
Knowledge Graph Construction from Library and Information Science Journal Articles

Partha Sarathi Mandal*, Sukumar Mandal**

ARTICLE INFO

Article history:
Online First 29 October 2025

Keywords:
Knowledge Graph,
Library and Information
Science,
Natural Language Processing,
Information Retrieval,
Semantic Search

ABSTRACT

Knowledge graphs (KGs) emerge as potential tools for information access and resource discovery in a structured format. It facilitates information retrieval, data integration, and semantic reasoning. Considering the rapid growth of literature publications, a high-relevance search is necessary for researchers and practitioners. There are numerous tools for data organization and knowledge extraction. The knowledge graph is one of them, which depicts structured information with nodes and relationships. Library and information science can help build KGs to grasp the intricate relationships between scholarly works, their authors, institutions, and topics. The knowledge graph technology produces more relevant search results, which makes it easy to explore accurately. This paper examines the construction of knowledge graphs from library and Information Science (LIS) journal articles. A systematic approach is followed to extract entities, relationships, and attributes from LIS literature. The accuracy of the constructed knowledge graph is 89.47% (Recall), 94.44% (Precision), and 91.80% (F1 Score). User satisfaction is 85% in rating their satisfaction with the search results, interface usability, and ease of exploring relationships between entities regarding Scopus and Google Scholar. This paper also discusses this research's potential areas and challenges in enhancing information organization and retrieval in the LIS domain.

1. Introduction

Knowledge graphs capture entities and their interrelations in a structured format. They are widely used in various domains, such as search engines, recommendation systems, and artificial intelligence applications. The best example is Google's Knowledge Graph, which enhances search results by providing contextual information. LIS is a multidisciplinary field that focuses on the organization, management, and dissemination of information. This Journal articles cover information retrieval, knowledge management, bibliometrics, and digital libraries. Building KGs from these articles can

* Research Scholar, Department of Library and Information Science The University of Burdwan, Burdwan, 713104, (mandalpsm@gmail.com) (First & Corresponding Author)

** Assistant Professor, Department of Library and Information Science The University of Burdwan, Burdwan, 713104, (sukumar.mandal5@gmail.com) (Co-Author)

facilitate better understanding and navigation of the vast body of knowledge within LIS. The growth of library and information science literature has increased day by day. They are breaking boundaries in digital libraries, focusing on user behavior and information retrieval. It becomes challenging for researchers and Library and Information Science (LIS) practitioners to access relevant information efficiently. In such cases, it is understandable that researchers and practitioners have sifted through a mountain of literature like Scopus and Web of Science. They carry out keyword-based search logic, which is inadequate in LIS because it tends to overload unrelated and remotely related keywords. In this regard, the emergence of graph technologies specifies the domain to increase the effectiveness of literature search in terms of comprehension of relational context among various concepts. The present paper explores the methodologies for building knowledge graphs from LIS journal articles to enable enhanced information retrieval and knowledge discovery.

2. Background and Related Work

Knowledge graphs (KGs) have entirely changed the information retrieval and resource management approach. Representations of structured data make semantic extraction easy. KGs are gaining popularity in different fields for their concepts, authors, institutions, and topics (Hogan, A. et al., 2022). It enhances query results to focus instead of being a specialized field. Databases are incapable of capturing the domain-specific relationships between topics and concepts. These are crucial for refining searches. Information is scattered in scientific papers, databases, and other sources. The assimilation of this information (Ward, Warren, & Hanisch, 2014) has become a challenge for all fields, like natural language processing (Danilevsky, 2020) and robotics. On the other hand, other sources of information like Wikidata, Google, Bing, Ask.com, and others have evolved into robust answer providers for specific questions (Mandal & Mandal, 2024a). Knowledge graphs (Uyar & Aliyu, 2015) encode information using a mesh of related nodes and edges, allowing the end user to search and browse data exhaustively. An essential application of knowledge graphs is to represent and integrate disparate data sources like scientific literature, databases, and ontologies (Rong, Yuan, & Yang, 2024). It facilitates data-driven discovery, establishes connections between entities, and allows the entities to be investigated and analyzed. They serve to represent and share information. They foster the rate of scientific development and enhance the comprehension of the environment. Knowledge graphs are applied in many areas and different applications, including search engines (Cheng, Yang, Wang, Zhang, & Zhang, 2020), social networks, recommendation systems (Zou, 2020), health care (An, Y, 2022), and finance (Paulheim, 2016). Younghee Noh and Woojeong Kwak (2023) addressed creating a database linking local culture and arts resources with the already registered cultural asset areas. The paper initially identified the nature and level of local resources that could be connected. The researchers supplied metadata about local culture and arts resources, which were categorized as material and publication data, document data, audiovisual data, oral recording data, village information, and personal information. These knowledge graphs (Xiong, Power, & Callan, 2017) require manual curation of data and the development of custom ontologies, which are laborious and time-consuming (Mandal & Mandal, 2024b). The algorithm is used for graph

representation and learning on articles' metadata in the online search engine Semantic Scholar (Wang, Mao, Wang, & Guo, 2017). KG embedding in searching academic literature improves the relevance of the returned documents due to the reliance on semantics and entity matching. Without graph embedding (Fensel et al., 2020), they devised a semantic matching approach based on support vector machines. From the above literature study, it is clear that there are several challenges in constructing knowledge graphs, such as a lack of semantic understanding, limited exploration of relationships, and broad generalization. As Chen et al. (2021) pointed out, both Library and Information Science (LIS) and Computer Science have become interested in knowledge graphs as a major field. They compared literature in SCI-Expanded and EI Compendex databases and visualized them in Citespace. It was discovered that LIS research is on knowledge graph plotting and Visualization. In contrast, Computer Science research concentrates on the construction method of knowledge, like entity recognition, knowledge integration, and Semantic Web applications. The domain knowledge graphs were found to be one of the most significant research directions that can provide new information about the development of this field. Hofer et al. (2023) investigated the increased demand for generalized pipelines to construct and renew knowledge graphs. Although generating knowledge graphs with structured and unstructured data is widely researched, its application in partial updates and integration is a little-investigated area. The paper has provided an overview of graph models, requirements of upcoming pipelines, and the most important processes like metadata management, ontology creation, and quality assurance. It also assessed existing tools and strategies, and identified gaps and areas that need additional research. The article by Bu et al. (2023) suggested a machine approach to creating knowledge graphs on large amounts of unstructured information. Their architecture is a hybrid of convolution neural networks (CNN) and attention-directed bidirectional layers of LSTM, and its performance is similar to that of RoBERTa but is more lightweight and faster. They also presented a paradigm to improve the knowledge graph by incorporating external data through relation linking and visualized using Neo4j. The method demonstrated that it can increase performance and be used in other fields, such as music education, where knowledge graphs may contribute to educational development.

3. Materials and Methods

Knowledge graphs and information in the library are designed to address the challenges outlined above by providing a domain-specific search engine. This system integrates these relationships into a search interface that allows users to search semantically across journals. The first step in constructing a KG is the collection of data. For this study, we selected a comprehensive dataset based on only 55 LIS journal articles from databases like JSTOR, IEEE Xplore, and Scopus. The dataset includes metadata like authors, journals, concepts, and institutions. These articles are from 2023 to 2024. Articles are chosen based on relevance, accessibility, and the presence of structured metadata. A user survey was also done among 20 research scholars who were selected randomly to rate their satisfaction with the search results, interface usability, and ease of exploring relationships between entities. For this purpose, a brief questionnaire with 11 questions is used based on general information,

literature search, and knowledge graph usage. We manually review 100 randomly selected triples to validate the KG's accuracy and evaluate precision, recall, and F1-score metrics.

3.1 Construction of Knowledge Graph

The knowledge graph construction in library and information science journal articles is based on four steps: data extraction, pre-processing, knowledge graph construction, and Visualization.

3.1.1 Data Extraction

The system uses natural language processing techniques to extract entities from journal articles, such as authors, research topics, keywords, and institutions. These entities are linked based on co-occurrence, citation networks, and topic similarities.

3.1.1.1 Defining Scope

At first, relevant LIS journals and the timeframe are defined for this process.

3.1.1.2 Query Design

Advanced search functionalities retrieve articles based on keywords, subjects, titles, and PIDs.

3.1.1.3 Data Export

Export metadata fields, such as title, abstract, authors, keywords, references, and citations, in a structured format (e.g., CSV or XML).

3.1.2 Data Pre-processing

Once the entities are extracted, they are organized into a knowledge graph where nodes represent entities (e.g., authors, journals, concepts) and edges represent relationships (e.g., co-authorships, citations, topic overlap). This graph structure allows for more nuanced queries and the ability to explore connections between research elements.

3.1.2.1 Data Cleaning

Incomplete, irrelevant, and duplicate records are removed from the datasets.

3.1.2.2 Normalization

Standardize author names, institution names, and keywords to ensure consistency.

3.1.2.3 Enrichment

The data is from many LIS journals indexed in prominent academic databases like Scopus, Web of Science, and DOAJ. Using APIs from CrossRef and ORCID ensures the accuracy and comprehensiveness of its entity data, providing users with reliable and up-to-date information. Incorporate additional

data sources, such as ORCID or institutional repositories, to enrich metadata.

3.1.3 Knowledge Graphs Construction

3.1.3.1 Selection of Tools

Use tools like RDFLib, Neo4j, or GraphDB to build the graph.

3.1.3.2 Ontology Development

Define the schema, including entities (e.g., articles, authors) and relationships (e.g., cites, authored by).

3.1.3.3 Semantic Annotations

Link entities to external ontologies or vocabularies like DBpedia or Wikidata.

3.1.4 Visualization

a visual graph representing topics and sub-topics where users can explore nodes (e.g., journals, topics, authors) and navigate through their relationships interactively.

3.1.4.1 Querying

It advances patterns and query discovery. Users can search for entities within the graph to explore related articles, authors, or concepts.

3.1.4.2 Analysis

Conduct a bibliometric and scientometric analysis to derive insights into research trends and collaboration patterns.

4. Results

After constructing the datasets, the following knowledge graphs are produced. Based on search results, they show various nodes, such as authors, places, co-authors, PIDs (Persistent Identifiers), and key terms.

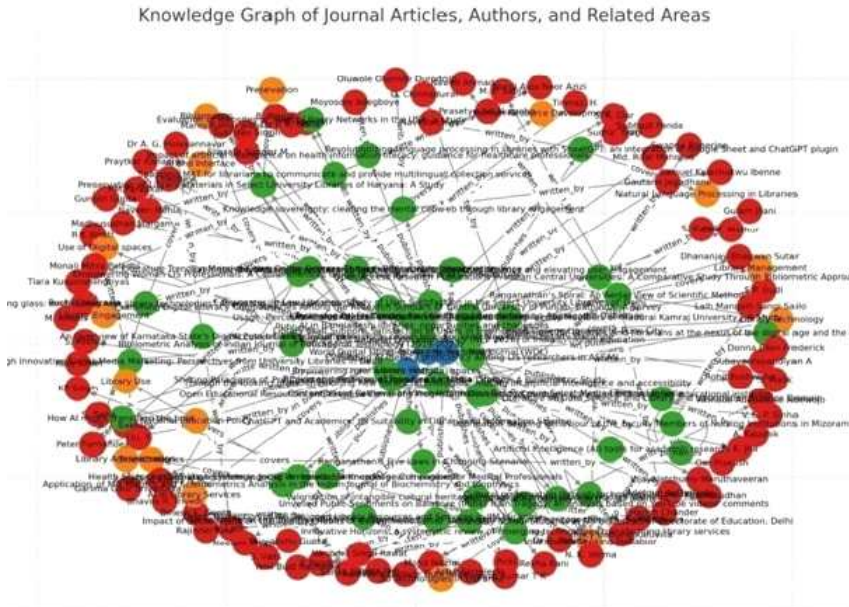


Fig 1. Knowledge graphs of LIS Journal Articles, Authors, and Related Areas

The above knowledge graph uses data from LIS journal articles related to authors and areas. It shows the connections between articles, authors, and key research topics. It shows journals publishing Articles and authors connected to articles. They have contributed to related areas. The colours differentiate between entities: journals, articles, authors, and associated areas. The graph helps to visualize the relationships and trends in LIS research.

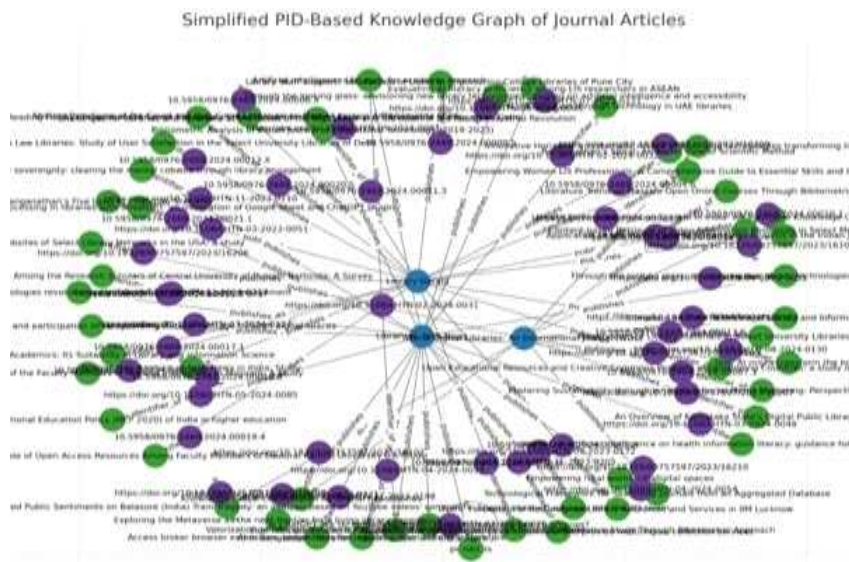


Fig 2. PID-Based Knowledge Graphs of Journal Articles

The above knowledge is created with available PIDs (Persistent Identifiers) from LIS journal articles. It links articles, authors, and topics through unique identifiers. The graph ensures accurate connections and helps track research easily. This PID-based knowledge graph focuses on relationships between PIDs and articles and their respective journals. Each PID is connected to its associated article, and each journal is shown as publishing its articles. This Visualization simplifies and effectively focuses on the PID-based relationship.

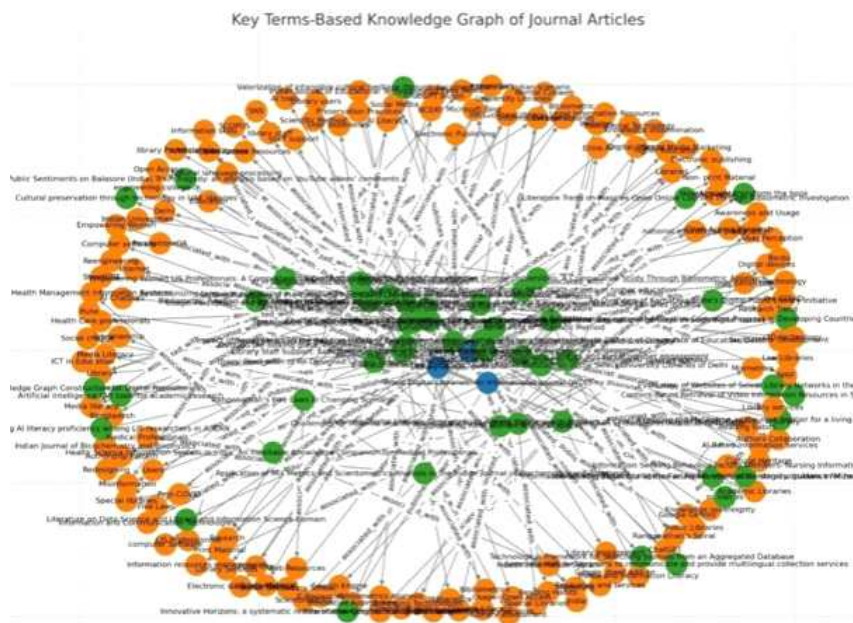


Fig 3. Key-term based Knowledge Graph of Journal Articles

The above figure shows knowledge graphs created using key terms from journal articles. It connects essential terms, topics, and concepts. The graph helps to explore research themes and relationships in LIS studies. This graph shows journal articles connect multiple key terms. Each entity and label illustrates the relationships and key concepts covered in each article.

5. Discussions

This knowledge graph is used for academic information retrieval. The metadata datasets of 55 journals provide clear and relevant search results that can be evaluated using three key metrics: precision, recall, and user satisfaction.

5.1 Precision and Recall

In a set of precision and recall tests, the knowledge graph provides more relevant articles concerning

other databases like Google Scholar and Scopus due to the system's ability to understand the contextual relationships between preservation, policy, and library systems. For instance, a search on "RK Bhatt" retrieved more relevant articles due to the system's ability to understand the contextual relationships between preservation, policy, and library systems. Figure 1 shows the various nodes, which refer to multiple fields and names of the authors, and it also indicates various relationships and nodes. We manually review 100 randomly selected triples to validate the KG's accuracy and evaluate precision, recall, and F1-score metrics. From this study, it is clear that 85 (True Positive) triples are correctly identified, 10 triples (False Positive) are identified incorrectly, and the knowledge graph misses five triples (False Negative). Hence, the results:

$$\text{Precision} = 85 / (85 + 10) = 0.8947 \text{ i.e. } 89.47\%$$

$$\text{Recall} = (85 / (85 + 5)) = 0.9444 \text{ i.e. } 94.44\%$$

$$\text{And F1 Score} = 2 \times (89.47 \times 94.44) / (89.47 + 94.44) = 0.9180 \text{ i.e. } 91.80\%$$

5.2 User Satisfaction

Knowledge graph structure allows faster retrieval of interconnected entities. A usability test was conducted among 20 researchers who used Knowledge Graphs, Scopus, and Google Scholar to perform literature searches. Participants are asked to rate their satisfaction with the search results, interface usability, and ease of exploring relationships between entities. Results show that users prefer Knowledge graphs' semantic search capabilities and intuitive interface. Almost seventeen researchers (85%) out of 20 reported higher satisfaction than those using Scopus and Google Scholar. Constructing them step-by-step knowledge graphs: first, entities (authors, articles, journals, PIDs, key terms) were chosen, then their names, identifiers, and terms were cleaned and normalized to make them less redundant and consistent. Fig. 1 presents connections between authors, articles, journals, and related fields to emphasize the research output, collaboration, and topic trends. The persistent identifiers (DOIs, ORCIDs, and ISSNs) allow an accurate aggregation of articles, authors, and publication journals, and this accuracy makes identity management reliable, reducing redundancy. These graphs in combination reveal who publishes, where the research is published, and what topics are under discussion, as well as revealing influential funders and fine-grained journals, areas of intersection, and areas of weak coverage, with PIDs augmenting the confidence and key terms supplying thematic content.

6. Implication

Knowledge graphs (KGs) become a solution to information access and resource finding in a new, structured format. They enable access to information, data fusion, and semantics-based reasoning. Given the high rate of increase in the number of literature publications, the search conducted by researchers and practitioners should be of high relative rank. The knowledge graph is among the most promising tools for data organization and knowledge extraction, and it boasts numerous

tools. It portrays the orderly information in the format of nodes and relationships, enabling more meaningful search capabilities amongst the academic sources. KGs can be applied relatively easily to understand the complexity of the interconnections between scholarly work, scholarly authors, and their associated institutions, as well as the subjects they address in the discipline of Library and Information Science (LIS). The technology boosts information categorization in that the search results are more relevant to the given query, and thus exploration becomes more accurate and less time-consuming. This paper discusses how knowledge graphs can be constructed from LIS journal articles through a systematic approach to extracting entities, relationships, and attributes from the literature. The resultant knowledge graph is promising in accuracy with an 89.47 percent recall, 94.44 percent precision, and 91.80 percent F1. Moreover, a user study revealed a satisfaction level of 85 percent, meaning that users consider the system effective in improving search relevance, interface usability, and relationship exploration options compared to the current platforms (such as Scopus and Google Scholar).

6.1 Scalability Aspects: Scalability is a significant consideration when building on a knowledge graph. Since the number of LIS publications is still rising, it should be able to cope with extensive, heterogeneous data collections. Scalable architecture demands a distributed storage system, incremental updating, and a real-time entity recognition pipeline. Lack of these could mean performance deteriorates when the corpus grows. Parallelized graph processing and cloud-based structures should be further implemented in the future in order to prioritize large datasets that are navigable at an acceptable response time.

6.2 Practitioner Adoption Pathways: To achieve the goal of extensive adoption by librarians, researchers, and practitioners, pathways should be well laid. This includes the integration of KGs in tri-fiber optic transmission, meaning the existing systems of the library catalogs, institutional repositories, and digital library systems through APIs and plug-in tools. Using training modules and easy-to-use interfaces can reduce the technical barriers to entry. Industry-wide events such as collaborative workshops and best practice guidelines for KG construction will also give LIS professionals confidence to switch to these systems. Compatible with existing and widely used tools, such as OPACs, discovery layers, and citation databases, may motivate practitioners to accept them.

6.3 Practitioner Adoption Pathways: To achieve the goal of extensive adoption by librarians, researchers, and practitioners, pathways should be well laid. This includes the integration of KGs in tri-fiber optic transmission, meaning the existing systems of the library catalogs, institutional repositories, and digital library systems through APIs and plug-in tools. Using training modules and easy-to-use interfaces can reduce the technical barriers to entry. Industry-wide events such as collaborative workshops and best practice guidelines for KG construction will also give LIS professionals confidence to switch to these systems. Being compatible with existing and widely used tools, such as OPACs, discovery layers, and citation databases, may motivate practitioners to accept them.

7. Conclusion and Future Directions

LIS journals' value lies in their ability to optimize the management of resources and the retrieval of information in Library and Information Science. The findings demonstrate the efficacy of KGs in representing scholarly knowledge to offer a scalable framework for academic libraries and improve domain-specific insights. The construction of these knowledge graphs strengthens resource management and enables more timely discoveries. Future directions are to enhance the natural language processing procedures concerning entity identification, address novel disciplines, and implement further sophisticated graph-based learning methods.

References

- An, Y., Greenberg, J., Hu, X., Kalinowski, A., Fang, X., Zhao, X., McClellan, S., Uribe-Romo, F. J., Langlois, K., Furst, J., Gomez-Gualdron, D. A., Fajardo-Rojas, F., Ardila, K., Saikin, S. K., Harper, C. A., & Daniel, R. (2022). Exploring pre-trained language models to build a knowledge graph for metal-organic frameworks (MOFs). *2022 IEEE International Conference on Big Data (Big Data)*, 3651-3658. <https://doi.org/10.1109/BigData55660.2022.10020568>
- Bu, F., Wang, Y., Li, Y., Zhang, M., & Sui, Y. (2023). Automatic knowledge graph construction over efficient information extraction networks. *2023 International Conference on Intelligent Education and Intelligent Research (IEIR)*, 1-7. <https://doi.org/10.1109/IEIR59294.2023.10391258>
- Chen, K., Xie, B., & Deng, S. (2021). Theme evolution of research on knowledge graphs based on visualization analyses of data. *Journal of Physics: Conference Series*, 1813(1), 012039. <https://doi.org/10.1088/1742-6596/1813/1/012039>
- Cheng, D., Yang, F., Wang, X., Zhang, Y., & Zhang, L. (2020). Knowledge graph-based event embedding framework for financial quantitative investments. *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*, 2221-2230. <https://doi.org/10.1145/3397271.3401427>
- Danilevsky, M., Qian, K., Aharonov, R., Katsis, Y., Kawas, B., & Sen, P. (2020). A survey of the state of explainable AI for natural language processing. *Proceedings of the 1st Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics and the 10th International Joint Conference on Natural Language Processing*, 447-459. <https://doi.org/10.18653/v1/2020.aacl-main.46>
- Fensel, D., Şimşek, U., Angele, K., Huaman, E., Kärle, E., Panasiuk, O., Toma, I., Umbrich, J., & Wahler, A. (2020). *Knowledge graphs: Methodology, tools, and selected use cases*. Springer International Publishing. <https://doi.org/10.1007/978-3-030-37439-6>
- Hofer, M., Obraczka, D., Saeedi, A., Köpcke, H., & Rahm, E. (2023). *Construction of knowledge graphs: Current state and challenges*. <https://doi.org/10.2139/ssrn.4605059>
- Hogan, A., Blomqvist, E., Cochez, M., D'Amato, C., Melo, G. D., Gutierrez, C., Kirrane, S., Gayo, J. E. L., Navigli, R., Neumaier, S., Ngomo, A.-C. N., Polleres, A., Rashid, S. M., Rula,
-

- A., Schmelzeisen, L., Sequeda, J., Staab, S., & Zimmermann, A. (2022). Knowledge graphs. *ACM Computing Surveys*, 54(4), 1-37. <https://doi.org/10.1145/3447772>
- Mandal, P. S., & Mandal, S. (2024a). Insights on tools for knowledge graphs in library and information science. *Pearl : A Journal of Library and Information Science*, 18(4), 244-248. <https://doi.org/10.5958/0975-6922.2024.00027.X>
- Mandal, P. S., & Mandal, S. (2024b). Multilingual knowledge graphs: Challenges and opportunities. *International Journal of Knowledge Content Development & Technology*, 14(4), 101-111. <https://doi.org/10.5865/IJKCT.2024.14.4.101>
- Noh, Y., & Kwak, W. (2023). A study on constructing a linked database for an integrated service platform of local culture and arts resources. *International Journal of Knowledge Content Development & Technology*, 13(4), 119-137. <https://doi.org/10.5865/IJKCT.2023.13.4.119>
- Paulheim, H. (2016). Knowledge graph refinement: A survey of approaches and evaluation methods. *Semantic Web*, 8(3), 489-508. <https://doi.org/10.3233/SW-160218>
- Rong, Z., Yuan, L., & Yang, L. (2024). Enhanced knowledge graph recommendation algorithm based on multi-level contrastive learning. *Scientific Reports*, 14(1), 23051. <https://doi.org/10.1038/s41598-024-74516-z>
- Uyar, A., & Aliyu, F. M. (2015). Evaluating search features of Google Knowledge Graph and Bing Satori: Entity types, list searches, and query interfaces. *Online Information Review*, 39(2), 197-213. <https://doi.org/10.1108/OIR-10-2014-0257>
- Wang, Q., Mao, Z., Wang, B., & Guo, L. (2017). Knowledge graph embedding: A survey of approaches and applications. *IEEE Transactions on Knowledge and Data Engineering*, 29(12), 2724-2743. <https://doi.org/10.1109/TKDE.2017.2754499>
- Ward, C. H., Warren, J. A., & Hanisch, R. J. (2014). Making materials science and engineering data more valuable research products. *Integrating Materials and Manufacturing Innovation*, 3(1), 292-308. <https://doi.org/10.1186/s40192-014-0022-8>
- Xiong, C., Power, R., & Callan, J. (2017). Explicit semantic ranking for academic search via knowledge graph embedding. *Proceedings of the 26th International Conference on World Wide Web*, 1271-1279. <https://doi.org/10.1145/3038912.3052558>
- Zou, X. (2020). A survey on the application of the knowledge graph. *Journal of Physics: Conference Series*, 1487(1), 012016. <https://doi.org/10.1088/1742-6596/1487/1/012016>

[About the authors]

Mr. Partha Sarathi Mandal is working as Librarian in shyamsundar Ramlal Adararsha Vidyalaya. He obtains M.Phil; MLIS and M.A from the University of Burdwan and presently pursuing his PhD in the Department of Library and Information Science, the University of Burdwan. He qualifies UGC-NET in 2019. He has published 35 research articles in various National and International journals. His area of interest includes: Machine Learning, Artificial Intelligence, Knowledge Graph, Internet of Things, Artificial Intelligence, Webometrics, Database, Information Retrieval System, Semantic Web, Linked Open Data and Open Knowledge System. Contribution in current study,

he selected the research problem, designed the research methodology and collected data for the study.

Dr. Sukumar Mandal is designated as Assistant Professor in the Department of Library and Information Science at The University of Burdwan. He obtains M.Com from the University of Burdwan. He also obtains MLIS and Ph.D from The University of Burdwan. His area of interest are Machine Learning, Artificial Intelligence, Knowledge Graph, Scientometrics, Bibliometrics, Integrated Library system, Digital Library System, Community Information System and Services, Institutional Digital Repository, Multilingual Information Retrieval System, Semantic Web, Library Administration and Automation, Thesaurus Construction, Visual Vocabulary, Linked Open Data, Open Access, Open Knowledge System etc.
