

A Keyword Based Scientometric Analysis of Semantic Web Research in Tourism

Shiv Shakti Ghosh*, Sunil Kumar Chatterjee

Department of Library and Information Science Jadavpur University, Jadavpur Kolkata, West Bengal, INDIA.

ABSTRACT

The economy of many nations sustain on tourism which is an information intensive industry and requires semantic handling of data. Semantic web research is an area that deals with technologies for semantic handling of data. In this study, the results of an investigation on the extent of semantic web research in the field of tourism have been presented. For this purpose, data was retrieved from the Scopus database. Scientometric methods that help in keyword-based analysis were applied for analyzing the data using the R software environment.

Keywords: Tourism, Semantic Web, Scientometrics, Conceptual structure, Thematic evolution.

Correspondence

Shiv Shakti Ghosh

Department of Library and Information Science Jadavpur University, Jadavpur Kolkata-700032, West Bengal, INDIA.
Email id: shivu@drct.isibang.ac.in
ORCID ID: 0000-0001-9850-8344

Received: 09-06-2021;

Revised: 27-10-2021;

Accepted: 03-05-2022.

DOI: 10.5530/jsires.11.1.12

INTRODUCTION

According to the United Nations World Tourism Organization (UNWTO),^[1] "Tourism is a social, cultural and economic phenomenon which entails the movement of people to countries or places outside their usual environment for personal or business/professional purposes. These people are called visitors (which may be either tourists or excursionists; residents or non-residents) and tourism has to do with their activities, some of which involve tourism expenditure." Quite naturally, tourism is a data intensive industry that requires semantic handling of data. Now, Semantic Web technologies help in creating Web data stores, building vocabularies and writing rules for data handling.^[2] Thus, applications of Semantic web technologies have made inroads into the tourism domain. To understand the state of semantic web research in the tourism domain, quantitative analysis of scholarly publication must be done. This is because quantitative analysis helps in predicting research topics at a macro level.^[3]

So, to get an understanding of the semantic web research scenario in the tourism domain, keyword based scientometric analysis of research publications on the domain was done. A scientometric study involves quantitative methods for studying the development of science.^[4-5] The results of the analysis have been presented in this study while illustrating about important

implications about the domain under consideration. In the following parts, the section titled Related work discusses about the different purposes for applying bibliometric/scientometric methods and associated relevant results from some studies that are mostly related to 'Tourism' or 'Semantic Web'; the section titled Data and Methodology discusses the processes by which data was collected and analyzed; the section titled Analysis discusses the observations made from the analysis of the data and the Conclusion section concludes the discussion while discussing about the implications of the presented study.

LITERATURE REVIEW

Garfield^[6] has traced the evolution of Scientometrics while discussing about the contributions of Nalimov and Price. It is one of the earliest studies that had demonstrated the process of visualizing a domain using software designed for bibliometric analysis and has influenced several similar studies. Shtovba and Petrychko^[7] have demonstrated the use of Jaccard index for assessing similarity of research areas and the process of identification of influential pairs of research areas. Eigenfactor,^[8] the Audience factor,^[9] *h*-index,^[10] local citation score^[11] are some of the indicators that are used in a scientometric analysis for getting overview of a domain under consideration. There are many studies like Vazquez *et al.*^[12] that have presented scientometric analysis of literature on application of domains like artificial intelligence on other domains. Liu *et al.*^[13] have done scientometric analysis to predict topics and trends related to tourism forecasting and have found an upward trend of research output in the domain with tourism demand related models being the hot topics. Sharafuddin and Madhavan^[14] have used a combination of scientometrics, citation-based and theme-based systematic

Copyright

© The Author(s). 2022 This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

literature review for assessing the research patterns and thematic evolution of research on blue tourism. The study notes a huge research growth in blue tourism and also sheds light on the conceptual relationship in the domain. Zhu *et al.*^[15] have used co-word analysis procedures including keyword clustering and strategic analysis for mapping the research progress on ontology research and have also proposed the disciplinary incidence index. The study shows that ‘applications in semantic web research’ is among the frequently used keywords in the domain. Fang *et al.*^[16] have presented the results of a scientometric analysis of publications climate change and tourism that the domain has become multi-disciplinary with a huge growth in publications. By doing a scientometric study of publications about cruise tourism, Vega-Muñoz *et al.*^[17] have identified important research areas related to cruise tourism. A knowledge map for pro-poor tourism, built using a scientometric approach has been presented by Yu *et al.*^[18] Qian *et al.*^[19] have studied co-occurrence and collaboration networks and presented a scientometric review of research on travel websites thereby identifying the research hotspots of the domain and shedding light on the importance of the websites. Zhang *et al.*^[20] have used semantic network analysis of subjects and social network analysis of networks together for mapping the knowledge domain of China’s smart tourism research and revealing the research trends. Influential institutes and scholars related to the smart tourism research have been identified. Johnson and Samakovlis^[21] have examined the development of concepts belonging to smart tourism knowledge through collaborative networks.

Ding’s^[22] study presents a bibliometric analysis of publications on Semantic Web retrieved from Scopus and Web of Science. Niknia and Mirtaheri^[23] have employed cluster analysis, network analysis, co-word analysis to study publications on Linked Data retrieved from Scopus. The study has revealed the new concepts and also the research trend in the domain after the contributions from librarians and information scientists. Mika, *et al.*^[24] have showed in their study on social network analysis of the sciences that the use of semantic web technologies can be beneficial for dealing with heterogeneity of data. Gandon^[25] has identified the main research trends by conducting a survey of the research field of Semantic Web, Linked Data and Web of Data. Li *et al.*^[3] have presented a bibliometric analysis of geo-ontology research with interesting visual analysis of research performance revealing the patterns of collaboration. An analysis of keywords, term words has been also presented. Bhattacharya^[26] has used co-word analysis and links between keywords in a bibliometric analysis of the machine learning domain. St-Germain and Mongeon^[27] have assessed the contribution of information science discipline in the field of Semantic Web research by studying the evolution of publications and discussion of topics

within different disciplines while also carrying out citation analysis.

DATA AND METHODOLOGY

The Semantic Web has evolved from the field of Artificial Intelligence, and semantics and ontology related research constitutes the core part of the Semantic Web research field.^[22] Though ontological research had started from the 1980s, research on Semantic Web gained pace from late 1990s due to the support of the European Commission and United States.^[28] In 2001, Tim Berners-Lee had indicated that the Semantic Web^[29] is the harbinger of the next wave of transformation through enabling content-based access interoperability. So, for the current study documents published after 2001 till 2020 were considered. Though some of the terms for emerging concepts of that time related to semantic web, for example SPARQL, RDF, OWL, OIL, DAML+OIL etc. Were not included in the search query for Ding’s study in 2010, some of them were included in the search query for the demonstrated study since they are quite popular now. But also some were avoided because it has been also noted by Ding that noisy data will be retrieved by the usage of the terms like OWL, OIL and DAML in the search query. The search was restricted to areas related to computer science and engineering. Conference reviews, editorial materials were excluded. The results were downloaded from Scopus on May 21, 2021. The results were reviewed manually and some articles were filtered based on a strategy similar to the strategy of Yu *et al.*^[18] for filtering items from the corpus of study. Items were filtered when the words used in the search query appeared only in the keywords provided by Scopus or if the abstract did not have any relationship with semantic web. The analysis was done using the Bibliometrix package^[30] for R.^[31] Information about the downloaded data has been presented in Table 1. The annual scientific production has been illustrated in Figure 1. It can be seen that scientific production in the domain has increased from 2004 onwards.

Table 1: Metadata about the data in the corpus.

Attribute	Value
Time span	2002: 2020
Sources	343
No. of documents	621
Average citations/document	8.209
Average citations/year/document	0.9352
References	13922
Scopus assigned keywords	3270
Author’s keywords	1372

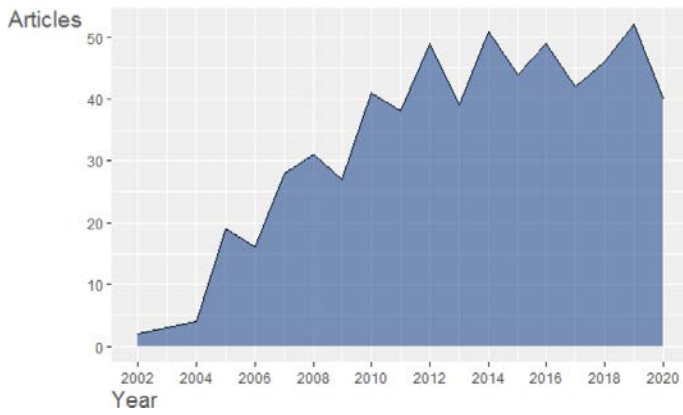


Figure 1: Annual scientific production in the domain of study plotted using data in the corpus.

Table 2: Authorship related information about the documents in the corpus.

Attribute	Value	Attribute	Value
Authors	1599	Single-authored documents	46
Author Appearances	2164	Documents per Author	0.388
Authors of single-authored documents	41	Authors per Document	2.57
Authors of multi-authored documents	1558	Co-Authors per Documents	3.48

RESULTS AND ANALYSIS

The authorship related information about the documents was analyzed to get an overview about collaboration in the domain. 92.6% documents in the corpus are multi-authored. So, it can be said that research in the field under consideration is mostly collaborative. Further, using the information from Table 2, the value of Collaboration Index^[30-32] is found to be 2.71, that is determined by dividing the total number of authors of multi-authored documents by the total number of multi-authored documents. Thus, in the field under consideration collaborative works mostly have a team of 2 to 3 authors per multi-authored document. Figure 2 shows the network^[30,32] formed by collaborations^[33] between countries. According to the degree of the vertex in the network, Italy, Spain, China, France, Germany are among the top countries with collaborations. Also, Italy and Spain are the top two countries in terms of total number of citations received.

Tracking the changes in research can help researchers to gain insights about the evolution of a research domain.^[16] Research status and development trends can be studied using keywords.^[3] The rest of the analysis is based on keywords associated with the articles. Both the Authors' keywords and Scopus assigned keywords were used for the study. Table 3 presents the top ten Authors' keywords and Scopus assigned keywords associated with the documents in the corpus. Since, keywords indicate a concise label to a document, keyword co-occurrence can

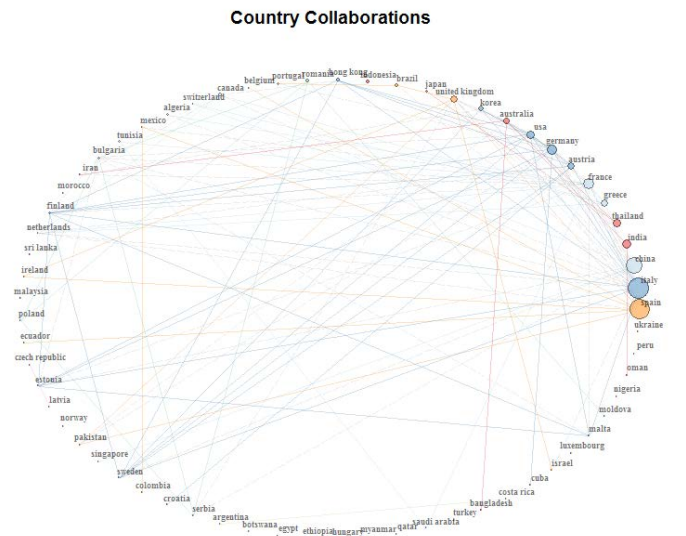


Figure 2: Country collaborations calculated from the data in the corpus.

Table 3: List of most used keywords by the documents in the corpus.

Author Keywords	Articles	Scopus assigned keywords	Articles
Ontology	150	Ontology	308
Semantic web	107	Semantic web	214
Tourism	76	Semantics	199
Ontologies	49	Tourism	75
Linked data	31	Recommender systems	60
E-tourism	25	Information retrieval	57
Information retrieval	20	Knowledge based systems	55
Linked open data	19	Web services	50
Recommender systems	18	Artificial intelligence	49
Cultural heritage	15	Search engines	49

effectively point out emerging topics of research.^[16] So, the network formed by co-occurrence of Scopus assigned keywords has been presented in Figure 3. The size of the nodes in the network correlate with the top Scopus assigned keywords list.

Figure 4 displays the keyword growth graph constructed from the Scopus assigned keywords to the documents. The graph clearly shows the high usage of the keywords 'ontology' and 'semantic web' among the other keywords by the documents in the corpus. To gain a better understanding about the usage of the keywords, the formulas for calculating Popularity Index^[34] and the Promising Index^[34] of entities proposed by Li *et al.*^[34] has been applied on selected keywords. Li *et al.*^[34] Popularity index helps in determining the percentage of publications discussing an entity among all publications within a specific period and the Promising index helps in determining the change in popularity of an entity in the research domain between two continuous periods. According

integration, linked open data, ontology mapping etc. And the third big cluster is made up of co-occurrence of concepts like taxonomies, named entities, machine learning techniques, question answering systems etc. By their size, these clusters symbolize about the areas where research in the domain under consideration is focused. Articles that mention words like context, intelligent systems, information sources, knowledge graph (belonging to different cluster than the previous words) are close to the middle of the graph and represent the centre of the research field.

Next the thematic evolution^[37] of the field was studied. The year breaks in the time span 2002–2020 for analyzing the thematic evolution of the graphs were at 2005, 2010 and 2015. The maps based on thematic evolution using the author-based keywords have been presented in Figure 6a, 6b, 6c, 6d. During 2002–2005 (see Figure 6a), semantic web was a generic theme in the research field. Silva and Rocha's^[38] and Chiu and Leung's^[39] works are among the top cited articles during this period. In the period 2006–2010 (Figure 6b), research related

to recommendation systems, twitter-based studies targeted at opinion mining were very well developed. Cambria *et al.*^[40] work is one of the top cited works of this period. Semantic web continued as a generic theme. Some other generic themes were ontology construction, context modeling. During 2011–2015 (see Figure 6c) cultural heritage research was among the developed themes. Goy *et al.*^[41] O'Keefe and Benyon's.^[42] Fermo *et al.*^[43] works are the notable works of this period. During 2016–2020 (see Figure 6d) e-tourism, recommender systems, big data etc. were among the generic themes. Chu *et al.*^[44] and Shi *et al.*^[45] works are some of the works related to the generic themes of this period.

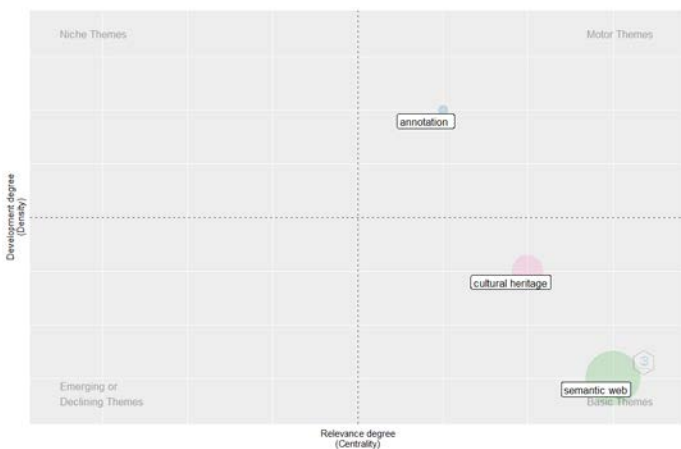


Figure 6a: Thematic evolution of the research field during 2002-2005 based on the documents in the corpus.

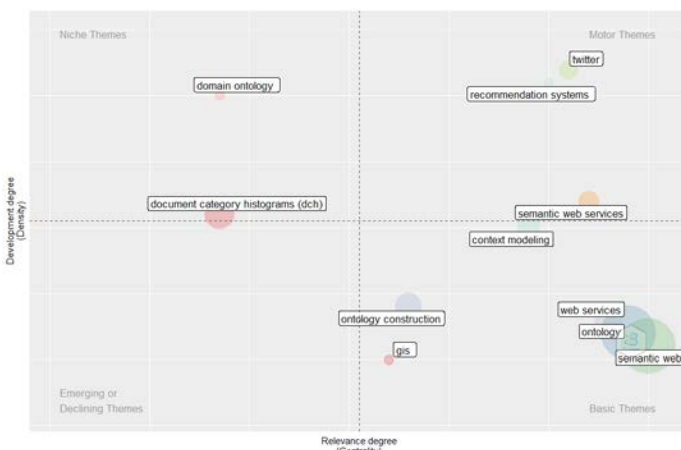


Figure 6b: Thematic evolution of the research field during 2006-2010 based on the documents in the corpus.

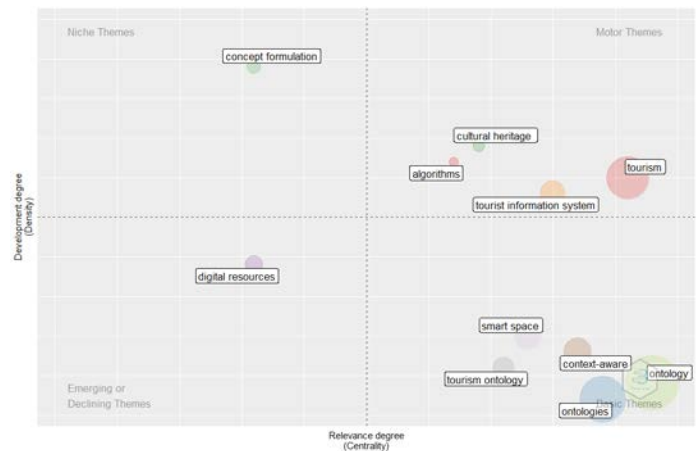


Figure 6c: Thematic evolution of the research field during 2011-2015 based on the documents in the corpus.

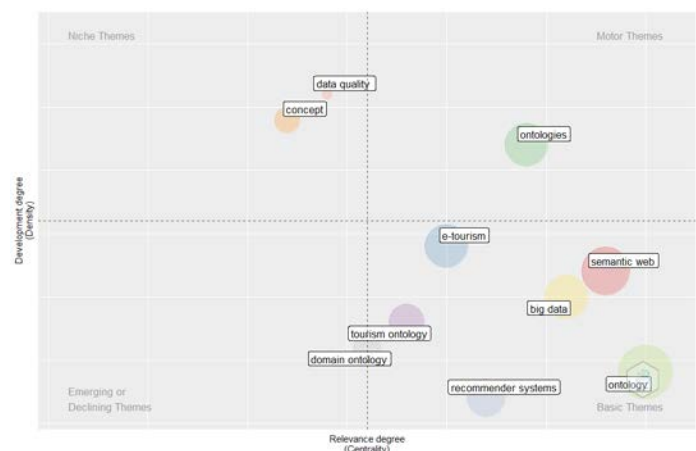


Figure 6d: Thematic evolution of the research field during 2016-2020 based on the documents in the corpus.

CONCLUSION

The study includes a review of selected scientometric studies on tourism and semantic web research. Building upon the reviewed methodologies the study presents some important findings by analyzing data about studies on the applications of

semantic web technologies in the tourism domain. The study highlights countries based on collaboration networks and also points out important areas of research based on keyword-based analysis and thus, can clearly provide a directional motivation to researchers intending to apply semantic web technologies in the tourism domain. The study clearly shows that semantic web started as a generic theme in the domain but specific semantic web technologies gained prominence in the later years. Research in the domain was found mostly to be collaborative. Intelligent systems, context, location-based services, point of interest, data integration, linked open data, ontology mapping, knowledge graph, taxonomies, machine learning techniques, question answering systems, etc. are some of the concepts around which research clusters can be found in the studied domain. It was found that major sources of the documents included in the corpus of the present study are Lecture Notes in Computer Science, CEUR Workshop Proceedings, ACM International Conference Proceeding Series, Communications in Computer and Information Science. This means that maximum documents are conference publications. A limitation of this study is that it presents the results from an analysis done over data collected from only one database. If data collected from other citation and indexing databases are also studied, then a better understanding of the research scenario of the domain under consideration may be obtained.

CONFLICT OF INTEREST

The authors declare that there is no conflict of interest.

REFERENCES

1. Unwto.org, UNWTO [cited May 27 2021]. Available from: <https://www.unwto.org/glossary-tourism-terms>.
2. W3.org. Semantic Web - W3C. [cited 2021 2 June]. Available from: <https://www.w3.org/standards/semanticweb>
3. Li L, Liu Y, Zhu H, Ying S, Luo Q, Luo H, et al. A bibliometric and visual analysis of global geo-ontology research. *Computers and Geosciences*. 2017;99:1-8. doi: 10.1016/j.cageo.2016.10.006.
4. Leydesdorff L, Milojevic S. *Scientometrics*. In: Lynch, M. editor, *International Encyclopedia of Social and Behavioral Sciences*. New York: Elsevier; 2015.
5. Mingers J, Leydesdorff L. A review of theory and practice in scientometrics. *European Journal of Operational Research*. 2015;246(1):1-19. doi: 10.1016/j.ejor.2015.04.002.
6. Garfield E. From the science of science to Scientometrics visualizing the history of science with HistCite software. *Journal of Informetrics*. 2009;3(3):173-9. doi: 10.1016/j.joi.2009.03.009.
7. Shtovba S, Petrychko M. Jaccard index-based assessing the similarity of research fields in dimensions. *DCSmart*. 2019:117-28.
8. Bergstrom CT, West JD, Wiseman MA. The Eigenfactor metrics. *Journal of neuroscience*. 2008;28(45):11433-4. doi: 10.1523/JNEUROSCI.0003-08.2008, PMID 18987179.
9. Zitt M, Small H. Modifying the journal impact factor by fractional citation weighting: The audience factor. *Journal of the American Society for Information Science and Technology*. 2008;59(11):1856-60. doi: 10.1002/asi.20880.
10. Hirsch JE. An index to quantify an individual's scientific research output. *Proceedings of the National Academy of Sciences*. 2005;102(46):16569-72. doi: 10.1073/pnas.0507655102, PMID 16275915.
11. Garfield E, Pudovkin AI, Istomin VS. Why do we need algorithmic historiography? *Journal of the American Society for Information Science and Technology*. 2003;54(5):400-12. doi: 10.1002/asi.10226.
12. Vazquez JP, Torres RS, Perez DB, Demarigny Y, Soldat V, Gemelas L, Panagiotakos DB, Bersimis FG. Scientometric analysis of the application of artificial

- intelligence in agriculture. *Journal of Scientometric Research*. 2021;10(1):55-62. doi: 10.5530/jscires.10.1.7.
13. Liu H, Liu Y, Wang Y, Pan C. Hot topics and emerging trends in tourism forecasting research: A scientometric review. *Tourism Economics*. 2019;25(3):448-68. doi: 10.1177/1354816618810564.
14. Sharafuddin MA, Madhavan M. Thematic evolution of blue tourism: A scientometric analysis and systematic review. *Global Business Review*. 2020. doi: 10.1177/0972150920966885.
15. Zhu Q, Kong X, Hong S, Li J, He Z. Global ontology research progress: A bibliometric analysis. *Aslib Journal of Information Management*. 2015;67(1):27-54. doi: 10.1108/AJIM-05-2014-0061.
16. Fang Y, Yin J, Wu B. Climate change and tourism: A scientometric analysis using CiteSpace. *Journal of Sustainable Tourism*. 2018;26(1):108-26. doi: 10.1080/09669582.2017.1329310.
17. Vega-Muñoz A, Arjona-Fuentes JM, Ariza-Montes A, Han H, Law R. In search of 'a research front' in cruise tourism studies. *International Journal of Hospitality Management*. 2020;85:102353. doi: 10.1016/j.ijhm.2019.102353.
18. Yu L, Wang G, Marcouiller DW. A scientometric review of pro-poor tourism research: visualization and analysis. *Tourism Management Perspectives*. 2019;30:75-88. doi: 10.1016/j.tmp.2019.02.005.
19. Qian J, Law R, Wei J. Knowledge mapping in travel website studies: a scientometric review. *Scandinavian Journal of Hospitality and Tourism*. 2019;19(2):192-209. doi: 10.1080/15022250.2018.1526113.
20. Zhang Q, Wang Q, Hao J-X, Yu Y. Mapping smart tourism research in China: A semantic and social network analysis using CiteSpace 13th International Conference on Service Systems and Service Management (ICSSSM). 2016; 2016. doi: 10.1109/icsssm.2016.7538622.
21. Johnson AG, Samakovlis I. A bibliometric analysis of knowledge development in smart tourism research. *Journal of Hospitality and Tourism Technology*. 2019;10(4):600-23. doi: 10.1108/JHTT-07-2018-0065.
22. Ding Y. Semantic Web: who is who in the field — a bibliometric analysis. *Journal of Information Science*. 2010;36(3):335-56. doi: 10.1177/0165551510365295.
23. Niknia M, Mirtaheri SL. Mapping a decade of linked data progress through co-word analysis. *Webology*. 2015;12(2):1-14.
24. Mika P, Elfiring T, Groenewegen P. Application of semantic technology for social network analysis in the sciences. *Scientometrics*. 2006;68(1):3-27. doi: 10.1007/s11192-006-0081-5.
25. Gandon F. A survey of the first 20 years of research on semantic Web and linked data. *Ingénierie des systèmes d'information*. 2018;23(3-4):11-38. doi: 10.3166/isi.23.3-4.11-38.
26. Bhattacharya S. Some salient aspects of machine learning research: A bibliometric analysis. *Journal of Scientometric Research*. 2019;8(2s):s85-92. doi: 10.5530/jscires.8.2.26.
27. St-Germain M, Mongeon P. The contribution of information science in the Semantic Web research landscape. *Proceedings of the Association for Information Science and Technology*. 2018;55(1):470-7. doi: 10.1002/pr2.2018.14505501051.
28. Shadbolt N, Berners-Lee T, Hall W. The semantic web revisited. *IEEE Intelligent Systems*. 2006 May 30;21(3):96-101. doi: 10.1109/MIS.2006.62.
29. Berners-Lee T, Hendler J, Lassila O. The semantic web. *Scientific American*. 2001 May 17;284(5):34-43. doi: 10.1038/scientificamerican0501-34.
30. Aria M, Cuccurullo C. Bibliometrix: an R-tool for comprehensive science mapping analysis. *Journal of Informetrics*. 2017;11(4):959-75. doi: 10.1016/j.joi.2017.08.007.
31. R Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing; 2020. Available from: <https://www.R-project.org/> [cited 5/6/2022].
32. Elango B, Rajendran P. Authorship trends and collaboration pattern in the marine sciences literature: A scientometric study. *International Journal of Information Dissemination and Technology*. 2012;2(3):166.
33. Batagelj V, Cerinšek M. On bibliographic networks. *Scientometrics*. 2013;96(3):845-64. doi: 10.1007/s11192-012-0940-1.
34. Li X, Rousseau JF, Ding Y, Song M, Lu W. Understanding drug repurposing from the perspective of biomedical entities and their evolution: Bibliographic research using aspirin. *JMIR Medical Informatics*. 2020;8(6):e16739. doi: 10.2196/16739, PMID 32543442.
35. Berners-Lee T. Linked data-design issues. p. w3; 2006. [org/DesignIssues/LinkedData.HTML](http://www.org/DesignIssues/LinkedData.HTML). [cited Jun 2 2021]. Available from: <http://www.org/DesignIssues/LinkedData.HTML>.
36. Cuccurullo C, Aria M, Sarto F. Foundations and trends in performance management: A twenty-five years bibliometric analysis in business and public administration domains. *Scientometrics*. 2016;108(2):595-611. doi: 10.1007/s11192-016-1948-8.
37. Cobo MJ, López-Herrera AG, Herrera-Viedma E, Herrera F. An approach for detecting, quantifying, and visualizing the evolution of a research field: A practical application to the fuzzy sets theory field. *Journal of informetrics*. 2011;5(1):146-66. doi: 10.1016/j.joi.2010.10.002.
38. Silva N, Rocha J. Semantic web complex ontology mapping. *Web Intelligence*. 2003:82-8.

39. Chiu DK, Leung HF. Towards ubiquitous tourist service coordination and integration: A multi-agent and semantic web approach. In: Proceedings of the 7th international conference on Electronic Commerce; 2005. p. 574-81.
40. Cambria E, Speer R, Havasi C, Hussain A. Senticnet: A publicly available semantic resource for opinion mining. In: AAAI fall symposium: Commonsense Knowledge; 2010.
41. Goy A, Magro D, Rovera M. Ontologies and historical archives: A way to tell new stories. *Applied Ontology*. 2015;10(3-4):331-8. doi: 10.3233/AO-150152.
42. O'Keefe B, Benyon D. Using the blended spaces framework to design heritage stories with schoolchildren. *International Journal of Child-Computer Interaction*. 2015;6:7-16. doi: 10.1016/j.ijcci.2016.02.001.
43. Feroso AM, Mateos M, Beato ME, Berjón R. Open linked data and mobile devices as e-tourism tools. A practical approach to collaborative e-learning. *Computers in Human Behavior*. 2015;51:618-26. doi: 10.1016/j.chb.2015.02.032.
44. Chu Y, Wang H, Zheng L, Wang Z, Tan KL. TRSO: A tourism recommender system based on ontology. *Lecture Notes in Computer Science International Conference on Knowledge Science*. 2016:567-79. doi: 10.1007/978-3-319-47650-6_45.
45. Shi M, Zhu W, Yang H, Li C. Applying semantic web and big data techniques to construct a balance model referring to stakeholders of tourism intangible cultural heritage. *International Journal of Computer Applications in Technology*. 2016;54(3):192-200. doi: 10.1504/IJCAT.2016.079873.