

Association for Information Systems

AIS Electronic Library (AISeL)

ICEB 2016 Proceedings

International Conference on Electronic Business
(ICEB)

Winter 12-4-2016

A Research Growth Study in Big Data field

Yao-Hung Lin

National Chengchi University, 101356510@nccu.edu.tw

Jiann-Min Yang

National Chengchi University, jmyang@nccu.edu.tw

Follow this and additional works at: <https://aisel.aisnet.org/iceb2016>

Recommended Citation

Lin, Yao-Hung and Yang, Jiann-Min, "A Research Growth Study in Big Data field" (2016). *ICEB 2016 Proceedings*. 51.

<https://aisel.aisnet.org/iceb2016/51>

This material is brought to you by the International Conference on Electronic Business (ICEB) at AIS Electronic Library (AISeL). It has been accepted for inclusion in ICEB 2016 Proceedings by an authorized administrator of AIS Electronic Library (AISeL). For more information, please contact elibrary@aisnet.org.

A Research Growth Study in Big Data field

Yao-Hung Lin, National Chengchi University, Taiwan, 101356510@nccu.edu.tw

Giann-Min Yang, National Chengchi University, Taiwan, jmyang@nccu.edu.tw

ABSTRACT

Responding to the diffusion and growth of big data research, this study adopted the bibliometric approach to describe the growth of the literatures, the distribution of journals, publication countries and subject area. This study collected the relative literature by querying the Social Science Citation Index (SSCI) of ISI Web of knowledge database, where we could collect the big data literatures in academic papers, systematically. Data from citation indexes can be analyzed to determine the popularity and impact of specific articles, authors, and publications. The results provided the distribution of core journals, and described the trends and feature of big data research for researchers interested in this field.

Keywords: Big data, bibliometrics, Social Science Citation Index

INTRODUCTION

Traditional data processing applications, such as data mining, statistical regression, and machine learning techniques are insufficient for big data, a term for data sets that are so large or complex and in the business perspective, big data simple refers to the use of predictive analytics or certain other advanced methods to extract value from data (Wikipedia, 2016). With the development of data science, big data is nowadays a novel method for research, military and business. Research and development of big data has grown very rapidly in recent years (see Figure 1) and becomes increasingly important in practice. This paper investigated the characteristics of big data literature and its implication during 2008 to 2015 by bibliometric approach. The bibliometric method adopts statistical techniques of bibliography counting to evaluate and quantify the growth of literature of a subject. The information includes the journal or other publication title, the authors' name and affiliation, document type, etc. This study used the Thomson Reuters' Citation database of Social Sciences Citation Index (SSCI) on ISI Web of Knowledge website to retrieve data. The query for general search was performed with keywords as big data with double quotation marks ("big data"). Consequently, total 918 bibliographic records were retrieved.

This study aimed: (1) to explore the growth of big data literature; (2) to identify major contributing countries that publish big data papers most; (3) to identify core journals that contain most part of big data literature.

THE GROWTH OF BIG DATA LITERATURE

Figure 1 showed the trend of academic growth from the collecting the literature of big data. In 2008, there were only 4 journal papers related to big data in the ISI web of science data base. Obviously, the tendency of literature growth of big data increased explosively from 2012 in social science.

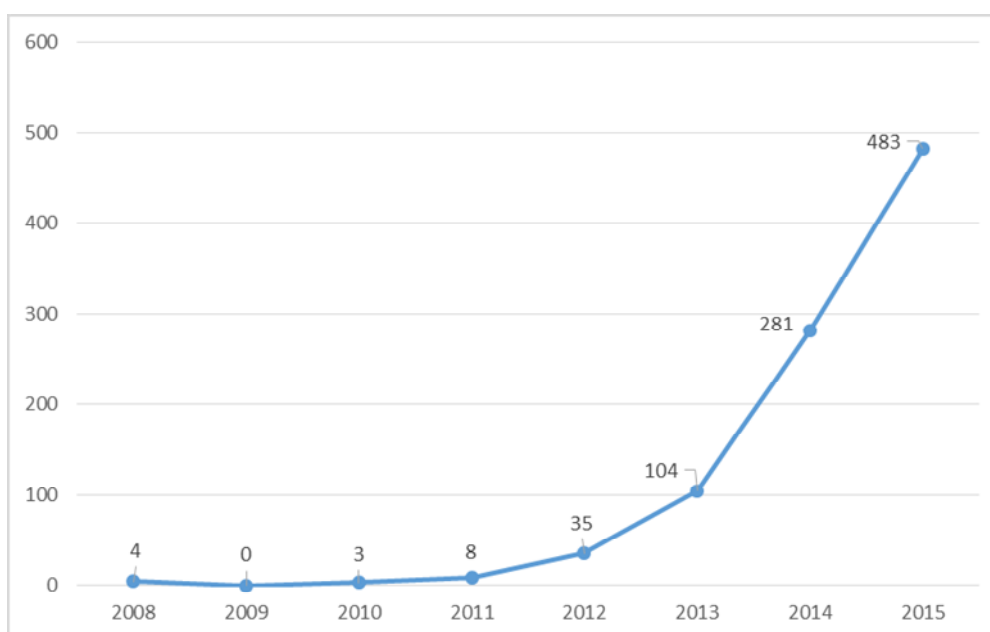


Figure 1: Annual journal production of big data literature

CHARACTERISTICS OF BIG DATA LITERATURE

Country Distribution

There were 45 countries publishing big data literature. Table 1 showed country distribution of big data literature and about 55%

of the big data literature was published from USA. England (14.82%) and People Republic of China (6.76%) contributed the 2nd and 3rd position, and Australia, Germany, Canada, and Netherlands each of them contributes more than 3% among the total literature. The researchers in these countries contributed on big data research has drawn the attention of the world. Additionally, researchers in Taiwan had published 9 (0.98 %) journal papers on big data field collected in this database.

Table 1:Country distribution

Country	Record Count	Percent
USA	507	55.23 %
ENGLAND	136	14.82 %
PEOPLES R CHINA	62	6.76 %
AUSTRALIA	45	4.90 %
GERMANY	41	4.47 %
CANADA	35	3.81 %
NETHERLANDS	28	3.05 %
SOUTH KOREA	27	2.94 %
ITALY	25	2.72 %
SPAIN	23	2.51 %

Subject Area

The mainly concerned area of big data literature includes “BUSINESS ECONOMICS”, “INFORMATION SCIENCE LIBRARY SCIENCE”, “COMPUTER SCIENCE”, “Business & Economics” and “Computer Science”. The major area of big data literature is business economics. However, the result also indicates that big data has been considered as an important issue in information science library science as well as computer science, health care science services and Electronic, that is, big data compose of many knowledge that are interdisciplinary. Table 2 showed the top 15 concerned subject area of big data literature.

Table 2:Subject area of big data literature

Rank	Subject area	Count	Percent
1	BUSINESS ECONOMICS	198	21.57 %
2	INFORMATION SCIENCE LIBRARY SCIENCE	135	14.71 %
3	COMPUTER SCIENCE	124	13.51 %
4	HEALTH CARE SCIENCES SERVICES	74	8.06 %
5	GOVERNMENT LAW	73	7.95 %
6	COMMUNICATION	61	6.65 %
7	GEOGRAPHY	57	6.21 %
8	SOCIAL SCIENCES OTHER TOPICS	57	6.21 %
9	PSYCHOLOGY	50	5.45 %
10	ENGINEERING	43	4.68%
11	SOCIOLOGY	43	4.68 %
12	ENVIRONMENTAL SCIENCES ECOLOGY	32	3.47 %
13	OPERATIONS RESEARCH MANAGEMENT SCIENCE	32	3.49 %
14	PUBLIC ADMINISTRATION	31	3.38 %
15	PUBLIC ENVIRONMENTAL OCCUPATIONAL HEALTH	31	3.38 %

Keywords

In addition to the analysis of subject area, this study also conducted the keyword analysis. Numbers of keyword from literatures not only reflect the research topic, but also provide a convenient way to search and retrieval. Table 3 also showed top 15

keywords. By collecting the most commonly listed of keywords, the most research issues in big data field from past research are identified clearly.

Table 3: Top 15 high frequency keywords in big data literature

Rank	Keyword (905 kinds)	Count	Percent
1	big data	231	10.7%
2	Social media	33	1.5%
3	Twitter	18	0.8%
4	Privacy	17	0.8%
5	Big Data Analytics	16	0.7%
6	Analytics	12	0.6%
7	Data science	11	0.5%
8	Data	11	0.5%
9	Internet	11	0.5%
10	Data mining	10	0.5%
11	Computational social science	9	0.4%
12	Social networks	9	0.4%
13	Predictive analytics	9	0.4%
14	Ethics	8	0.4%
15	Data protection	7	0.3%

BRADFORD LAW AND CORE JOURNALS

Bradford (1934) proposed a law that numerous relevant papers would be concentrated in a small group of journal titles, and the rest papers would be scattered over the vast account of journals. Therefore, the journals could be divided into three groups according to the formulation and the proportion of the number of journals in each group would be $1:n:n^2$. Through the law, some core journals in the research field could be identified. The Bradford's law has been widely used to study author productivity (Chung, 1994), and journal literature distribution (von Ungern-Sternberg, 2000). In total, there are 457 journals, publishing 918 big data papers. According to zone of Bradford's law, table 4 showed the amount of publication in each group and relevant information listed by the amount of publishing papers. The cumulative number of papers for each group against its ranks for literature of big data was illustrated in figure 2. Table 4 also offered the comparisons of these three zones (3:34:420), which equals to $1: 11.33: 11.33*12.35$. That is, $1: 11.33: 11.33*12.35 = 1:n:n^2$, approximately. The result approached the explanations of Bradford's Law. Table 5 specifies 3 core journals title and top 5 relevant journals title in big data.

Table 4: The distribution of big data literature

	Articles(a)	Journals(b)	Amount of articles (b)*(c)	Accumulated Journals (c)	Accumulated articles (c)
(A)	18	1	18	1	18
Core	17	1	17	2	35
	14	1	14	3	49
	(B)	13	2	26	5
Relevant	12	2	24	7	99
	11	2	22	9	121
	10	3	30	12	151
	9	3	27	15	178
	8	7	56	22	234
	7	5	35	27	269
	6	4	24	31	293
	5	6	30	37	323
(C)	4	16	64	53	387
marginal	3	29	87	82	474
	2	69	139	151	612
	1	306	306	457	918

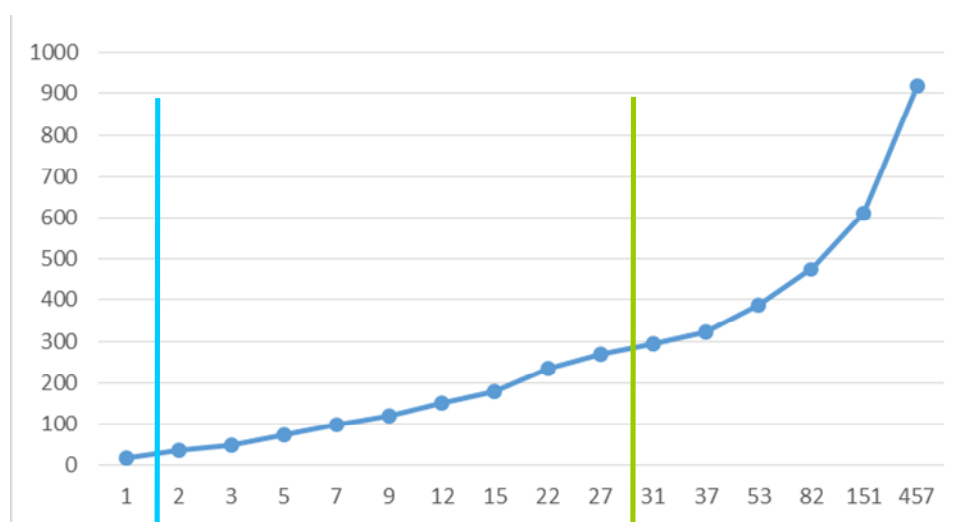


Figure2: The Bradford plot of big data journal literature

Table 5: Core journals and Top 5 relevant journals title

Journal category	Journal title	Count	Percent
Core journal	HEALTH AFFAIRS	18	2.0%
	VALUE IN HEALTH	17	1.9%
	JOURNAL OF THE AMERICAN MEDICAL INFORMATICS ASSOCIATION	14	1.5%
Relevant journal	HARVARD BUSINESS REVIEW	13	1.4%
	INTERNATIONAL JOURNAL OF COMMUNICATION	13	1.4%
	COMPUTER LAW & SECURITY REVIEW	12	1.3%
	INFORMATION COMMUNICATION & SOCIETY	12	1.3%
	ECONTENT	11	1.2%

CONCLUSION

This work investigated the growth of big data literature in the Social Science Citation Index (SSCI) of ISI Web of knowledge database, and reviewed some characteristics of big data literature through bibliometric techniques. According to the analysis, there are several findings: First, big data is considered as an important issue that the volume of relevant researches increases dramatically from 2012. Secondly, there are 3 core journals identified by the Bradford law, about 5.4% of big data literature is concentrated in top 3 journals, and the rest 94.6% is widely spread on other 452 journals. Finally, the analysis of the core journals also demonstrates that the big data core journals are high health-medical relevant. This work is the preliminary stage that researchers would probe the trend of big data and future study may adopt the different analysis such as co-citation to get the deeper comprehension and find more useful information behind the big data filed.

REFERENCES

- [1] Bradford, S.C. (1934). Sources of Information on Specific Subjects. *Engineering: An Illustrated Weekly Journal* (London), 137, 85-86.
- [2] Chung, Y.K. (1994). Bradford distribution and core authors in classification systems literature. *Scientometrics*, 29 (2), 253-269.
- [3] von Ungern-Sternberg, S. (2000). Bradford's Law in the context of information provision. *Scientometrics*, 49 (1), 161-186.
- [4] Wikipedia (2016). Big data. https://en.wikipedia.org/wiki/Big_data