**ICCEPM 2022**
The 9th International Conference on Construction Engineering and Project Management
*Jun. 20-23, 2022, Las Vegas, NV, USA*

# The Status Quo of Graph Databases in Construction Research

Kahyun Jeon[1]*, Ghang Lee[1]

[1] *Department of Architecture and Architectural Engineering, Yonsei University, Yonsei-ro 50, Seodaemoon-gu, Seoul, South Korea,* E-mail address: jeonkh0310@yonsei.ac.kr, glee@yonsei.ac.kr

**Abstract:** This study aims to review the use of graph databases in construction research. Based on the diagnosis of the current research status, a future research direction is proposed. The use of graph databases in construction research has been increasing because of the efficiency in expressing complex relations between entities in construction big data. However, no study has been conducted to review systematically the status quo of graph databases. This study analyzes 42 papers in total that deployed a graph model and graph database in construction research, both quantitatively and qualitatively. A keyword analysis, topic modeling, and qualitative content analysis were conducted. The review identified the research topics, types of data sources that compose a graph, and the graph database application methods and algorithms. Although the current research is still in a nascent stage, the graph database research has great potential to develop into an advanced stage, fused with artificial intelligence (AI) in the future, based on the active usage trends this study revealed.

**Key words:** graph database, graph model, knowledge graph, construction information management, Neo4j

## 1. INTRODUCTION

Construction information has been exponentially growing in a highly interconnected manner. For example, a large amount of structural health monitoring data has been accumulated using digital sensors in a close relationship with structural building information modeling (BIM) data. Graph databases have been attracting attention as a means to manage the scalability and complex relationships of such construction information. A graph database can model the intricate relationships of real-world data as a graph, and it is easy to add or remove the data without a schema. In addition, knowledge inferences through graph algorithms can provide insights into problem solving in construction management. Thanks to these advantages, quite a few studies deployed graph databases in managing construction information. However, no study has yet reviewed how these studies utilized the graph database in construction research. This study conducted a critical review of 42 papers, both quantitatively and qualitatively. This study focuses on four research questions: 1) What is the research trend? 2) In which research topics were graph databases used? 3) How were data sources translated into graph database components? and 4) Which graph algorithms were deployed?

The following section introduces the research method. The overall research trend is described in the third section. The analysis of the research topic and the data resources are explained in the fourth and fifth sections. The applied graph algorithms are discussed in the sixth section. This paper will be concluded in the seventh section with a summary of the current status and future research directions.

## 2. RESEARCH METHOD

This study collected papers from academic databases, such as Web of Science, Science Direct, and the American Society of Civil Engineers (ASCE) library of the domain-specific database. The combinations of two keyword groups—one about the graph and the other about construction—were used to search queries. Through two steps of filtering, 42 papers in total were reviewed in this study. The overall data acquisition process is described in Figure 1**.**

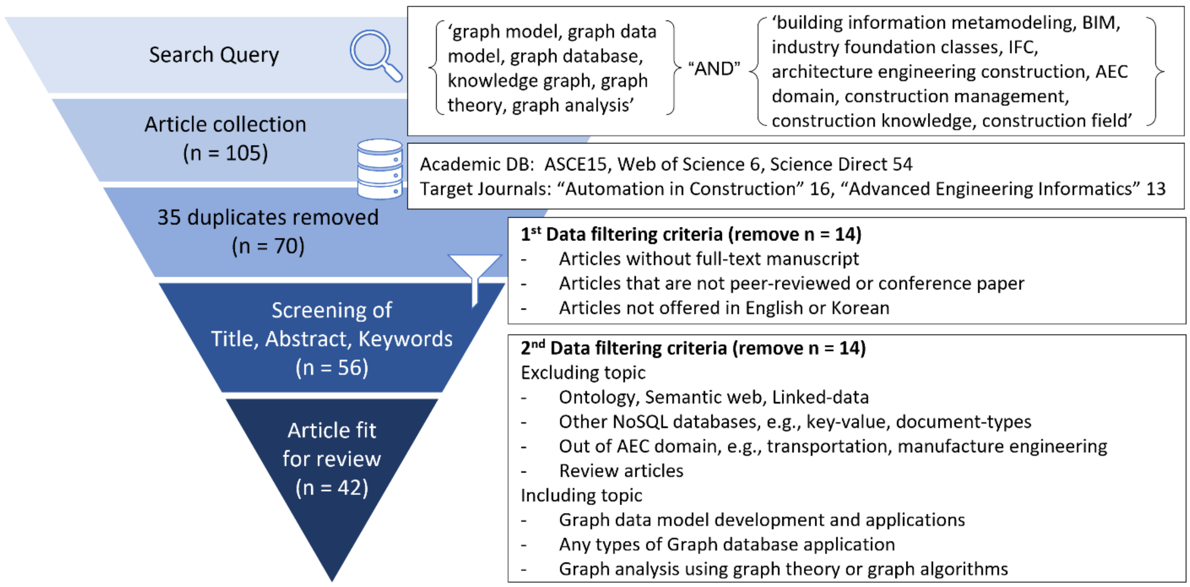


**Figure 2.** Paper selection process

This study conducted a statistical method, keyword analysis, and topic modeling as the quantitative analysis. Based on the statistical method, the number of papers with the publication type and year and the distribution of graph database management systems were analyzed. To conduct a keyword analysis and topic modeling, the NLP procedure was applied to the corpus, including the title, abstract, and keywords. In the NLP procedure, tokenization, lemmatization, and part-of-speech tagging were sequentially conducted. Both bi-gram and tri-gram keyword analyses were performed on a total of 1,360 non-redundant keywords. To identify the detailed research topics, Latent Dirichlet Allocation (LDA)—an unsupervised learning-based topic modeling method [1]—was used. The content analysis was manually conducted as a qualitative analysis to investigate the following issues in more detail: 1) how construction data sources were converted into graph database components, such as nodes, edges, and properties, and 2) the application methods of graph databases and graph algorithms.

## 3. RESEARCH TREND OF GRAPH DATABASES

The research trend was investigated through a descriptive statistical analysis. Figure 3 shows the yearly research trend by publication type. A small number of studies was performed from 2011 to 2016. Since 2017, graph databases have been actively adopted in many studies in the construction sectors. In total, 81% of papers were published in construction and civil engineering-related journals, such as Automation in Construction and the Journal of Computing in Civil Engineering.

The most dominantly used graph database management system was Neo4j (82%), followed by OrientDB, ArangoDB, and others.

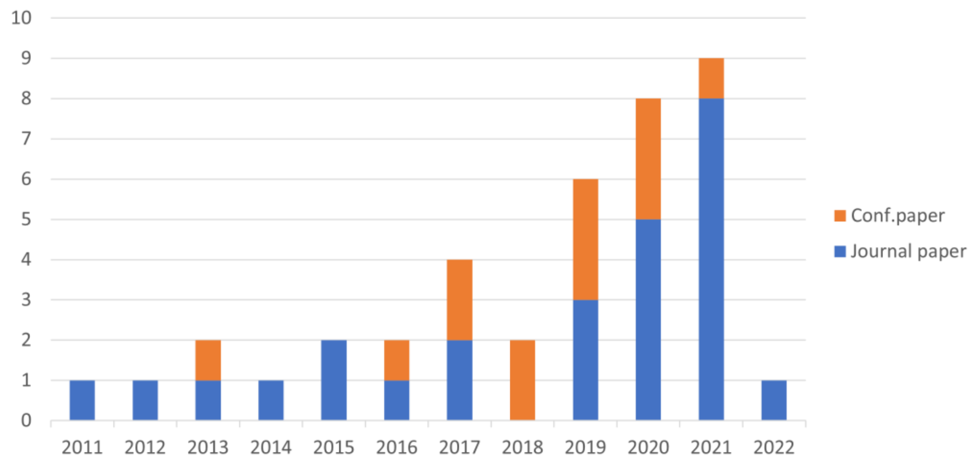


**Figure 3.** Annual research trend by publication type

## 4. RESEARCH TOPICS ACHIEVED BY GRAPH DATABASE

To classify papers with similar research objectives, topic modeling was conducted. Instead of providing a single keyword, topic modeling yields a set of keywords, which are closely associated with a topic. To group topic modeling keywords into a single representative topic, the top ten keywords used in the papers were derived through the n-gram analysis of the words included in the title, abstract, and keywords of the 42 papers first. Then, the topic modeling keywords were grouped using the top ten keywords as a headword (Table 1).

In examining the topic keywords, such words as ‘graph,’ ‘database,’ ‘data,’ and ‘model’ appear in all papers were removed first. As single keywords such as ‘web’ and ‘clash’ often do not represent a research topic, bi-gram and tri-gram analyses were conducted to identify keywords composed of two or three words. The top ten keywords were ‘IFC-based,’ ‘building element,’ ‘semantic web,’ ‘prefabricated building,’ ‘BIM GIS integrated,’ ‘clash detection,’ ‘improve clash correction,’ ‘BIM enabled SCM,’ ‘information flow construction,’ and ‘design rule knowledge’.

The most dominant research topic, which a ratio of 26.2%, was ‘support design rules and decision making,’ such as proposing optimal design alternatives, finding similar floorplans, and tracking design changes in BIM. The purpose of converting IFC-based BIM data to a graph database was secondly dominant at 16.7%. The graph-represented BIM has been used for the purposes of space or structural topology analysis and clash detection. As for other popular research topics, graph-based semantic information extraction, the construction of domain-specific knowledge graphs, and schedule analysis using graph theory were addressed as major topics.

**Table 1**. Research purposes of graph database applied research in construction

| Research Topic | Matching top ten keywords | Topic Modeling Keywords | Ratio (%) | References |
|---|---|---|---|---|
| Design rule knowledge | design rule knowledge | Decision, Drawing, Design, Rules Knowledge | 26.2 | [2,5,6,8,11–16] |
| IFC | IFC-based | BIM, IFC, Object | 16.7 | [17–23] |

| | | | | |
|---|---|---|---|---|
| BIM topology | building element | Structure, Topology, Hierarchy, Flexibility | 9.5 | [3,4,24,25] |
| BIM GIS integrated | BIM GIS integrated | Urban, city, Advanced, Performance | 9.5 | [26–29] |
| clash detection | clash detection, improve clash correction | Clash, Detection, Resolution, Impact, Minimize | 9.5 | [30–33] |
| Semantic information extraction | semantic web | Semantic, Information, Extraction, Feature | 7.1 | [34–36] |
| Knowledge graph | design rule knowledge, information flow construction | Knowledge, Extraction, Natural language Unstructured, Document | 7.1 | [9,37,38] |
| BIM enabled SCM | BIM enabled SCM, prefabricated building | Network, Schedule, Graph theory, Machine learning, Evaluation | 4.8 | [39,40] |
| Others | information flow construction | COBie, RDF, Web, Complex, GNN, Room | 9.5 | [10,41–43] |

## 5. DATA SOURCES FOR GRAPH DATABASES

The types of data sources used to construct a graph database were also analyzed through the content analysis. In Table 2, BIM/IFC data were used most frequently as a data source for a graph database. Examples of BIM/IFC data include IFC data, structural BIM models, model view definition data, and the geometric data of a BIM model. Because BIM data is already in a highly structured format, IFC classes and relationships, BIM objects, and attribute information could be systematically mapped to nodes and edges. Design sources, including drawings, sketches, space images, or diagrams, have also been widely used as a data source in graph database studies. In converting a design into a graph, a space, such as a 'kitchen' or 'lobby,' was defined as a node, and the adjacency or accessibility between spaces was defined as a relation.

The text data from various documents, such as specifications, reports, design rules, or regulations, have been also used to construct a knowledge graph. To convert text to a graph, NLP technologies, such as named entity recognition (NER), have been used to extract the information entities. In general, nouns were defined as nodes, and verbs were defined as relationships.

In the case of geographic information system (GIS) data, CityGML, a geographic data format that can be easily imported into a graph database, is often used. In addition, Construction Operations Building information exchange (COBie), scheduling data, videos and images, and urban data were used to construct a graph database.

**Table 2.** Construction data sources for graph database components

| **Data sources** | **Nodes** | **Edges** | **Properties** | **References** |
|---|---|---|---|---|
| BIM, IFC | IFC classes, Entities | IfcRel objects | IfcType, IFC properties, | [5,18,26,28, 34,35,40] |
| | Structural components | Connection, Dependency | Geometry properties | [14,20,25,30 ,31,33,41] |

| | | | | |
|---|---|---|---|---|
| | Building rooms | Accessibility, Adjacency | Space position, Area, Room types or size | [7,17] |
| Design knowledges | Space, Place, Design rules, Material, Person | Accessibility, Adjacency | GlobalID, Cost, Schedule, Quality | [2,6,16,43] |
| Text | Entities (nouns) | Relations (verbs) | - | [11,36,37] |
| GIS data | Geographic entities, | Time interval, Place boundary | - | [22,23] |
| | CityGML keys | CityGML values | - | [26] |
| Others: *COBie* | Building components | Actions with verbs | Description, types | [41] |
| *Construction schedules* | A set of tasks, possible roles | Sequence, Use | - | [39] |
| *Urban data* | Urban elements | Proximity relationships | Height, Area, and Time-series data | [27] |
| *Videos and images* | Construction site objects | Spatial distance, Safety criteria | - | [9] |

## 6. APPLIED GRAPH ALGORITHMS

The application method of the graph database in the research was classified based on the applied graph algorithms. First, the studies aimed at data converting and knowledge graph construction were in the early stages. These studies were focused on the representation of construction data as a graph database or on the automatic transformation method rather than graph analysis using graph algorithms. Second, most studies used graph databases for the purpose of graph analysis. In addition, the type and distribution of graph algorithms were examined. Among a total of 42 papers, 19 studies used the graph algorithms, and half of them used ‘Pathfinding and Search’, such as the shortest path [2], depth-first search [3] or, graph traversal algorithm [4]. The similarity measurement [5–7] was frequently used for information extraction and design decision support. However, only a few studies tried advanced analysis methods, including graph embedding based on generative adversarial networks (GAN) [8] and graph neural network (GNN) [9,10].

## 7. CONCLUSION

This study reviewed the research trend of graph databases in construction. The major findings of this study are as follows. First, graph databases have been used for construction research since 2011 and have been rapidly increasing in use from 2017 until now. Second, graph databases have been mainly used for design decision making, BIM data transformation, semantic information extraction, clash resolution, and relational data search. Third, the main sources of information included BIM/IFC data, design knowledge, text data, and other construction-related data like COBie, GIS data, or construction scheduling data. Fourth, although the number of studies that utilize a graph database has been continuously increasing, most are still in a nascent stage. A few studies attempted to fuse a graph database with AI, such as GNN or zero-shot learning in analyzing BIM data. As a future research direction, previous studies addressed that the construction of a robust and reliable knowledge graph is the most important step in applying a graph database on an advanced level. Moreover, graph reasoning can discover new knowledge and recommend ideal solutions, which will help solve various problems in the construction industry. In addition, the multi-source heterogeneity of construction data has been identified as a challenge. This study contributes to the understanding of the status quo of the application of graph databases in construction research. It also contributes to understanding the future research direction of construction information management using graph databases.

## ACKNOWLEGEMENTS

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. 2021R1A2C3008209).

## REFERENCES

[1] D.M. Blei, A.Y. Ng, M.I. Jordan, Latent Dirichlet Allocation, Journal of Machine Learning Research. 3 (2003) 993–1022.

[2] G. Ślusarczyk, A. Łachwa, W. Palacz, B. Strug, A. Paszyńska, E. Grabska, An extended hierarchical graph-based building model for design and engineering problems, Automation in Construction. 74 (2017) 95–102. https://doi.org/10.1016/j.autcon.2016.11.008.

[3] A. Khalili, D.K.H. Chua, IFC-Based Graph Data Model for Topological Queries on Building Elements, Journal of Computing in Civil Engineering. 29 (2015) 04014046. https://doi.org/10.1061/(ASCE)CP.1943-5487.0000331.

[4] A. Khalili, D.K.H. Chua, IFC-Based Framework to Move beyond Individual Building Elements toward Configuring a Higher Level of Prefabrication, Journal of Computing in Civil Engineering. 27 (2013) 243–253. https://doi.org/10.1061/(ASCE)CP.1943-5487.0000203.

[5] S. Esser, S. Vilgertshofer, A. Borrmann, Graph-based version control for asynchronous BIM level 3 collaboration, in: 28th EG-ICE International Workshop on Intelligent Computing in Engineering, 2021.

[6] J. Roith, C. Langenhan, F. Petzold, Supporting the building design process with graph-based methods using centrally coordinated federated databases, Visualization in Engineering. 5 (2017) 20. https://doi.org/10.1186/s40327-017-0057-y.

[7] C. Langenhan, M. Weber, M. Liwicki, F. Petzold, A. Dengel, Graph-based retrieval of building information models for supporting the early design stages, Advanced Engineering Informatics. 27 (2013) 413–426. https://doi.org/10.1016/j.aei.2013.04.005.

[8] N. Nauata, K.-H. Chang, C.-Y. Cheng, G. Mori, Y. Furukawa, House-GAN: Relational Generative Adversarial Networks for Graph-Constrained House Layout Generation, in: A. Vedaldi, H. Bischof, T. Brox, J.-M. Frahm (Eds.), Computer Vision – ECCV 2020, Springer International Publishing, Cham, 2020: pp. 162–177. https://doi.org/10.1007/978-3-030-58452-8_10.

[9] Z. Pan, C. Su, Y. Deng, J. Cheng, Video2Entities: A computer vision-based entity extraction framework for updating the architecture, engineering and construction industry knowledge graphs, Automation in Construction. 125 (2021) 103617. https://doi.org/10.1016/j.autcon.2021.103617.

[10] Z. Wang, R. Sacks, T. Yeung, Exploring graph neural networks for semantic enrichment: Room type classification, Automation in Construction. (2021) 104039. https://doi.org/10.1016/j.autcon.2021.104039.

[11] J. Hao, L. Zhao, J. Milisavljevic-Syed, Z. Ming, Integrating and navigating engineering design decision-related knowledge using decision knowledge graph, Advanced Engineering Informatics. 50 (2021) 101366. https://doi.org/10.1016/j.aei.2021.101366.

[12] V.J.L. Gan, BIM-based graph data model for automatic generative design of modular buildings, Automation in Construction. 134 (2022) 104062. https://doi.org/10.1016/j.autcon.2021.104062.

[13] M.H. Rasmussen, M. Lefrançois, P. Pauwels, C.A. Hviid, J. Karlshøj, Managing interrelated project information in AEC Knowledge Graphs, Automation in Construction. 108 (2019) 102956. https://doi.org/10.1016/j.autcon.2019.102956.
[14] J. Nikolic, Open Configuration Model for Collective Housing Building Structure: Graph Model, International Journal for Housing Science & Its Applications. 38 (2014).
[15] B. Strug, Automatic design quality evaluation using graph similarity measures, Automation in Construction. 32 (2013) 187–195. https://doi.org/10.1016/j.autcon.2012.12.015.
[16] E. Grabska, G. Ślusarczyk, Knowledge and reasoning in design systems, Automation in Construction. 20 (2011) 927–934. https://doi.org/10.1016/j.autcon.2011.03.009.
[17] A. Ismail, B. Strug, G. Ślusarczyk, Building Knowledge Extraction from BIM/IFC Data for Analysis in Graph Databases, in: L. Rutkowski, R. Scherer, M. Korytkowski, W. Pedrycz, R. Tadeusiewicz, J.M. Zurada (Eds.), Artificial Intelligence and Soft Computing, Springer International Publishing, Cham, 2018: pp. 652–664. https://doi.org/10.1007/978-3-319-91262-2_57.
[18] A. Ismail, A. Nahar, R. Scherer, Application of graph databases and graph theory concepts for advanced analysing of BIM models based on IFC standard, in: 24th International Workshop on Intelligent Computing in Engineering (EG-ICE 2017), 2017.
[19] E.S. Malinverni, B. Naticchia, J.L. Lerma Garcia, A. Gorreja, J. Lopez Uriarte, F. Di Stefano, A semantic graph database for the interoperability of 3D GIS data, Appl Geomat. (2020). https://doi.org/10.1007/s12518-020-00334-3.
[20] J. Werbrouck, P. Pauwels, M. Bonduel, J. Beetz, W. Bekers, Scan-to-graph: Semantic enrichment of existing building geometry, Automation in Construction. 119 (2020) 103286. https://doi.org/10.1016/j.autcon.2020.103286.
[21] Z. Chen, Y. Pu, D.R. Shelden, A Graph Database and Query Approach to IFC Data Management, Future of Information Exchanges and Interoperability. (2019) 28.
[22] L. Zheng, L. Zhou, X. Zhao, L. Liao, W. Liu, The Spatio-Temporal Data Modeling and Application Based on Graph Database, in: 2017 4th International Conference on Information Science and Control Engineering (ICISCE), 2017: pp. 741–746. https://doi.org/10.1109/ICISCE.2017.159.
[23] J.-E. Lee, G.-I. Rho, H.-M. Jang, K.-U. Yu, System Design and Implementation for Building a Place Information based on Crowdsourcing Utilizing the Graph Data Model, Journal of Cadastre & Land InformatiX. 46 (2016) 117–131. https://doi.org/10.22640/LXSIRI.2016.46.1.117.
[24] S. Isaac, F. Sadeghpour, R. Navon, Analyzing Building Information Using Graph Theory, in: ISARC. Proceedings of the International Symposium on Automation and Robotics in Construction, IAARC Publications, Waterloo, 2013: pp. 1–8.
[25] A. Khalili, D.K.H. Chua, Framework for an IFC-Based Tool for Implementing Design for Deconstruction (DfD), (2012) 619–626. https://doi.org/10.1061/41182(416)76.
[26] A.-H. Hor, G. Sohn, P. Claudio, M. Jadidi, A. Afnan, A Semantic Graph Database for BIM-GIS integrated Information Model for an Intelligent Urban Mobility Web Application, in: ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Copernicus GmbH, 2018: pp. 89–96. https://doi.org/10.5194/isprs-annals-IV-4-89-2018.
[27] K. Gupta, Z. Yang, R.K. Jain, Urban Data Integration Using Proximity Relationship Learning for Design, Management, and Operations of Sustainable Urban Systems, Journal of Computing in Civil Engineering. 33 (2019) 04018063. https://doi.org/10.1061/(ASCE)CP.1943-5487.0000806.
[28] A.-H. Hor, G. Sohn, Design and Evaluation of a Bim-GIS Integrated Information Model Using RDF Graph Database, ISPRS Annals of the Photogrammetry, Remote Sensing and

Spatial Information Sciences. VIII-4/W2-2021 (2021) 175–182. http://dx.doi.org/10.5194/isprs-annals-VIII-4-W2-2021-175-2021.
[29] E. Papadonikolaki, R. Vrijhoef, H. Wamelink, Supply chain integration with BIM: a graph-based model, Structural Survey. 33 (2015) 257–277. https://doi.org/10.1108/SS-01-2015-0001.
[30] Y. Hu, D. Castro-Lacouture, C.M. Eastman, S.B. Navathe, Component Change List Prediction for BIM-Based Clash Resolution from a Graph Perspective, J. Constr. Eng. Manage. 147 (2021) 04021085. https://doi.org/10.1061/(ASCE)CO.1943-7862.0002092.
[31] Y. Hu, D. Castro-Lacouture, C.M. Eastman, S.B. Navathe, Automatic clash correction sequence optimization using a clash dependency network, Automation in Construction. 115 (2020) 103205. https://doi.org/10.1016/j.autcon.2020.103205.
[32] W.-T. Chien, S.-H. Hsieh, A Web-Based Approach to Dynamically Assessing Space Conflicts by Integrating BIM and Graph Database, in: ISARC. Proceedings of the International Symposium on Automation and Robotics in Construction, IAARC Publications, Waterloo, Canada, 2020: pp. 307–312.
[33] Y. Hu, D. Castro-Lacouture, C.M. Eastman, Holistic clash detection improvement using a component dependent network in BIM projects, Automation in Construction. 105 (2019) 102832. https://doi.org/10.1016/j.autcon.2019.102832.
[34] E. Tauscher, K. Smarsly, Generic BIM queries based on the IFC object model using graph theory, in: ICCCBE2016, 2016.
[35] S.A. Morshed, X. Lv, R.B. Tanvir, Network-Based Information Extraction from IFC Files to Support Intelligent BIM Companion (iBcom) Technology, in: Construction Research Congress 2020, 2020: pp. 427–435. https://doi.org/10.1061/9780784482865.046.
[36] C. Wu, X. Wang, P. Wu, J. Wang, R. Jiang, M. Chen, M. Swapan, Hybrid deep learning model for automating constraint modelling in advanced working packaging, Automation in Construction. 127 (2021) 103733. https://doi.org/10.1016/j.autcon.2021.103733.
[37] S. Leng, Z.-Z. Hu, Z. Luo, J.-P. Zhang, J.-R. Lin, Automatic MEP knowledge acquisition based on documents and natural language processing, in: Proceedings of the 36rd CIB W78 Conference, 2019.
[38] H. Zhao, Y. Pan, F. Yang, Research on Information Extraction of Technical Documents and Construction of Domain Knowledge Graph, IEEE Access. 8 (2020) 168087–168098. https://doi.org/10.1109/ACCESS.2020.3024070.
[39] K. Sigalov, M. König, Recognition of process patterns for BIM-based construction schedules, Advanced Engineering Informatics. 33 (2017) 456–472. https://doi.org/10.1016/j.aei.2016.12.003.
[40] B. Yang, M. Dong, C. Wang, B. Liu, Z. Wang, B. Zhang, IFC-Based 4D Construction Management Information Model of Prefabricated Buildings and Its Application in Graph Database, Applied Sciences. 11 (2021) 7270. https://doi.org/10.3390/app11167270.
[41] V. Kumar, E.A.L.E. Teo, Exploring the application of property graph model in visualizing COBie data, Journal of Facilities Management. 19 (2021) 500–526. https://doi.org/10.1108/JFM-09-2020-0066.
[42] A.J.A. Donkers, D. Yang, N. Baken, Linked Data for Smart Homes: Comparing RDF and Labeled Property Graphs, (2020) 14.
[43] R. Pinquié, P. Véron, F. Segonds, T. Zynda, A Property Graph Data Model for a Context-Aware Design Assistant, in: C. Fortin, L. Rivest, A. Bernard, A. Bouras (Eds.), Product Lifecycle Management in the Digital Twin Era, Springer International Publishing, Cham, 2019: pp. 181–190. https://doi.org/10.1007/978-3-030-42250-9_17.