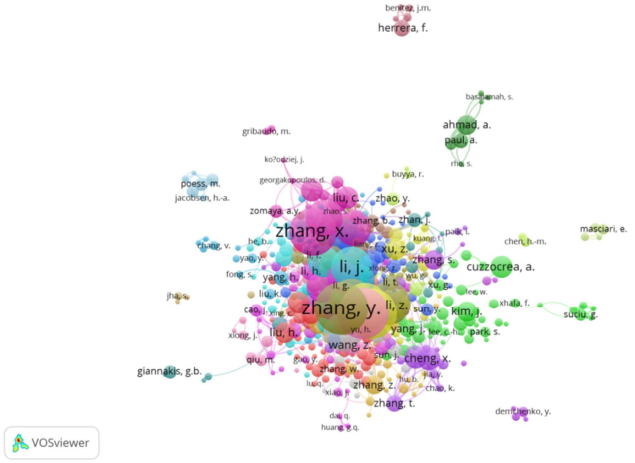

Figure 1 Co-authorship Network

9. Data Analysis by Subjects:

Table No. 3 presents the subject-wise categorization of the documents retrieved. Subject-wise analysis indicates that maximum number of contributions was in the area of Computer Science i.e. 5708 followed by Engineering with 2254 documents .The document contribution in the area of Dentistry is less i.e. 3. However , if compared the total documents with Table 1 & 3 it shows there are unmatched the figure ; that means considering the interdisciplinary approach of subject same documents were reflect in more than one subject Therefore, it can be possibility of subject overlapping.

Table No.3

Subject wise Documents Availability on Big Data

Sr. No.	Subject	Documents
1	Computer Science	5708
2	Engineering	2254
3	Mathematics	1142
4	Social Sciences	943
5	Medicine	916
6	Business, Management and Accounting	701
7	Decision Sciences	677
8	Biochemistry, Genetics and Molecular Biology	266
9	Materials Science	255
10	Energy	245
	Total=	**13107**

10. Term Occurrence

The below graph constructed on the minimum number of 15 occurrences terms out of 15107 terms.The below graph terms considered from title field of all documents. The big data term occurred in 5093 times followed by big data analytic term 583 times.

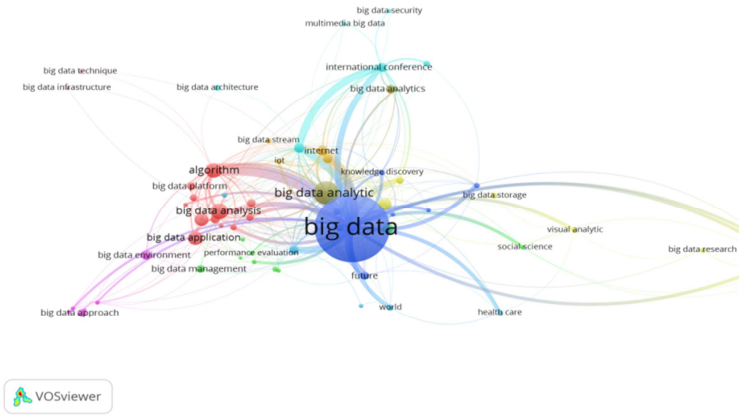

Figure 2 Term Co-occurrence

11. Documents by Sources

The below Table 4 indicates that highest ranking sources in which documents was published. It shows that Lecture Notes in Computer Science ranks first with 361 documents to its credit, followed by ACM International Conference Proceeding Series ranking on second with 146 documents. International Journal of Applied Engineering Research is on Tenth rank with 57 documents.

Table No. 4 Top Ten Source Titles

Sr. No.	Source Title	Documents	Rank
1	Lecture Notes in Computer Science	361	1
2	ACM International Conference Proceeding Series	146	2
3	Procedia Computer Science	117	3
4	Ceur Workshop Proceedings	79	4
5	Proceedings 2016 IEEE International Conference on Big Data 2016	77	5
6	Communications in Computer And Information Science	72	6
7	IBM Data Management Magazine	68	7
8	Advances In Intelligent Systems And Computing	60	8
9	Proceedings 2014 IEEE International Conference on Big Data IEEE Big Data 2014	59	9
10	International Journal of Applied Engineering Research	57	10

12. Classification by Types of Documents

The documents on Big Data were classified in twelve different types of documents which are classified by Scopus & it is presented in Table No.5.

Table No.5

Types of Documents available on Big Data

Sr. No.	Document Type	Documents	Percentage
1	Conference Paper	4290	46.68
2	Article	3002	32.66
3	Book Chapter	438	4.77
4	Review	411	4.47
5	Editorial	397	4.32
6	Note	226	2.46
7	Short Survey	121	1.32
8	Conference Review	108	1.17
9	Article in Press	65	0.71
10	Book	63	0.68
11	Letter	53	0.58
12	Erratum	17	0.18
	Total	**9191**	**100**

Table No.5 shows that the maximum number of papers published under the category of Conference Paper i.e.4290 (46.68%), whereas 3002 (32.66%)under the category Article. There are438 (4.77%)Book Chapter and the Review are411 (4.47%). A small number of contributions are categorized under Erratum i.e. 17(0.18%).

13. Geographical Analysis of Documents

The Twelve types of documents on Big Data were written by various authors and published

from different countries. However author has presented geographical distribution of top twenty countries & presented in Table No.6.

Table No.6

Top 20 Country credits highest documents

Sr. No.	Country Name	Documents	Rank
1	United States	2732	1
2	China	1826	2
3	India	717	3
4	United Kingdom	567	4
5	Germany	448	5
6	South Korea	357	6
7	Australia	333	7
8	Canada	322	8
9	Italy	285	9
10	Spain	261	10
11	Japan	251	11
12	France	219	12
13	Taiwan	135	13
14	Netherlands	132	14
15	Hong Kong	116	15
16	Switzerland	101	16
17	Singapore	85	17
18	Brazil	82	18
19	Greece	78	19
20	Malaysia	77	20

Table No. 6 depicts the geographical distribution of authors. Among 9191 documents, United States tops the list with 2732 documents, followed by China with 1826documents to its credit. India's contribution to 'Big Data' is 717 documents during 2012–2016 which is ranked on 3rd&Malaysiapublished 77documents with rank 20thposition.

14. Data Analysis by Affiliation

The Compiled data has been analyzed by affiliation institute & presented in Table No.7. It reflects the status of contribution of documents written on big data.

Table No.7

Affiliation wise Distribution of Documents

Sr. No.	Affiliation Name	Documents	Rank
1	Chinese Academy of Sciences	191	1
2	Tsinghua University	99	2
3	Wuhan University	62	3
4	Shanghai Jiao Tong University	50	4
5	Massachusetts Institute of Technology	49	5
6	Beijing University of Posts and Telecommunications	47	6
7	IBM Thomas J. Watson Research Center	45	7
8	Vellore Institute of Technology	44	8
9	University of Southern California	44	8
10	University of Toronto	42	9
11	University of Technology Sydney	42	9
12	Huazhong University of Science and Technology	42	9
13	CNRS	41	10
14	New York University	41	10

Table No. 7 presents the list of top ten Affiliation contributions on the subject Big Data. The institution affiliation from the address field is taken as the data for this categorization. Chinese Academy of Sciencescontributed 191documents which is the highest while Tsinghua University has 99 documents to its credit &CNRS & New

York Universitycontributed41 documents with rank tenth.

15. Source Type wise Documents

Table No.8 Source Type wise Documents on Big Data

Sr. No.	Source Type	Documents	Percentage
1	Journals	3895	42.38
2	Conference Proceedings	3715	40.42
3	Book Series	825	8.98
4	Books	465	5.06
5	Trade Publications	288	3.13
6	Undefined	3	0.03
	Total	**9191**	**100**

Table No.8 shows that the maximum number of documents published under the Source type of Journals is 3895 (42.38%), whereas 3715 (40.42%) under the Conference Paper. There are 825 (8.98%) Book Series and the Books are 465 (5.06%). A small number of contributions are categorized under Undefined is 3 (0.03%).

16. Languages wise documents

There are hundreds of languages in the world, the data are analysed by language to know the languages in which highest documents contributed in Scopus on Big Data. The related information indicates by Table No.9.

Table No.9

Top 5 Preferred Languages for writing documents on Big Data

Sr. No.	Languages	Documents	Rank
1	English	8743	1
2	Chinese	267	2
3	German	100	3
4	French	39	4
5	Japanese	30	5
6	Spanish	30	5

The Table No.9 indicates that English is the most preferred language for publication of 8743 documents on Big Data followed by in Chinese publishing 267 documents and In Japanese and Spanish language has published 30 documents on Big Data with rank fifth.

17. Major Inferences

- In Scopus, under the category Big Data, 9191 documents were retrieved among which maximum number of documents was contributed in the year 2016 and minimum in 2012.
- Subject-wise analysis indicates that maximum number of contributions was in the area of Computer Science i.e. 5708 documents& Dentistry is less i.e. 3 documents.
- Document-wise study reveals that the maximum numbers of documents published are under the category “Conference Paper”.
- Institution-wise distribution shows that “Chinese Academy of Sciences” contributed 191 documents which are the highest while Tsinghua University has 99 documents to its credit & CNRS and New York University contributed 41 documents with rank tenth.

- Country-wise analysis indicates that United States tops the list with 2732 documents, followed by China with 1826 documents to its credit. India's contribution to 'Big Data' is 717 documents during 2012–2016 which is ranked on 3rd& Malaysia published 77 documents with rank 20th position.

18. Conclusion

The data suggest that there was a significant research activity in the field of Big Data during the study period. The contributions of by year on Big Data indicate that there is healthy pattern of progress in this field. As per data retrieved, overall Big Data H-index is 80 and g-index is 137.

References

1. Kalantari, A., Kamsin, A., Kamaruddin, H.S. et al. (2017)Abibliometric approach to tracking big data research trends, *J Big Data, 4*(30). Retrieved from https://117.232.125.7:2726/10.1186/s40537-017-0088-1
2. Khiste G.P. (2017) Publication Productivity of 'Consortia' by Scopus during 1989-2016, *International Journal of Current Innovation Research*, 3(11), 879-882.
3. Khiste G.P. &Paithankar R.R. (2017) Analysis of Bibliometric Term in Scopus. *International Journal of Library Science and Information Management, 3*(3), 81-88.
4. Mishra, D., Mishra, D., Luo, Z., Luo, Z., Jiang, S., Jiang, S., .&Dubey, R. (2017). A bibliographic study on big data: concepts, trends and challenges. *Business Process Management Journal*, *23*(3), 555-573.
5. Nagarkar, S., Veer, C., &Kumbhar, R. (2015). Bibliometric Analysis of Papers Published by Faculty of Life Science Departments of SavitribaiPhule Pune University during 1999-2013. *DESIDOC Journal of Library & Information Technology, 35*(5). doi:http://dx.doi.org/10.14429/djlit.35.5.8429
6. Nobre, G.C. & Tavares, E. (2017). Scientific literature analysis on big data and internet of things applications on circular economy: a bibliometric study.Scientometrics 111: 463. Retrieved from https://117.232.125.7:2726/10.1007/s11192-017-2281-6
7. Veer D.K. & Khiste G.P. (2017). Mapping of Intellectual Assets of Agricultural Scientists with special Reference to Indian Citation Index. *National Conference on Emerging Library and Information Science and Technologies*. Rahuri, 2017. 181-189.
8. Scopus. (2017, December 2017) Retrieved from https://www.scopus.com/
7. Xian, H. &Madhavan, K. (2014). Anatomy of scholarly collaboration in engineering education: A big-data bibliometric analysis. *Journal of Engineering Education*, *103*(3), 486-514. doi:10.1002/jee.20052