

OVERVIEW OF DBLP- COMPUTER SCIENCE BIBLIOGRAPHY

Manas Kumar Ray

Lecturer

Department of BCA

Banwarilal Bhalotia College, Asansol

Sudipta Ghosh

Departmental Librarian

(BBA, BCA)

Banwarilal Bhalotia College,

Asansol

Abstract : The paper aims to highlight the Database and Logic Programming Bibliography of Computer Science (DBLP). It is mainly designed to cater information to the computer science people to provide open access to all computer science research community. DBLP data has been released under the Open Data Commons ODC-BY 1.0 license. As per condition of licenses anyone is free to 'copy, distribute, use, modify, transform, build upon, and produce derived works' from the retrieved data. DBLP advisory board consists of many renowned computer scientists from different areas of computer. It caters information on popular books, journals and proceedings of Computer Science. It is a joint venture of University of Trier in Germany and Schloss Dugstuhl, Leibniz centre for informatics Computer science research centre in Germany.

Keywords: Open access (OA), e- books, e- theses, Database and Logic Programming Bibliography of Computer Science (DBLP).

1.0 Introduction

Due to popularity of open access in the present decade, most of the people are familiar with open access publication. In this respect DBLP (Database and Logic Programming Bibliography of Computer Science) , a computer science bibliographical database, hosted at University of Trier, in Germany is noteworthy. (Wikipedia, 2017) It was inaugurated as an experimental server basis in 1993 to examine web technology and was created by Michael Ley. DBLP (DBLP , 2017) is a searchable index with links to the full-texts of the publications of theses, books, edited books and conference proceedings and it is mainly designed to cater information to the computer science books, theses, conference proceedings, etc. to provide open access to research community of computer science. DBLP data has been released under the Open Data Commons attribution licenses (ODC-BY) 1.0 license. As per condition of this license, anyone is free to 'copy, distribute, use, modify, transform, build upon, and produce derived works' from the retrieved data. DBLP advisory board consists of many renowned computer scientists from different corners of the world. The huge XML, JSON, rdf, bibTeX file in DBLP contains the bibliographic records of books, theses, conference proceedings comprising of two million bibliographic records. DBLP is not restricted to database management system but it is limited to file management system. For academicians, researchers and students of computer science, the DBLP website is a search tool to find out the contemporary research papers. At first it was "Digital Bibliography & Library Project" intended to help academicians, researchers of computer science for helping free access to the e-publications. Now DBLP is a joint project of Schloss Dagstuhl - Leibniz Centre for Informatics and the University of Trier. At present DBLP has total 3978379 numbers of publications, 1999891 numbers of authors sorted by the last name, 5250 number of conferences alphabetically arranged, 1562 Journals alphabetically arranged are indexed in DBLP as on 30.10. 2017. The huge number of e-publications indicates the popularity of DBLP among the academicians, researchers of the field of Computer Science.

2.0 Related works

To find out the pattern of the study, the related works have been recorded to observe the undefined area in the previous study. Konings in his study in 1985 compared seven bibliographies in the area of computer science and observed the merits and demerits of different bibliographies (Konings , 1985).Michael Ley in his review paper throws light on evolution of DBLP in general (Ley, 2002). In Computer Science and Computing: A Guide to the Literature is a study on the literature of computer science and computing and presented the importance of computer

and computing when the book is going to vanish (Computer Science and Computing, 2006). Mr. Ley in his work, he pointed out the functioning of DBLP and how it works in web environment (Ley, 2009). Orr and Ortiz identified that DBLP is a repository and the pattern of presenting information in DBLP is not always simple in nature (Orr and Ortiz, n.d.). Biryukov and Dong have completed analysis on computer science communities as per DBLP data. They studied to evaluate various facets of each research area (Biryukov and Dong, 2010). Pedersen, L. A., & Arendt, J. Found that in google scholar the papers related to computer science are decreasing day by day and they showed 52% conference papers and 55% journal articles were retrieved from google scholar (Pedersen & Arendt, 2014). S. Kumar in his paper studied a review of literature on co-authorship networks and found that co-authorship is the main person of writing a literature (Kumar, 2015).

3.0 Statement of the problem

A lot has been written about bibliography in general but there are very few studies that concentrated on bibliography of computer science and it is to be noted that no such study has been undertaken to present an overview of Computer Science books indexed in DBLP. This paper will be beneficial for the publishers, authors and librarians. To increase the visibility of e books, theses, conference proceedings, edited books, etc of Computer Science indexed in DBLP, the study has been conducted.

4.0 Objective of the study

The main objective of the study is to identify the publication trend in DBLP. The detailed objectives are stated as follows----

- To identify year-wise growth of e-documents indexed in DBLP.
- To find out number of publications per author.
- To know the number of co-authors per author.
- To find out the number of authors per publication.
- To visualize the graph relating to publication traits of DBLP.

5.0 Scope and Limitations

The coverage of the study is restricted to open access e-books, theses, conference proceedings in Computer Science indexed in DBLP and the period of coverage is up to from 01.01.1997 to 31.12.2016.

6.0 Methodology

The requisite data was retrieved from the DBLP on the 14th November, 2017 after retrieval of data, the data was entered in a predefined excel format for better understanding of data. The requisite data was then analyzed and it was presented in tabular and graphical forms to throw light on the latest trend of e-publication in DBLP.

7.0 Analysis and interpretation of data

From the above mentioned objectives the following tables and figures have been prepared for better understanding and interpretation of DBLP database. The understanding of database of DBLP is not an easy matter so it is necessary to draw table and figures. Total three tables and figures are stated below:-

The above table-01 shows the growth of Computer Science e-books, theses, data artefacts, etc. in DBLP. It is evident from the above table that a steady growth of Computer and information science e resources has been observed from the year of 1997 to 2016. In 1997, total number of e-documents was 68593 whereas in 2016 total number of e-documents is 3607787. So the growth in number is 3539194 e-documents in two decades. It is also observed that journal articles were the largest in number in DBLP followed by the book and theses. The notable attribute of the study is that data are available in DBLP on and from 01.01.2016.

Table 1-: Year-Wise Growth of e-documents indexed in DBLP

Content	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006
Journal Articles	38990	4948	63725	78692	99420	113781	167386	214909	254296	299728
Book and Theses	212	303	425	455	521	869	1037	1157	1176	1191
Data and Artifacts	00	0	0	0	0	0	0	0	0	0
Editorship	39	65	216	297	1184	3093	4572	6023	7193	8412
Parts in Books or Collection	5	66	645	682	928	1020	1350	1538	1793	2514
Informal Publication	2	5	5	5	5	5	5	5	5	7
Conf. Workshop paper	29345	47137	82479	106281	150026	221298	286089	364011	434202	510094
Ref. Works	0	0	0	0	0	0	0	0	0	0
Withdraws item	0	0	0	0	0	0	0	0	0	0
Total	68593	97224	147495	186412	252084	340066	460439	587643	698665	821946

Content	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016
Journal Articles	361026	424566	509121	607403	746624	892217	1042924	1174358	1327217	1475714
Book and Theses	1216	1261	1354	1440	9361	16043	16829	17546	18368	51193
Data and Artifacts	0	0	0	0	0	0	0	0	0	14
Editorship	9750	11604	13164	14670	17105	20018	22613	25957	29853	33567
Parts in Books or Collection	2594	8326	14510	15101	19643	21779	24953	27080	29465	32969
Informal Publication	8	8	8	8	13	13646	29202	47811	68366	92981
Conf. Work shop paper	598439	687183	784963	894743	1049454	1201001	1349326	1544474	1725569	1914272
Ref Works	0	0	0	0	0	961	961	1992	5807	7076
Withdraws item	0	0	0	0	0	0	01	01	01	01
Total	973033	1132948	1323120	1533365	1842200	2165665	2486809	2839219	3204646	3607787

Table 2- Number of Publications per author

Number of Publications	Number of Authors
02	309640
03	153944
04	95505
05	65721
06	48922
07	37869
08	30296
09	24700
10	20794
11	17532
12	14758
13	12602
14	11035
15	9684
16	8679

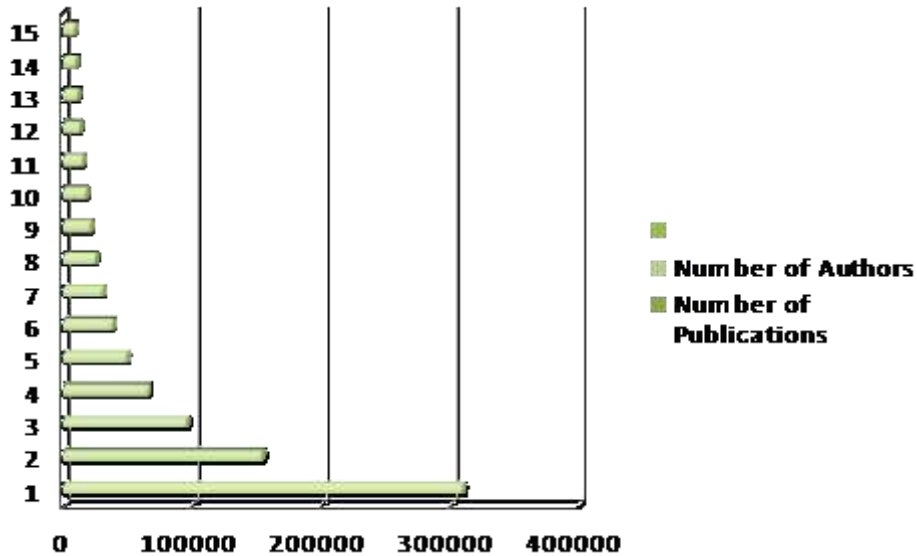


Figure-1: Number of publications per author

The above table-02 explicates that the double authors publish the highest number of papers followed by triple authors then followed by four authors and it is also noteworthy that the number of authors are indirectly proportionate to number of publications. The notable attribute of the study is that 8679 authors publish 16 publications. On the other hand 309640 authors have only two publications.

Table 3- Number of co-authors per author

Number of co-authors	Number of authors
01	225302
02	296132
03	264512
04	201998
05	150528
06	109790

07	85531
08	66543
09	54402
10	43322
11	36614
12	30539
13	26714
14	23059
15	19740
16	17717
17	15567
18	13794
19	12644
20	11458
21	10187
22	9304
23	8753
24	7768
25	7081
26	6773
27	6252
28	5635
29	5259
30	4780
31	4625
32	4410
33	4190
34	3834
35	3704
36	3396
37	3287
38	3020
39	2878
40	2765
41	2633
42	2430
43	2302
44	2286
45	2220

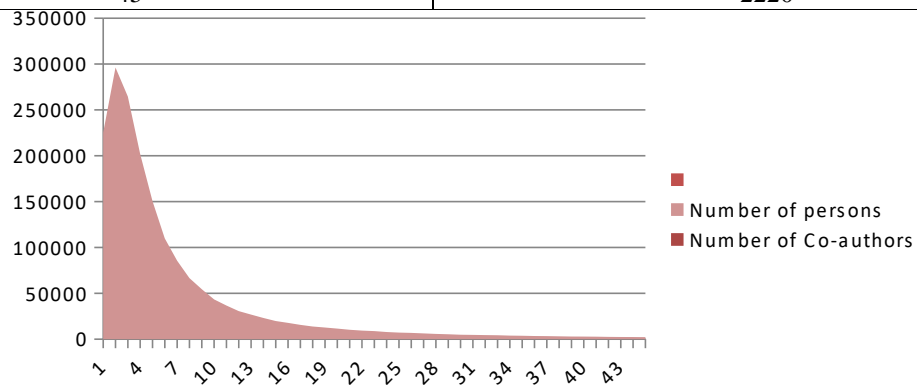


Figure 2- Number of Co-authors per author

It is retrieved from the table 3 and figure 2 that the three co-authors are the highest in number in terms of author and forty-five co-authors are the lowest in number in terms of one author in DBLP. Two co-authors are lying with the 296132 number of authors and Forty-five co-authors are lying with the 2220 number of authors.

Table 4- Number of authors per publication

Number of Authors	Book and Theses	Conference and Workshop Papers	Editorship	Journal Articles	Informal Publications	Reference works	Parts in Books or Collections
01-20	76966	2074899	36372	1568718	172641	20873	22941
21-40	0	371	0	807	88	0	0
41-60	0	28	0	67	8	0	0
61-80	0	6	0	8	3	0	0
81-100	0	2	0	3	1	0	0
101-120	0	2	0	0	0	0	0
121-140	0	1	0	0	0	0	0

The table 4 presents number of authors per publication. Total 2074899 conference and workshop papers are sorted when number of author lies between 1 to 20. Then it is followed by 1568718 journal articles indexed in DBLP. The notable matter of the study is that the number of authors increases and publication per author decreases accordingly. The aforementioned matter is also true in case of books and theses, conference and workshop, informal publications, etc. According to type of publications, it is noteworthy to state that conference and workshop papers (2074899) are the highest in number followed by journal articles(1568718). Informal publications, reference works, parts in book or collection must be mentioned to be part of DBLP database.

8.0 Discussions on findings of the study

From table-01, it exports record through BibTex, RIS, RDF N-Triples, RDF/XML, and XML. The record can be shared through twitter, Raddit, Mendeley, BibSonomy, LinkedIn, Google+, facebook. It provides link through which some documents can be retrieved from Google, Google Scholar, MSAcademic, CiteSeerX, CORE, Semantic Scholar. In the table-02, it is observed that the authors are available holding 02 to 16 publications. From table-03, it is found that 01 to 45 co-authors are present in different publication indexed in DBLP. An innovative news feed is also available in the website of DBLP. From the table-04, it is clear that number of publication increases when the number of authors decreases. In the last two figures, the visualization of the graph relating to distribution of publication type, publications per year, number of authors per publication and number of publication per author and publication traits in DBLP i.e. number of co-authors per person, records in DBLP, new records per year, new records per month is observed.

9.0 Suggestions & Conclusions

The term Open Access was popularized in the past decade but various open access channels have been prevalent now. In the arena of open access, DBLP is an open access repository of reference works, books and theses, conference and workshop papers and some informal publications on Computer Science. As on 31.12.2016, total 3607787 e-documents are indexed in DBLP. Among them journal articles are the largest in number followed by books and theses. From the above interpretation and analysis the following suggestions have been enumerated below:

- (i) Though the year wise growth is satisfactory, authors and publishers should submit more works in the DBLP.
- (ii) A huge number of documents are indexed in DBLP so proper indexing and infrastructural facilities should be upgraded in the DBLP to increase the accessibility of e-documents.
- (iii) There is a controversy regarding the Ranking and profiling of scholars, institutions, journals, or conferences indexed in DBLP. The system of ranking and profiling of scholars, institutions, journals, or

conferences indexed in DBLP must be stopped.

- (iv) The presentation of visual graph is very much helpful for retrieval of data indexed in DBLP but it is suggested that the presentation of visual graph should be clearer.

In short, DBLP is an online open access collection of valuable e-resources on computer science which will be helpful not only for academicians and researchers but also for anyone interested in computer science.

10.0 References

- Biryukov, M & Dong, C. Analysis of computer science communities based on DBLP . In *Proceedings of the 14th European conference on Research and advanced technology for digital libraries*, ECDL'10, Berlin, Heidelberg, 2010. Springer-Verlag.
- Computer Science and Computing: A Guide to the Literature. (2006). *Reference Reviews*, 20(6), 41–41. <https://doi.org/10.1108/09504120610687308>
- DBLP. (2017). About DBLP. Retrieved from <http://DBLP.uni-trier.de/> (Accessed on 22.11.2017)
- Konings, C. A. G. (1985). Comparison and evaluation of nine bibliographies/ bibliographic databases in the field of computer science. *Online Review*, 9(2), 121–133. <https://doi.org/10.1108/eb024178>
- Kumar, S. (2015). Co-authorship networks: a review of the literature. *Aslib Journal of Information Management*, 67(1), 55–73. <https://doi.org/10.1108/AJIM-09-2014-0116>
- Ley, Michael. (2002). The DBLP Computer Science Bibliography: Evolution, Research Issues, Perspectives. String Processing and Information Retrieval. LNCS.Springer-Verlag. 2000: 481–486. doi:10.1007/3-540-45735-6_1 (Accessed on 12.11.2017)
- Ley, Michael . (2009). DBLP: Some Lessons Learned (PDF). VLDB. *Proceedings of the VLDB Endowment*. 2 (2).1493–1500. doi:10.14778/1687553.1687577. (Accessed on 30.11.2017)
- Laurel Orr & Jennifer Ortiz. (n.d.) Retrieved from <https://homes.cs.washington.edu/~jortiz16/images/MLProjectPaper.pdf> (Accessed on 10.11.2017)
- Pedersen, L. A., & Arendt, J. (2014). Decrease in free computer science papers found through Google Scholar. *Online Information Review*, 38(3), 348–361. <https://doi.org/10.1108/OIR-07-2013-0159>
- Wikipedia (2017). About DBLP. Retrieved from <https://en.wikipedia.org/wiki/DBLP> (Accessed on 22.11.2017)