Review of automatic and semi-automatic creation of knowledge graphs from structured and unstructured data

Auritë Bytyçi, Lorik Ramosaj and Eliot Bytyçi

University of Prishtina "Hasan Prishtina", Prishtina, 10000, Kosovo

Abstract

In today's data-driven world, data storage and information extraction are key processes where the focus of researchers and industries is centralized. While relational databases are widely used, knowledge graphs have emerged as a cutting-edge idea and have begun to compete with them as a result of utilization from renowned companies. Nevertheless, creating knowledge graphs is a time-consuming process that needs domain experts' support. This motivated us to research automatic or semi-automatic creation of knowledge graphs from structured or unstructured data. To do that and comprehend the newest advancements in this field, we have analyzed papers published in the recent five years in well-known libraries such as ACM, IEEE, Springer Link, and ScienceDirect. The analysis takes into account the process of building knowledge graphs, which encompasses the steps involved in managing unstructured texts, defining nodes and relationships between them, and developing ontologies. Moreover, widely used machine learning algorithms including support vector machines, neural networks, random forests, and logistic regression, and other algorithms such as K-Means, TF-IDF, BERT, and skip-gram, the usage of graph database platform Neo4j and Python scripting language, were considered. Conclusive to the study, there exist some semi-automatic approaches but the fully automatic ones remain as ideas.

Keywords

Knowledge graph, machine learning, automatic creation

1. Introduction

An enormous amount of data, in terms of large velocity and large volume, is being produced worldwide [1], with an estimated daily data production of 2.5 billion terabytes [2]. Using data analysis for various data architectures, especially ones that accelerate machine learning, it is possible to identify efficient ways to turn this volume of data into knowledge. On the other hand, this diverse data ecosystem has been a key point in the transition from standard machine architecture to other specialized architectures, aiming for satisfactory efficiency, particularly the effective use of available resources. Nevertheless, even though hardware architecture and machine learning algorithms are crucial components of resource efficiency, a smart combination of both is required to accomplish this efficient use of resources [1].

A novel idea, called a knowledge graph, emerged as a result of the entire process. Considering that sensors may extract a lot of data that needs to be analyzed, knowledge graphs are suitable for application in a variety of contexts. One such context could be automated driving, where data analysis conclusions from an initial state of datasets offer less semantic and poorer structure than the application of knowledge graphs. Knowledge graphs could be used to represent driving scenes and thus concluding that knowledge graphs are more sophisticated in capturing different relations between entities in

Proceedings of RTA-CSIT 2023, April 26–27, 2023 Tirana, Albania
EMAIL: aurite.bytyci@student.uni-pr.edu (A. 1); lorik.ramosaj@student.uni-pr.edu (A. 2); eliot.bytyci@uni-pr.edu (A. 3)
ORCID: 0009-0005-6907-0896 (A. 1); 0009-0000-6557-1857 (A. 2); 0000-0001-7273-9929 (A. 3)
© 2023 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).
CEUR Workshop Proceedings (CEUR-WS.org)
CEUR-WS.org/Vol-3402/paper09.pdf

driving scenarios [3]. Other examples of knowledge graphs application alongside sensors include augmenting weather sensor data with remote knowledge graphs [4]. Yet another example is sensor usage to detect consumers' health indicators to obtain a large amount of data for knowledge graph creation based on food science and industry [5].

The Internet of Things (IoT) data is an adequate illustration of the tremendous explosion of data and the usage of knowledge graphs. The network of physical objects known as the IoT includes tools, instruments, cars, buildings, and other things equipped with electronics, circuits, software, sensors, and network connectivity, allowing these objects to gather and share data [6]. The IoT data segment is expanding rapidly, and according to Statista, there will be 25.44 billion IoT devices in use worldwide by 2030, a prediction based on forecasts made by DataSphere and StorageSphere showing that IoT data is the fastest-growing data segment [1]. Data produced needs to be stored somewhere and today we mainly use cloud storage to store this enormous quantity of data as the most adaptable and convenient approach [7].

The increasing demand for knowledge graphs in various domains has led to the need for methods to try to generate them automatically. Experts frequently manually construct domain-specific knowledge graphs, which process can be time-consuming and error-prone. Furthermore, manual creation may not be possible in large-scale or quickly evolving domains where the volume of information is excessively vast to be done by hand. This served as motivation to check recent developments involving the automatic or semi-automatic generation of knowledge graphs [8].

The remainder of the paper is organized as follows: Section 2 explains the rationale of relational databases and their relation to knowledge graphs, while Section 3 describes the methodological approach used in the review and Section 4 presents a detailed analysis of the obtained papers. In Section 5, the paper is concluded.

2. Relational databases and knowledge graphs

In the 1970s, research at IBM and the University of California Berkeley led to the development of relational databases. They were initially a response to rising expenses for implementing and maintaining sophisticated systems [9]. Edgar Codd was the first to advance the idea of a relational model [10]. He proposed an idea where in a data table, the rows correspond to unique things (such as students), and the columns to numerous qualities (such as student ID, first name, last name, and GPA) [11].

The fundamental principle of relational database architecture has two advantages: it reduces the amount of storage space required for the database and it allows the management of one or more relationships between collection items [12]. Structured data, which easily fits in well-organized tables, performs well in these databases. As opposed to that, relational databases have trouble working with unstructured and distributed data, due to the difficulty of linking their tables across a distributed system [13]. Relational databases are usually used by businesses and often for transactions that require great precision. They thus support ACID (atomicity, consistency, isolation, and durability) constraints [13]. Moreover, their popularity can be attributed to the relational data model's simplicity and adaptability, which enable effective and efficient data management. Due to the relational data model's straightforward and dependable data management capabilities, healthcare organizations can store and access data; such as patient demographics, medical histories, test results and treatment records. Thus, relational databases are used in medical informatics to support the confidentiality and security of sensitive patient information by providing robust access control methods. The advantages of using relational databases also extend to the field of GIS (Geographic Information Systems), including consistent data management and analysis, real-time data querying (especially in route planning), location-based services, and spatial data analysis [14]. Another benefit of using relational databases is to manage financial data, such as transaction records, portfolio holdings, and market data [15].

Since Euler introduced Graph Theory as a novel mathematical idea in 1735, graphs have played a significant role in computational sciences and mathematics [16]. Given that the focus of our work is on evaluating certain entities and the relationships that they have with one another, incorporating graph application into different approaches could be quite beneficial when considering the analogy of associating entities to nodes and relationships to edges. That choice is further strengthened by the notion that knowledge graphs possess methods for large-scale data extraction from many sources of data. Moreover, standard relational procedures (such as join, union, etc.) employed in other No-SQL models are empowered by specialized graph query languages by providing the possibility of finding entities connected by various paths of different lengths [17].

Since their big debut in 2012, when Google used their knowledge graph, knowledge graphs have seen extensive use in both industry and academia [18]. Google Knowledge Graph has been followed by further announcements of knowledge graphs by Airbnb [19], Amazon [20], Microsoft [21], LinkedIn [22] and Facebook [23], among others. Applications for data representation as graphs can be found in a wide range of fields, including food science and industry [5], medicine and healthcare [24], economics, banking, and finance [25] and plenty others. Knowledge graphs use the nodes and edges technique of items and interactions providing intuitive abstraction in a variety of domains [26]. The use of ontologies, a tangible representation of a term's meaning within the context of usage in the computational setting, lends support to this positive argument [17].

Some aspects of knowledge graph structure are very important for data mining and knowledge discovery usage. Considering a graph-based database, knowledge discovery is further improved and the results derived from specific cases show significantly good benefits with the main focus on a novel approach using knowledge graphs and machine learning [27].

3. Methodology

Our research is based on recommendations from Prisma [28] and follows their checklist for performing a systematic review. The search was conducted in four libraries: IEEExplore, ScienceDirect, Springer Link and ACM digital library. The query used for the search was: (“databases” OR “relations databases”) AND “knowledge graphs” AND “automatic” AND “creation”. With this query we intended to locate any relevant papers that would explain generation of knowledge graphs, automatically or semi automatically from tabular data or relational databases. Due to recency of the topic, we have decided to limit the search to the last five years, resulting in a period from 2017 until December 2022, when the search was concluded.

Thus, the strategy used has identified 1027 potential papers in all four libraries. In IEEExplore search resulted in 232 papers, while in Springer Link the search resulted in 288 papers. In ScienceDirect and ACM digital library the searches resulted in 342 papers and 165 papers, respectively. Diagram presented in Figure 1 explains the selection process.

After the initial collection of the papers, the selection process started. Initially the collection was checked for duplicates that might have come up in different libraries. After that books and book chapters were removed from the collection, followed by surveys and reviews. The selection concentrated in Journal and Conference papers, since those would report the most recent progress in the field. The selection process continued with the procedure of selecting the most appropriate papers by reviewing each paper on the keywords and abstracts. If the reviewer had doubt, another reviewer would help to filter and select or not the paper. That concluded the first round of the reviews, which resulted in 56 papers selected. The second round involved reading the whole paper and deciding whether the paper is still interesting and fits to the criteria defined in the beginning of the review. A final table with all selected papers was created. The table was filled with different characteristics of the papers, such as: domain, methodology used, conclusions and future suggestions of the authors.

The final 21 papers chosen for the review were analyzed in depth and their analysis is presented in Section 3. Papers were selected by consensus of the reviewers and their analysis described also according to the overall consensus. It should be noted that we were not able to find online 2 papers as complete documents, therefore we were not able to include them, resulting in 19 papers to be analyzed.

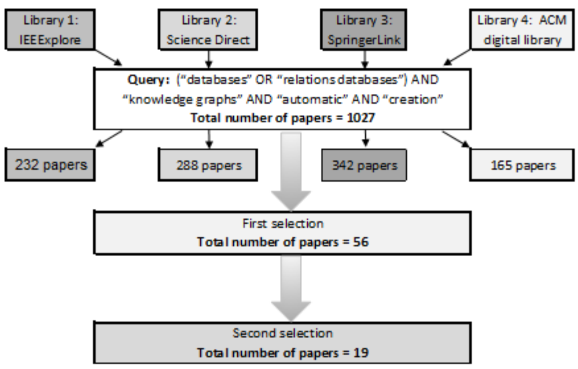

Figure 1. Graph explaining the selection process

4. Research analysis of the selected papers

4.1. Meta data analysis

Preliminary findings regarding the quantity of the papers published and reviewed by the study, are presented in Figure 2. Even though there was a drop-down tendency, in papers reviewed, which we believe was due to unforeseen worldwide situations, the trend kept growing for the future. This we believe is related to the growing usage of knowledge graphs in different domains

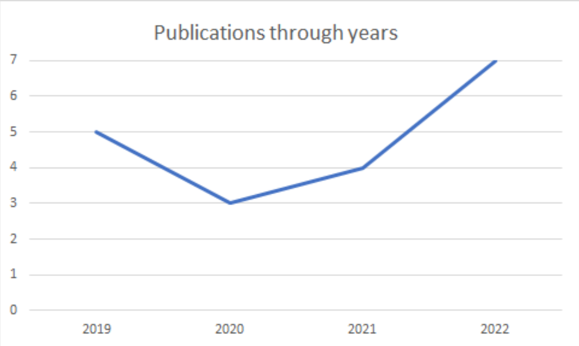

Figure 2. Number of reviewed papers according to the publication years

As per venues of the paper presentations, presented in Figure 3, one specific conference stands out: International joint conference on knowledge graphs. Nevertheless, there are also well-established journals present such as The Semantic Web, Expert Systems with Applications and Journal of Biomedical Semantics, which we believe relates to the domain of the research.

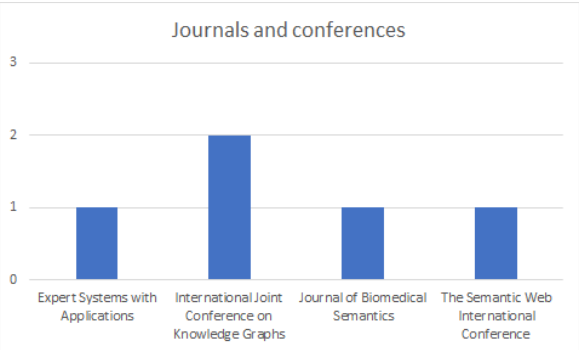

Figure 3. Journals and conference venues of the reviewed papers

The most used research domain, in the analyzed papers, is medicine [29, 40, 42, 45] and education [34, 35, 39, 46], followed by research conducted with graphs in research papers [32, 38], followed by other diverse domains [30, 31, 32, 33, 36, 37, 41, 43, 44, 47] as shown in Figure 4.

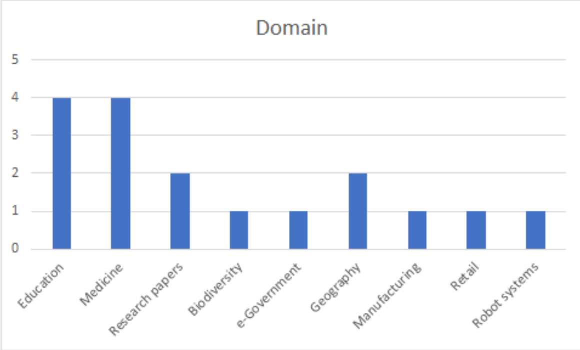

Figure 4. Reviewed papers according to the domain of study

4.2. Thorough paper analysis

Researchers in [29] suggest that by bypassing nearly completely the supervision of knowledge extraction in a biomedical application, it can be generalized for other domains as well. Yet, another semi supervised knowledge graph creation was proposed by [30] but for a different industry, welding. Knowledge graph creation in [34] was performed with the combination of manual and automatic methods, while authors in [36] focus on creating biodiversity knowledge graphs in semi-automatic fashion, with usage of data mining.

Few other authors such as [31] suggest construction of an automatic tool for knowledge graph creation in retails, and similarly authors in [35] suggest automatic construction methods.

In several papers, a crucial part of the process involves handling unstructured texts. One approach involves using information extraction to retrieve facts from text, with each fact comprising a subject, predicate, and object [29, 34], while other strategies include using knowledge extraction techniques to extract information from unstructured texts [33]. Moreover, there are approaches that incorporate unstructured data from free texts as data subject to extraction of electronic medical records of patients for vector representation [40] and using unstructured texts as input that is transformed into cognitive script understandable by humans for the development of a computing model for a graph-based repository [42]. Entities and relationships between them in the dataset are among the important beginning points in the knowledge graph creation procedure. Entity linking, which allows the linking of text spans to previously known entities, is described in some papers as the first step in the construction of knowledge graphs [29] and managing mappings between cells and entities in a knowledge graph [44]. Besides entity linking, information extraction is another crucial component in building knowledge graphs, involving extracted entities and relationships [34] alongside the preprocessing phase that includes a model with entities and relationships for knowledge graph [36]. Developing a neural architecture for the identification of entities and relationships between them in order to construct knowledge graphs [38], carrying out manual analysis for detected entities [40], executing Cypher statements to create entities with properties and corresponding relationships after they have been generated with Python that was applied to cognitive scripts [42] and using named entity recognition to create more machine-readable sentences [32] are other important approaches used in the remaining papers. Two other relevant concepts mentioned and used are: data cleaning - which identifies anomalies from existing data and newly imputed values [31] and provides possibility of achieving an effective form of data expression when data cleaning of triple data is applied [34], and handling synonyms -creating a lexical redirection table to keep track of synonym relationships between tags [39], enriching knowledge graph with synonyms [36] and associating synonyms between product types and attribute values [31]. Important aspect of knowledge graph creation are ontologies, too, and several approaches have been used in the papers: using many ontologies for knowledge graph creation that are created for particular tasks in order to model the general knowledge, connect datatype properties with classes with the same name and produce upper level schema [30], as a comprehensive suite in the architecture consisting of taxonomy enrichment and relation discovery [31] and particular ontology facilitated by automatic class suggestions at the sentence annotation stage [32]. Other strategies employed include constructing the knowledge graph ontology using Protégé tool [34], using the prefixed structure of the DISK ontology to transform sentences into readable texts [43], et cetera. Utilizing an ontology that was suggested in another article is a different strategy [39], whereas some papers don't mention any specific ontologies [35, 36, 38, 42].

According to our research, most of the papers that dealt with generation of graphs from tabular data / relational databases, used machine learning algorithms such as logistic regression [31, 40], random forest [31, 40, 46], support vector machine [31, 40, 46], neural networks (mainly to train the respective models) [33, 35, 38, 43, 46], and other algorithms [30, 32, 34, 39, 41]. They have been an important aspect of knowledge graph creation, used as a trained model for future predictions [40], helping manual processes on knowledge graph creation [32], building the ontology [41], integrating data from multiple sources to create an unified knowledge representation (with the help of semantic web technologies) [46], and providing analytical functionality [30]. Alongside machine learning algorithms, K-means algorithm is used in Knowledge Graph embeddings algorithm [43] and also finds an important role in creating the knowledge graph from extracted concepts and relationships using natural language processing [34], while natural language processing (among the most famous mentioned are TF-IDF [34, 39], BERT and skip-gram [38]) and mining the data are found to be very useful for identifying spans and other necessary elements that need to be provided in order to create a knowledge graph. Other algorithms used involve cognitive scripting to define knowledge graph representation [42], entity linking [29, 32, 44] and Viterbi algorithm [33]. Some papers might have performed manual conversion from relational databases to knowledge graphs [46] or no specific algorithm was mentioned [37, 47].

Knowledge storage is another essential component of the process of the creation of knowledge graphs [34]. In some of the publications we examined, Neo4j is a unique graph database platform [39] used as a storage tool that in some cases incorporates the visualization process [34] with the D3.js visualization tool from Neo4j used to complete the visualization. Loading the original data, specifying the data into the document elements, determining the scope of components, and setting attributes to control the transition and change process of the elements are the steps of using D3.js based on the data document to address the limitation of data presentation [34]. Another paper refers to Neo4j as a non-relational graph store, a structured repository for the extracted knowledge [42]. The nodes in a graph store are connected by relationships between other nodes in the domain, which are all machine-readable entities [42]. It contains the characteristics of the entities and their relationships as a key-value pair to describe specific pieces of information, as an essential component of knowledge graph storing and visualization [42].

The use of Python, is a significant finding from the papers we analyzed, too. Python was used to create an innovative program that produces Cypher statements from a cognitive script to create nodes and edges in Neo4j [42]. In a different study, during the creation of a knowledge graph by utilizing a pre-created ontology two Python libraries are used [37]. The first one is the Osmium Python library, which is used to retrieve all OSM nodes from the newest OSM dumbs, while the other library, known as RDFlib, is responsible for creating RDF triples.

4.3. Limitations of the review

Knowledge graphs have seen a recent increase in usage by major players in industry and thus also a vast number of papers are dealing with them in the last few years. This makes the review even harder, especially taking into account the number of fields dealing with knowledge graphs.

Even though we have used major libraries to retrieve the number of papers, we might have missed any of the other relevant papers from other different online databases. Moreover, non-English papers were not included in the review, and we were not able to retrieve a couple of them as mentioned in the methodology part.

Despite the limitations, to the best of our knowledge, the majority of interesting papers for this review were retrieved in the last five years. Those papers were classified and analyzed to identify a few interesting aspects tackled by the researchers related to knowledge graphs.

4.4. Future challenge directions

Regarding future research, although many of the papers discuss enhancing their models, it is noteworthy that BERT is mentioned in few of the papers [32, 35, 38, 42] as the way forward. Nevertheless, GPT-3 is seen as a possibility to use for the pre-trained models, too [42].

5. Conclusion

Machine learning algorithms, along with semantic data mining, deep reinforcement learning, crowdsourcing, and natural language processing methods, are heavily utilized in the building of knowledge graphs. Even though some papers suggest semi-automatic and automatic methods for producing knowledge graphs, manual methods are less popular because of their shortcomings, which include the inability to easily combine data from various sources into a single knowledge graph, the potential for bias in knowledge graph creation, and the requirement for ongoing maintenance.

Applications for knowledge graphs can be found in a variety of industries, including education, metallurgy, healthcare, and the food sector, as depicted in Figure 4 from the reviewed papers. The results obtained in the publications studied for this review show that it is possible to extract important data and relationships from these created knowledge graphs by using machine learning methods, underpinned by the observation that a majority of the scrutinized papers employ at least one variant of machine learning algorithms in their analyses. Another conclusion that can be inferred is that knowledge graphs can be effectively and usefully constructed when the data and relationships in a knowledge graph are properly structured, utilizing well-designed ontologies and cutting-edge algorithms.

6. References

[1] Morik, Katharina. "Machine Learning under Resource Constraints."

[2] Reddy, AV Rama Krishna, T. Murali Krishna, and J. Mohan Kumar. "A Short Review on News Articles Clustering." Mathematical Statistician and Engineering Applications 71.1 (2022): 167-172.

[3] Halilaj, Lavdim, et al. "Knowledge graphs for automated driving." 2022 IEEE Fifth International Conference on Artificial Intelligence and Knowledge Engineering (AIKE). IEEE, 2022.

[4] Wu, Jiantao, et al. "Augmenting weather sensor data with remote knowledge graphs." IGARSS 2022-2022 IEEE International Geoscience and Remote Sensing Symposium. IEEE, 2022.

[5] Min, Weiqing, et al. "Applications of knowledge graphs for food science and industry." Patterns 3.5 (2022): 100484.

[6] Gokhale, P., Bhat, O., & Bhat, S. (2018). Introduction to IOT. International Advanced Research Journal in Science, Engineering and Technology, 5(1), 41-44.

[7] Trikha, B. (2010). A journey from floppy disk to cloud storage. International Journal

on Computer Science and Engineering, 2(4), 1449-1452.
[8] Wang, T., Gu, H., Wu, Z., & Gao, J. (2020). Multi-source knowledge integration based on machine learning algorithms for domain ontology. Neural Computing and Applications, 32, 235-245.
[9] Seltzer, M. (2008). Beyond relational databases. Communications of the ACM, 51(7), 52-58.
[10] Codd, E. F. (1983). A relational model of data for large shared data banks. Communications of the ACM, 26(1), 64-69.
[11] Strawn, G., & Strawn, C. (2016). Relational databases: Codd, Stonebraker, and Ellison. IT Professional, 18(2), 63-65.
[12] Morris, P. J. (2005). Relational database design and implementation for biodiversity informatics. The Academy of Natural Sciences1900 Ben Franklin Parkway, Philadelphia.
[13] Leavitt, N. (2010). Will NoSQL databases live up to their promise?. Computer, 43(2), 12-14.
[14] Tan, K. S., Egenhofer, M. J., & Laurentijn, R. J. G. (1994). Relational Databases in Geographic Information Systems. Geographical Systems, 1(1), 1-16.
[15] Seeber, R. J., Johnson, D. W., & Thompson, G. L. (1998). Relational Databases in Financial Management. Financial Management, 27(1), 1-15.
[16] Trudeau, R.J. Introduction to Graph Theory; Courier Corporation: North Chelmsford, MA, USA, 2013.
[17] Hogan, Aidan, et al. Knowledge Graphs. Springer, 2019.
[18] Shaoxiong, Ji, et al. "A survey on knowledge graphs: Representation, acquisition, and applications." IEEE transactions on neural networks and learning systems (2021): 494-514.
[19] Chiticariu, Laura, Yunyao Li, and Frederick Reiss. "Rule-based information extraction is dead! long live rule-based information extraction systems!." Proceedings of the 2013 conference on empirical methods in natural language processing. 2013.
[20] Hitzler, P. "Modeling with ontology design patterns: Chess games as a worked example." Ontology Engineering with Ontology Design Patterns: Foundations and Applications 25 (2016): 3.
[21] Skjæveland, Martin G., et al. "Practical ontology pattern instantiation, discovery, and maintenance with reasonable ontology templates." The Semantic Web–ISWC 2018: 17th International Semantic Web Conference, Monterey, CA, USA, October 8–12, 2018, Proceedings, Part I 17. Springer International Publishing, 2018.
[22] Heindorf, Stefan, et al. "Vandalism detection in wikidata." Proceedings of the 25th ACM International on Conference on Information and Knowledge Management. 2016.
[23] Page, Lawrence, et al. The PageRank citation ranking: Bringing order to the web. Stanford InfoLab, 1999.
[24] Rajabi, Enayat, and Somayeh Kafaie. "Knowledge Graphs and Explainable AI in Healthcare." Information 13.10 (2022): 459.
[25] Wang, Shuai. "On the Analysis of Large Integrated Knowledge Graphs for Economics, Banking and Finance." EDBT/ICDT Workshops. 2022.
[26] Angles, Renzo. "The Property Graph Database Model." AMW. 2018.
[27] Muñoz, Emir. Knowledge graph mining with latent shape graphs. Diss. NUI Galway, 2020.
[28] Page, Matthew J., Joanne E. McKenzie, Patrick M. Bossuyt, Isabelle Boutron, Tammy C. Hoffmann, Cynthia D. Mulrow, Larissa Shamseer et al. "The PRISMA 2020 statement: an updated guideline for reporting systematic reviews." Systematic reviews 10, no. 1 (2021): 1-11.
[29] Kroll, Hermann, Jan Pirklbauer, and Wolf-Tilo Balke. "A toolbox for the nearly-unsupervised construction of digital library knowledge graphs." 2021 ACM/IEEE Joint Conference on Digital Libraries (JCDL). IEEE, 2021.
[30] Zhou, Baifan, et al. "Scaling usability of ML analytics with knowledge graphs: exemplified with a Bosch welding case." The 10th International Joint Conference on Knowledge Graphs. 2021.
[31] Dong, Xin Luna, et al. "Autoknow: Self-driving knowledge collection for products of thousands of types." Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. 2020.
[32] Oelen, Allard, Markus Stocker, and Sören Auer. "Crowdsourcing scholarly discourse annotations." 26th International Conference on Intelligent User Interfaces. 2021.
[33] Deng, Jianfeng, et al. "Research on event logic knowledge graph construction method of robot transmission system fault

diagnosis." IEEE Access 10 (2022): 17656-17673.
[34] Bai, Junhua, and Lei Che. "Construction and Application of Database Micro-course Knowledge Graph Based on Neo4j." The 2nd International Conference on Computing and Data Science. 2021.
[35] Su, Ying, and Yong Zhang. "Automatic construction of subject knowledge graph based on educational big data." Proceedings of the 2020 The 3rd International Conference on Big Data and Education. 2020.
[36] Sharafeldeen, Dina, Alsayed Algergawy, and Birgitta König-Ries. "Towards knowledge graph construction using semantic data mining." Proceedings of the 21st International Conference on Information Integration and Web-based Applications & Services. 2019.
[37] Dsouza, Alishiba, et al. "Worldkg: A world-scale geographic knowledge graph." Proceedings of the 30th ACM International Conference on Information & Knowledge Management. 2021.
[38] Trivedi, Ishdutt, and Sudhan Majhi. "Span level model for the construction of scientific knowledge graph." 2020 5th International Conference on Computing, Communication and Security (ICCCS). IEEE, 2020.
[39] Qiao, Lei, et al. "Automated constmction of course knowledge graph based on China MOOC platform." 2019 IEEE International Conference on Engineering, Technology and Education (TALE). IEEE, 2019.
[40] Gazzotti, Raphaël, et al. "Extending electronic medical records vector models with knowledge graphs to improve hospitalization prediction." Journal of Biomedical Semantics 13.1 (2022): 1-20.
[41] Wang, Ting, et al. "Multi-source knowledge integration based on machine learning algorithms for domain ontology." Neural Computing and Applications 32 (2020): 235-245.
[42] Fahd, Kiran, et al. "Knowledge graph model development for knowledge discovery in dementia research using cognitive scripting and next-generation graph-based database: a design science research approach." Social Network Analysis and Mining 12.1 (2022): 61.
[43] de Haan, Rosaline, Ilaria Tiddi, and Wouter Beek. "Discovering research hypotheses in social science using knowledge graph embeddings." The Semantic Web: 18th International Conference, ESWC 2021, Virtual Event, June 6–10, 2021, Proceedings 18. Springer International Publishing, 2021.
[44] Cremaschi, Marco, et al. "A fully automated approach to a complete semantic table interpretation." Future Generation Computer Systems 112 (2020): 478-500.
[45] Sloep, Matthijs, et al. "A knowledge graph representation of baseline characteristics for the Dutch proton therapy research registry." Clinical and Translational Radiation Oncology 31 (2021): 93-96.
[46] Paneque, Manuel, María del Mar Roldán-García, and José García-Nieto. "e-LION: Data integration semantic model to enhance predictive analytics in e-Learning." Expert Systems with Applications 213 (2023): 118892.
[47] Alobaid, Ahmad, and Oscar Corcho. "Balancing coverage and specificity for semantic labelling of subject columns." Knowledge-Based Systems 240 (2022): 108092.