DESIDOC Journal of Library & Information Technology, Vol. 40, No. 6, Nov 2020, pp. 375-381, DOI : 10.14429/djlit.40.6.15680

Research Collaboration and Authorship Pattern in the field of Semantic Digital Libraries

Shriram Pandey[#,*] and Sidhartha Sahoo[$]

[#]*Banaras Hindu University, Varanasi - 221 005, India*

[$]*Indian National Centre for Ocean Information Services, Ministry of Earth Sciences, Hyderabad - 500 090, India*

[*]*Email: shriram.lib@bhu.ac.in*

ABSTRACT

This study aims to explore research collaborations and authorship patterns in the field of semantic digital libraries(SDL). The data is extracted (N=2075) from the Scopus database using keywords related to semantic digital libraries by considering all types of publications during 1983-2019. The analysis of each document is based on the following scientometrics indicators: author productivity, degree of collaboration, collaboration index, collaboration coefficient and modified collaboration coefficient. Correlation matrices were also calculated and inferences drawn in terms of authors and publications. A network visualisation tool VOSviewer was used to present authorship correlation network strength and keyword mapping for a better insight into the emerging areas in the field of SDL. The resulting average degree of collaboration of 0.898 indicates that a large number of publications are multi-authored and that there is a higher level of collaborative research in the field of semantic digital libraries. Meghini C from the Institute of Information Science and Technologies, Italy has produced the highest number of research paper(n=18) whereas Egenhofer MJ found to be a profoundly impacted author with 851 citations on in the studied domain. Results also reveal that the focus areas of research related to SDL include digital libraries, semantic web, ontology, metadata and information retrieval. However, keywords such as natural language processing system, computational linguistics, linked data are also repeated frequently in the published literature, thus revealing the emerging areas of the future research in the domain of SDL.

Keywords: Authorship metrics; Collaborative index; Semantic digital library; Semantic web.

1. INTRODUCTION

Digital libraries technological research has opened a fresh possibility to add semantic web applications for information storage, access and retrieval. The semantic web offers various tools and applications deployed in the digital library domain. The semantic web integrates common standards and technologies in a collaborative way which provides the opportunity to the better showcase of valuable digital resources (Macgregor, 2008)[1]. Semantic web technologies offer a valuable add-on for digital libraries and there are numerous academic and commercial semantic tools available which can be applied to the digital libraries(Sure & Studer, 2005)[2]. Library professionals core expertise such as cataloguing and classification of resources for the creation and maintenance of metadata and taxonomies are equally helping in designing the semantic web applications (Burke, 2009)[3].Semantic-based digital libraries termed as 'Semantic Digital Library(SDL)' is the combinations of technologies which integrates information resources, vocabulary controlled devices, bookmarks, taxonomies and user profiles. SDL works with RDF(Resource Description Framework) and provides interoperability with other systems and data. SDL offers an adaptive, robust and user-friendly search and browsing interface for information retrieval(Kruk, 2010)[4]. The emergence of social media tools offers tremendous opportunities to create, share, annotate and collaborate. Integration of these tools to the digital libraries, the ever-changing relationship between libraries and users can be more strengthen. Semantic technologies provide a framework to describe objects, for instance, the need to establish typical schemes in the form of ontologies(Sure &Studer, 2005b, p.3)[5]. Users community, as well as developers widely adopt the integration of semantic web technologies with digital libraries. Research in the field of semantic digital libraries discovered that there is a progressive trend of scientific research output in the field of semantic digital libraries. The scope of the study is limited to presents an insightful authorship collaborative patterns of the research in the domain of SDL. The research timeline is limited to the year 1983 to 2019.

2. LITERATURE REVIEW

Semantic web application to digital libraries is a broad area of research. Semantic digital library based documents in the context of bibliometrics or scientometrics are rare. Development of library technology would remain significantly incomplete until deployment in the practice of semantic web technologies

Received : 8 April 2020, Revised : 27 June 2020
Accepted : 8 July 2020, Online published : 03 December 2020

Table 1. Authors productivity in the related study area

Authors productivity				
Year	**Total number of articles**	**Total number of Authors**	**AAPP**	**Productivity per Author**
1983	1	1	1.000	1.000
1984	1	1	1.000	1.000
1986	1	1	1.000	1.000
1988	1	1	1.000	1.000
1989	1	1	1.000	1.000
1992	1	1	1.000	1.000
1995	2	2	1.000	1.000
1996	8	7	0.875	1.143
1997	14	13	0.929	1.077
1998	26	24	0.923	1.083
1999	21	19	0.905	1.105
2000	23	22	0.957	1.045
2001	38	34	0.895	1.118
2002	38	34	0.895	1.118
2003	54	52	0.963	1.038
2004	82	79	0.963	1.038
2005	99	92	0.929	1.076
2006	97	90	0.928	1.078
2007	107	101	0.944	1.059
2008	131	120	0.916	1.092
2009	122	120	0.984	1.017
2010	120	113	0.942	1.062
2011	115	106	0.922	1.085
2012	113	107	0.947	1.056
2013	135	126	0.933	1.071
2014	122	107	0.877	1.140
2015	114	106	0.930	1.075
2016	104	99	0.952	1.051
2017	135	121	0.896	1.116
2018	136	131	0.963	1.038
2019	113	104	0.920	1.087
Total	2075	1935	29.287	32.868

Table 2. Correlation between number of publications and number of authors

		Number of Publications	**Number of Authors**
Number of Publications	Pearson Correlation	1	0.999**
	Sig. (2-tailed)		0.000
	N	31	31
Number of Authors	Pearson Correlation	0.999**	1
	Sig. (2-tailed)	0.000	
	N	31	31

**. Correlation is significant at the 0.01 level (2-tailed).
Mean (Total Papers(TP)=66.94, Total author(TA)=62.42)
Std. Deviation(Total Papers(TP)=53.248, Total author(TA)=49.707)

(Hsiao & Bomhold, 2013)[6]. Contemporary digital libraries face challenges to publish the resources on web using XML enabled metadata sets. Deployment of semantic technologies to digital libraries add better meaning and visualisation to the resources and collections for both users and machines (Raja, Mahmood, Warraich, 2019)[7]. Adopting semantic web technologies to digital libraries is the need of the current generation users. Semantics Webs need to be implemented to digital library domain with the feeling that semantic web has lots of potentials and may change the way information of information access and retrieved (Roy & Arora, 2011)[8]. Semantic web technologies play a significant role in knowledge and content representation of digital library resources which leads to better visibility and retrieval(Prasad &Madalli, 2008)[9]. Semantic and linked-data technologies are heavily oriented toward data reuse and integration. These technologies play a significant role in typical search and retrieval activities (Rico et al., 2019)[10]. Recently, so much attention has been oriented towards semantic web applications, digital library use and usability, organisational and economic issues, as well as legal issues of digital library research(Liew, 2009)[11]. Couto, 2010 explored the effectiveness of classification algorithms and linked information inherent to different document collections based on co-citation and bibliographic coupling (Couto et al., 2010)[12]. Calaresu, 2010 developed a prototype system which assists how the data and associated metadata boost the retrieval efficiency in the context of the Semantic Web (Calaresu& Shiri, 2015)[13]. Ahmad, 2018, explored the digital library research from 2002 to 2016 and calculated annual productivity & citation, highly cited articles, prolific authors, and eminent sources of the subjects and found that year 2016 was the most productive year of publication as well as for the growth rate of citations. The top source journal was the Electronic library. USA was the highest contributions in terms of research output in the field of digital libraries(Ahmad, Jian Ming, & Rafi, 2018)[14]. It has been observed from the literature review as well as documents retrieved through Scopus that no study conducted so far about analyse the authorship pattern and collaborative research in the field of semantic digital libraries. However, some of the studies presented a bibliometric analysis in the field of digital libraries.

2. OBJECTIVE OF THE STUDY

The most significant objectives of the study are:

- To examine correlation matrices and level of collaborative research in the field of semantic digital libraries;
- Evaluate the collaborative authors on SDL using the Degree of Collaboration (DC);
- To visualise authorship correlation network strength and degree of collaboration;
- To understand the keywords clustering map and research trend;and
- Identify prolific authors at the global level and to know about the Average Productivity Per Author (APPA) on SDL.

3. METHODOLOGY AND APPROACH

Datasets for this study were retrieved from the SCOPUS

Table 3. Authorship pattern

	Authors													
Year	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**Total**	**CI**	**CC**	**MCI**
1983	0	0	0	0	0	1	0	0	0	0	1	6.000	0.833	0.000
1984	0	0	1	0	0	0	0	0	0	0	1	3.000	0.667	0.000
1986	0	1	0	0	0	0	0	0	0	0	1	2.000	0.500	0.000
1988	0	0	0	0	0	1	0	0	0	0	1	6.000	0.833	0.000
1989	0	0	0	0	0	1	0	0	0	0	1	6.000	0.833	0.000
1992	0	1	0	0	0	0	0	0	0	0	1	2.000	0.500	0.000
1995	0	0	1	1	0	0	0	0	0	0	2	3.500	0.708	1.417
1996	0	1	2	2	1	1	0	0	0	0	7	3.857	0.710	0.828
1997	3	2	2	3	3	0	0	0	0	0	13	3.077	0.537	0.582
1998	4	5	5	6	3	0	0	0	0	1	24	3.250	0.568	0.593
1999	3	8	4	2	2	0	0	0	0	0	19	2.579	0.514	0.543
2000	2	5	7	5	1	2	0	0	0	0	22	3.182	0.608	0.637
2001	2	12	10	4	3	0	0	2	1	0	34	3.294	0.609	0.627
2002	7	7	7	6	3	1	1	0	1	1	34	3.324	0.545	0.562
2003	7	15	13	5	7	1	3	1	0	0	52	3.192	0.573	0.584
2004	13	23	16	13	4	5	4	0	0	1	79	3.127	0.552	0.559
2005	13	27	30	14	3	5	0	0	0	0	92	2.804	0.550	0.556
2006	11	19	29	17	7	4	1	2	0	0	90	3.178	0.590	0.597
2007	13	23	27	20	12	1	5	0	0	0	101	3.178	0.586	0.592
2008	18	34	32	23	9	2	1	1	0	0	120	2.883	0.552	0.556
2009	11	31	32	22	11	6	5	0	1	1	120	3.350	0.610	0.615
2010	14	28	40	17	8	4	2	0	0	0	113	2.973	0.574	0.579
2011	13	30	22	26	11	2	2	0	0	0	106	3.057	0.579	0.584
2012	17	26	25	24	6	7	1	1	0	0	107	3.056	0.561	0.566
2013	17	32	28	23	13	8	3	0	0	2	126	3.262	0.582	0.587
2014	27	21	29	13	7	7	1	0	0	2	107	2.916	0.502	0.506
2015	13	25	26	25	8	4	1	4	0	0	106	3.245	0.591	0.597
2016	8	25	28	21	13	1	2	1	0	0	99	3.222	0.614	0.620
2017	14	29	24	31	12	6	1	2	1	1	121	3.355	0.601	0.606
2018	20	40	26	25	13	5	2	0	0	0	131	2.954	0.552	0.557
2019	6	23	39	20	7	3	4	1	1	0	104	3.337	0.633	0.639
Total	256	493	505	368	167	78	39	15	5	9	1935	104.153	18.768	15.689

Note: CC(Collaboration Coefficient);MCI(Modified Collaboration Coefficient(MCI);CI(Collaborative Index)

database using key terms pertaining to semantic digital libraries. It was found that the first article on semantic digital libraries was published in 1983. Hence, the time between 1983 and 2019 was fixed for the literature search. In sum, 2075 research documents were published in the field of semantic digital libraries from 1983 to 2019. The extracted bibliographic records (N=2075 publications) were exported and analysed for further investigation using VOSviewer data visualisation tool. Authorship correlation network and keywords cluster map were created for a better understanding of research trend and insights to the emerging areas. We have also used various scientometric parameters such as author productivity, degree

Table 4. Degree of collaboration among authors

Year	Single author(Ns)	Multiple author (Nm)	Total (Nm+Ns)	Degree of collaboration
1983	0	1	1	1.000
1984	0	1	1	1.000
1986	0	1	1	1.000
1988	0	1	1	1.000
1989	0	1	1	1.000
1992	0	1	1	1.000
1995	0	2	2	1.000
1996	0	7	7	1.000
1997	3	10	13	0.769
1998	4	20	24	0.833
1999	3	16	19	0.842
2000	2	20	22	0.909
2001	2	32	34	0.941
2002	7	27	34	0.794
2003	7	45	52	0.865
2004	13	66	79	0.835
2005	13	79	92	0.859
2006	11	79	90	0.878
2007	13	88	101	0.871
2008	18	102	120	0.850
2009	11	109	120	0.908
2010	14	99	113	0.876
2011	13	93	106	0.877
2012	17	90	107	0.841
2013	17	109	126	0.865
2014	27	80	107	0.748
2015	13	93	106	0.877
2016	8	91	99	0.919
2017	14	107	121	0.884
2018	20	111	131	0.847
2019	6	98	104	0.942
Total	256	1679	1935	27.834

of collaboration, and explored statistical techniques for this study.

4. RESULTS AND DISCUSSION

4.1 Author Productivity

The formula given by Yoshikane, Nozawa, Shibui, & Suzuki, 2009[11] is used to calculate the author productivity over time.

$$Average \text{ author per paper} = \frac{Number \text{ of authors}}{Number \text{ of papers}}$$

$$\text{Productivity per author} = \frac{Number \text{ of papers}}{Number \text{ of authors}}$$

Results about author productivity and average author per paper are shown in Table 1. It is revealed that the total average number of authors per paper is 0.945, and the average productivity per author is 1.059. The highest number of author's productivity 136 (1.038) was in year 2018.

4.2 Correlation between Number of Publications and Number of Authors

Pearson correlation analysis was conducted to examine the correlation between number of publications and number of authors. The results inferred a significant and positive relationship ($r = 0.999$, $N = 31$, $p = 0.000$). As p-value is <0.05, it is highly significant. The correlation was strong in strength. A higher number of publications were associated with a higher number of authors (Table 2). It means that that higher number of co-authors are contributed to a higher number of papers.

4.3 Authorship Pattern

From the results presented in Table 3, it is clear that only 12.33 per cent (256) of the total publications are single-authored, whereas remaining 1819 papers are predominantly multi-authored. Hence, it may be concluded that collaborative research has dominated the field of Semantic Digital Libraries.

4.4 Degree of Collaboration

Degree of collaboration among the authors is calculated using following formula:

$$\text{Degree of Collaboration(C)} = \frac{Nm}{Nm + Ns}$$

Where,

C = degree of collaborative authorship among authors
Nm = number of multiple-authored research papers
Ns = number of single-authored research papers

The result indicates that the degree of collaboration ranges from 0.748 to 1.000, and the mean value of collaboration is 0.898. It clearly shows that there are more multi-authored publications and a higher level of collaborative research in the field of semantic digital libraries.

Visualisations of co-authorship networks are given in Fig. 1. Each node is representing the authors and the edges representing a co-authorship. The size of the nodes presents relative frequency in a network structure, and the width of links illustrates the strength of the relationship between each pair.

4.5 Collaboration Index

Collaboration Index(CI) helps to draw the quantitative collaboration patterns. We have used following formula to

calculate CI:

$$\text{Collaborative Index (CI)} = \frac{\sum_{j=1}^{k} fj}{N}$$

Where j = the number of the author(s), fj = the number of j-authored research papers published in the discipline during a certain period, N= the total number of research papers published in the discipline during a certain period of time and K= the greatest number of collaborated authors per paper in a discipline. The mean value of the collaborative index is 3.360 (Table 3), which indicates the better collaboration rate among the authors in the field of semantic digital libraries.

4.6 Collaboration Coefficient and Modified Collaboration Coefficient

The following formula is adopted for calculation of Collaboration Coefficient (CC) :

$$\text{Collaborative Coefficient (CC)} = 1 - \frac{\sum_{j=1}^{k} \left(\frac{1}{j}\right) fj}{N}$$

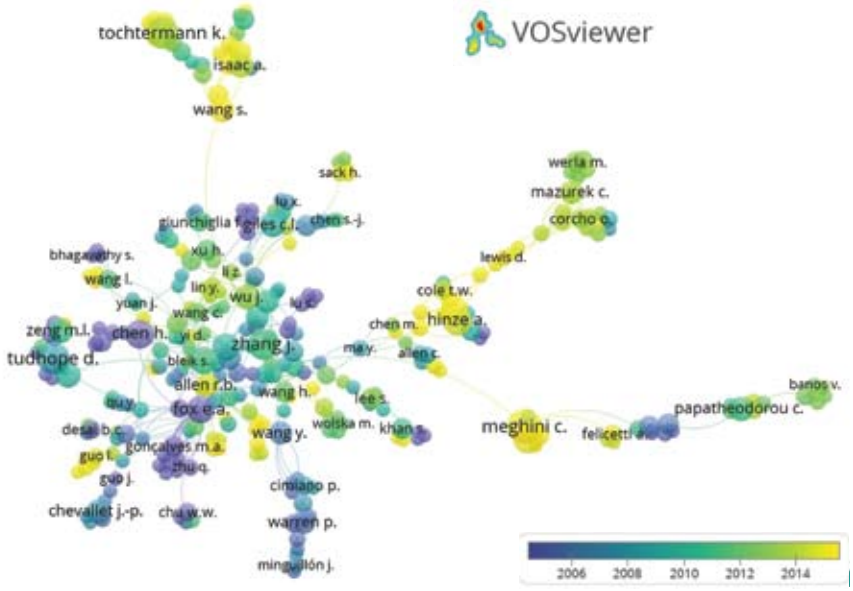

Figure 1. Network visualisation of co-authors.

Where j = authorship, fj = Number of j - authored research papers, N= the total number of research papers and K= the greatest number of authors per paper. The mean value of the collaborative coefficient is 0.605 as per Table 3, which indicates the better collaboration rate among the authors in this field. The value of CC lies between 0 to 1. Modified collaboration coefficient (MCI) values have been calculated. The mean value of MCI is 0.506 (Table3), which is highly significant and represents better authorship collaborations.

4.7 Most Productivity and Highly Cited Author

Top productivity author(based on publication) & highly impacted author(based on citations) in the studied research domain is represented in Fig. 2. The resultant data shows that Meghini C from Institute of Information Science and Technologies, Italy has produced the highest number of research paper(n=18) followed by S.R. Kruk (n=16, Knowledge Hives, Poland) and WT Balke of Institute for Information Systems, Braunschweig, Germany with 14 publications. As far as the most cited author is concerned, it has been observed that Egenhofer MJ has found to highly impact author with 851 citations followed by Rodrguez MA(690) and Jung JJ(358).

Collaboration patterns of these prolific authors are better visualised through a Sankey chart diagram which was constructed with parameter author-country-keywords. Figure 3 shows the most prolific authors with their affiliating countries and studied the research domain.

5. KEYWORD MAPPING

An analysis of co-occurrence of keywords was constructed using VOSviewer.Out of 8757 keywords, we have chosen those keywords whose minimum occurrence is 20. In total, 113 keywords are qualified for the network and grouped into five clusters. The size of the bubble is as per the strength of the keyword in terms of their frequency, association and influence. The top key terms having the largest total link strength were as follows: digital libraries

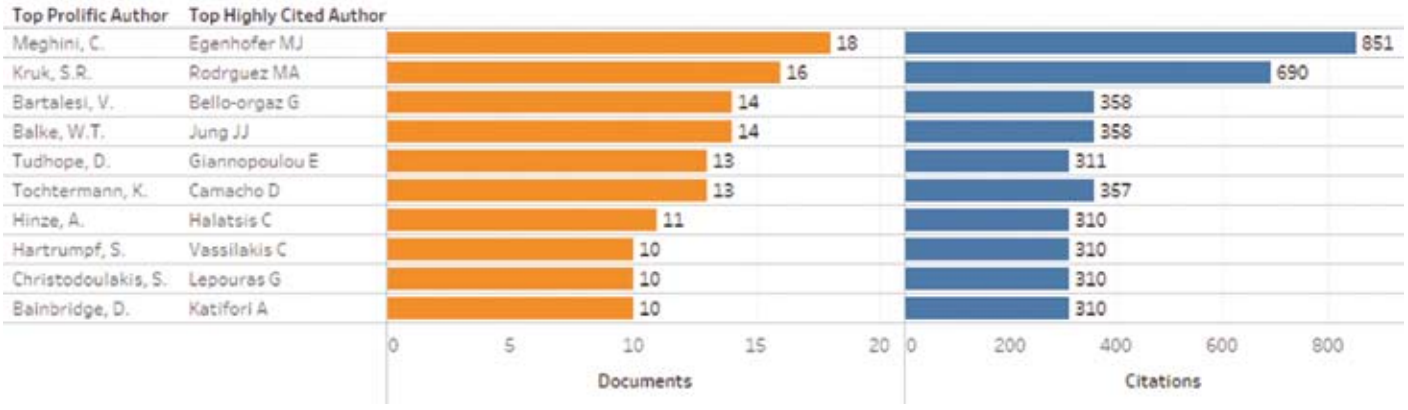

Figure 2. Top prolific author(based on publication) & top highly impacted author(based on citations) in the studied research domain.

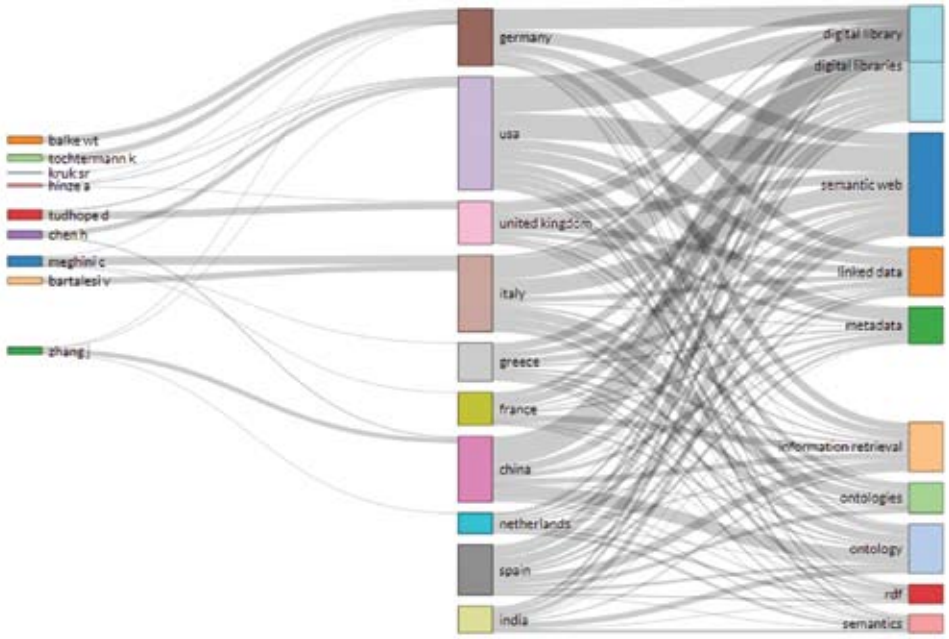

Figure 3. Most prolific authors in the studied research domain.

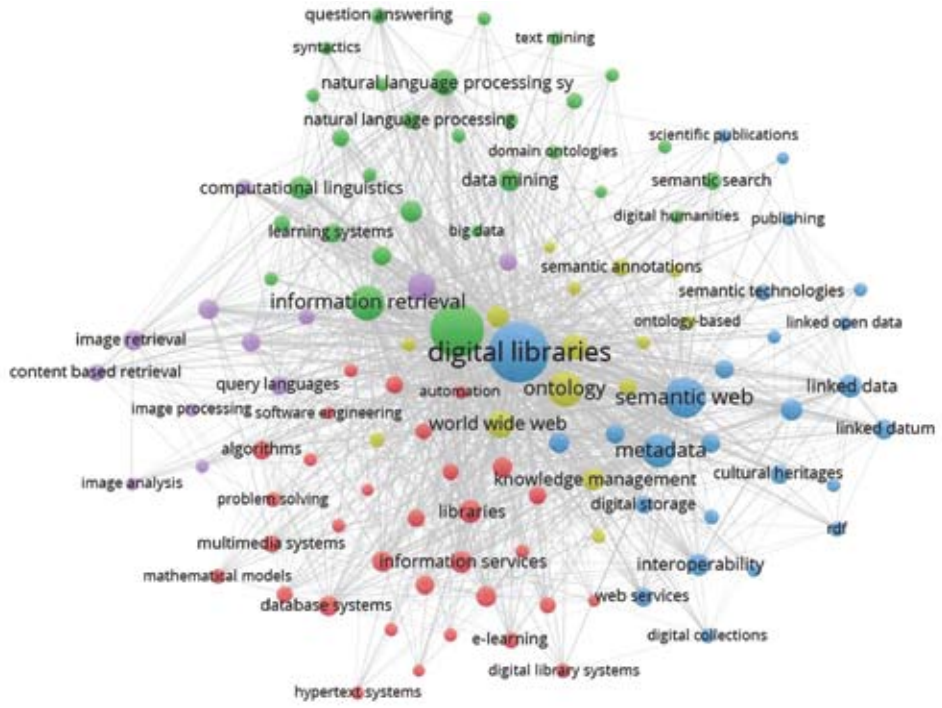

Figure 4. Keywords clustering map.

(frequency = 1453, TLS = 6846 times), semantics (frequency = 1069, TLS = 5036 times), semantic web (frequency = 493, TLS = 2602 times), information retrieval (frequency = 321, TLS = 2003 times), ontology (frequency = 308, TLS = 1795 times). These most repeated keywords are positioned in the central area of the network, which indicates their influences. As depicted in Fig. 4 of visualisation network; cluster third, first, four and second are positioned close to each other in the visualisation network. At the same time, the fifth cluster is a little far away. It shows the close association between the keywords in these four clusters compared with cluster five. Results also reveal that the focus areas of research related to SDL include digital libraries, semantic web, ontology, metadata, information retrieval. However, keywords such as natural language processing system, computational linguistics, linked data are also repeated frequently in the published literature thus revealing the emerging areas of the future research in the domain of SDL

6. CONCLUSIONS

In scientometrics based research, the publication trend can better visualised through authors collaborations. The study provides a complete view of authorship and collaborative pattern of global semantic digital libraries research. Results and findings of the study will not only help researchers to know about the collaborative authorship pattern and correlation matrics but also establish directions for future research in the untouched domain of SDL. Study finds that the average degree of collaboration (DC) is 0.898, which shows there are more multi-authored publications and a higher level of collaborative research in the field of semantic digital libraries. Further, the collaborative coefficient (CC) is 0.605, which indicates the better collaboration rate among the authors in this field. The current data show positive correlation and authorship pattern upright SDL research productivity in the past 38 years. The results show that the focus of researchers is digital libraries, semantic web, ontology, metadata, information retrieval. However, keywords such as natural language processing system, computational linguistic, linked data are also repeated frequently in the published literatures, which reveals the emerging areas of future research in the domain of semantic digital libraries.

REFERENCES

1. Macgregor, G. Introduction to a special issue on digital libraries and the semantic web: Context, applications and research. *Libr. Rev.*, 2008, **57**(3), 173–77. doi: 10.1108/00242530810865457.
2. Sure, Y. & Studer, R. Semantic Web technologies for digital libraries. *Libr. Manage.*, 2005., **26**(4/5), 190–95. doi: 10.1108/01435120510596044.

3. Burke, M. The semantic web and the digital library. Aslib Proceedings: New Information Perspectives., 2009, **61**(3), 316–22. .
doi: 10.1108/00012530910959844.
4. Kruk, S. Semantic digital libraries: Improving usability of information discovery with semantic and social services. Lulu, 2010.
5. Sure, Y. & Studer, R. Semantic web technologies for digital libraries. *Libr. Manage.*, 2005, **26**(4/5), 190–95.
doi: 10.1108/01435120510596044.
6. Hsiao, K. & Bomhold, C.R. Library Hi Tech news article information, *Libr. Hi. Tech.*, 2013, **32**(3), 409–22.
doi: 10.1108/LHTN-04-2015-0024.
7. Raza, Z.; Mahmood, K. & Warraich, N. Application of linked data technologies in digital libraries: A review of literature, *Libr. Hi. Tech. News*, 2019, **36**(3), 9-12.
doi: 10.1108/LHTN-10-2018-0067.
8. Roy, P. & Arora, D. Social semantic digital library: The future. *DESIDOC J. Libr. Inf. Technol.*, 2011, **31**(4), 226-33.
doi: 10.14429/djlit.31.4.1101.
9. Prasad, A. & Madalli, D. Faceted infrastructure for semantic digital libraries. *Libr. Rev.*, 2008, **57**(3), 225-34.
doi: 10.1108/00242530810865493.
10. Rico, M.; Vila-Suero, D.; Botezan, I. & Gómez-Pérez, A. Evaluating the impact of semantic technologies on bibliographic systems: A user-centred and comparative approach. *Jr. Web Semantics*,2019, **59**, 100-500.
doi: 10.1016/j.websem.2019.03.001.
11. Liew, C.L. Digital library research 1997-2007: Organisational and people issues. *J. Doc.*, 2009, **65**(2), 245–66.
doi: 10.1108/00220410910937606.
12. Couto, T.; Ziviani, N.; Calado, P.; Cristo, M.; Gonçalves, M.; de Moura, E.S. & Brandão, W. Classifying documents with link-based bibliometric measures. *In. Inf. Retr.,* **13**, 2010.
doi: 10.1007/s10791-009-9119-7.
13. Calaresu, M. & Shiri, A. Understanding semantic web: A conceptual model. *Lib. Rev.*, 2015, **64**(1), 82–100.
doi: 10.1108/LR-09-2014-0097.
14. Ahmad, K.; Jian Ming, Z. & Rafi, M. Assessing the digital library research output: Bibliometric analysis from 2002 to 2016. *Ele. Lib.*, 2018, **36**(4), 696–704.
doi: 10.1108/EL-02-2017-0036.
15. Yoshikane, F.; Nozawa, T.; Shibui, S. & Suzuki, T. An analysis of the connection between researchers' productivity and their co-authors' past attributions, including the importance in collaboration networks. *Scientometrics.*, 2009, **79**(2), 435–449.
doi: 10.1007/s11192-008-0429-8.

CONTRIBUTORS

Dr Shriram Pandey is currently working as Asst. Professor at the Department of Library and Information Science, Banaras Hindu University, Varanasi. He is the recipient of the prestigious Commonwealth Professional Fellowship at the University of East London, UK. UGC-Junior Research Fellowship, and BHU-Gold Medal. His area of interest includes semantic web, digital libraries &, Scientometric.
His individual contribution in the current study: Conceptualisation, data retrieval and tabulation and formal analysis;

Mr Sidhartha Sahoo is currently working as Scientific Officer at Knowledge Resource Centre (KRC), Indian National Centre for Ocean Information Services (INCOIS), Ministry of Earth Sciences, Government of India, Hyderabad.
His individual contribution to the current study is validation, visualisation, review & editing.