Journal of Theoretical and Applied Information Technology
29th February 2024. Vol.102. No 4

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

STRATEGIC ADVANCEMENT OF BUILDING INFORMATION MODELLING PRACTICES THROUGH KNOWLEDGE GRAPH CONSTRUCTION USING ONTOLOGY AND DEEP LEARNING

HAYAT EL ASRI[1], ASMAA RETBI[2], SAMIR BENNANI[3]

[1,2,3] RIME TEAM-Networking, Modelling, and e-Learning Team, MASI Laboratory, Engineering 3S Research Center, Mohammadia School of Engineering, Mohammed V University in Rabat, Morocco

E-mail: [1]hayat.elasri@research.emi.ac.ma, [2]retbi@emi.ac.ma, [3]sbennani@emi.ac.ma

ABSTRACT

The integration of data from diverse sources continues to present a significant challenge in the Architecture, Engineering, and Construction (AEC) industry. To address this issue, our proposed approach leverages ontologies to structure and organize the disparate data, while employing deep learning algorithms to extract pertinent information and relationships. The resultant knowledge graph serves as a valuable resource that facilitates decision-making, enhances the management of construction projects, and fosters expert knowledge sharing within the domain. To evaluate the effectiveness of our method, we conducted experiments on a manually built and annotated dataset, and the results demonstrated the approach's commendable accuracy and efficiency. The proposed knowledge graph, founded upon the synergistic amalgamation of ontology and deep learning, shows the potential to significantly elevate the efficiency and effectiveness of the construction industry by streamlining data integration, fostering knowledge exchange, and empowering stakeholders with informed decision-making capabilities.
Keywords: *AEC Industry, Building Information Modeling (BIM), Knowledge Graph, Ontology, Deep Learning.*

1. INTRODUCTION

Architecture, Engineering, and Construction (AEC) organizations are constantly seeking innovative approaches, tools, and strategies to enhance organizational effectiveness. Knowledge Graphs (KGs) are one of these strategies. These latter have been increasingly popular in the recent decade.

According to [1], "a knowledge graph, also known as a semantic network, represents a network of real-world entities—that is, objects, events, situations, or concepts—and illustrates the relationship between them". This data is typically kept in a graph database and represented as a graph structure, thus the name knowledge "graph." KGs are an effective way to not only organize, but also visualize massive volumes of complicated data. They make it simple to browse and evaluate data, as well as swiftly establish links between various parts. KGs may also be used to build interactive visualizations that aid in data comprehension and exploration. The ability to see patterns and correlations swiftly, even the hidden ones, can be beneficial for organizations wanting to make better decisions.

KGs have been used in a variety of disciplines including but not limited to artificial intelligence, natural language processing, and data integration, to facilitate smarter and more context-aware decision-making processes. Figure 1 is an example of an application field taxonomy.

Moreover, KGs play a crucial role in enhancing organizational performance by providing a structured and interconnected representation of information that is relevant to the entity's operations.

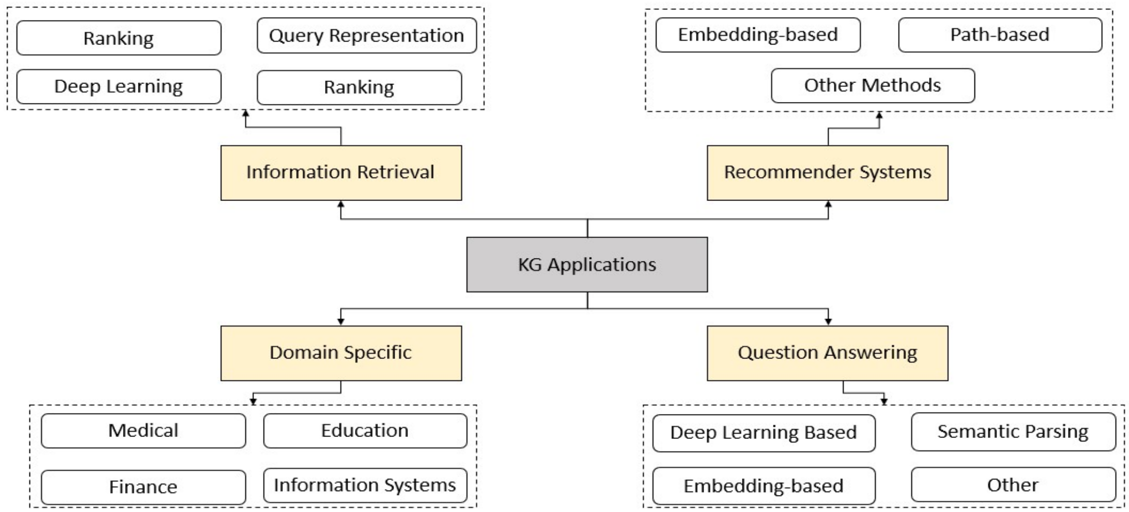

Figure 1: Application Fields Of Kgs (Adapted From [2])

Despite the transformative capabilities of BIM in the construction industry, a prevalent challenge emerges from the extensive and disparate data contributed by various stakeholders. We conducted a prior survey among BIM professionals that revealed a common concern: the cumbersome task of managing and organizing scattered data, hindering the seamless utilization of BIM. The data, originating from diverse sources as previously mentioned, lacks cohesion and structure, leading to inefficiencies in decision-making processes. Recognizing this challenge as a significant impediment to optimal BIM utilization, we aim to address this issue by developing a KG. This solution aims to streamline data integration and enhance accessibility for BIM professionals, fostering a more coherent and efficient BIM ecosystem.

This article is organized as follows. Section 2 discusses some of the pertinent research publications on KG building. Section 3 describes the methodology followed and the technique used in the framework of this research. Moreover, the methods, training dataset, and full procedure are detailed and discussed in this part. Section 4 concludes with the findings and discussion.

2. RELATED WORKS

The literature on KGs for the AEC industry emphasizes their transformational potential in reinventing information management, decision-making, and collaboration. Researchers and practitioners have identified KGs as a strong tool for integrating varied data sources from many stakeholders, hence solving semantic heterogeneity and data fragmentation difficulties in the AEC domain [3]. KGs provide an organized and interconnected representation of domain-specific information, including project data, construction materials, laws, and industry best practices, by utilizing approaches such as ontology modeling, data integration, and semantic reasoning [4]. Furthermore, research have stressed the importance of machine learning methods in developing comprehensive KGs for knowledge extraction, entity resolution, and connection prediction [5]. These KGs improve project management efficiency, risk reduction, and informed decision-making processes by enabling advanced analytics, predictive modeling, and expert knowledge exchange. While several studies have demonstrated the feasibility and effectiveness of knowledge graphs in the AEC industry, there is still room for more research into optimizing graph construction algorithms, broadening the scope of the graph to accommodate emerging technologies such as BIM, and exploring collaborative knowledge-sharing platforms to foster industry-wide innovation and best practices [6].

KGs have been the subject of several research investigations in both industry and academia. According to [7], KGs can aid in the integration of diverse data sources and the improvement of data interoperability in BIM. Likewise, [8] presented a BIM-based KG framework for building performance analysis that may give a more accurate and complete knowledge of building energy usage. Some methodologies, such as semantic web technologies

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

and rule-based reasoning, are employed in this study. However, other restrictions have been overlooked, such as data integration issues and scalability. Other research has focused on the use of KGs for BIM data management, such as [9], who presented a KG-based strategy to simplify BIM data integration and retrieval. The application of KGs in BIM has also been investigated in the context of project management, where KGs can aid in the identification of possible hazards and the facilitation of decision-making processes [10].

This study addresses a significant gap in current knowledge by leveraging ontologies, deep learning, and a KG for the BIM industry.

3. METHODOLOGY

3.1 Research Framework

The research framework followed in this work is presented in figure 2.

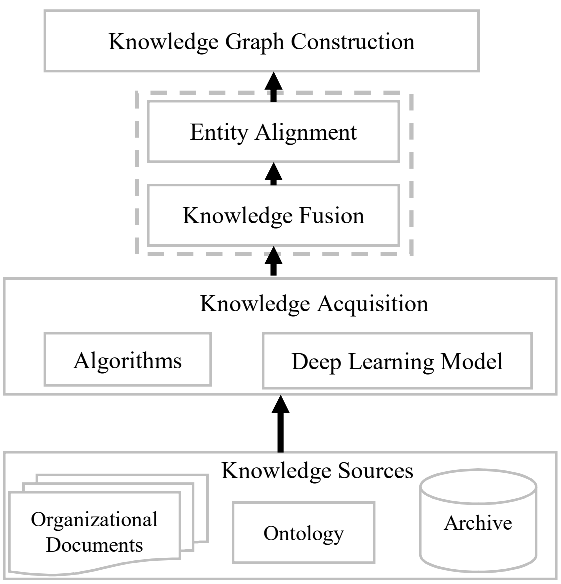

Figure 2: Research Framework

The research framework provided in this study consists of five major blocks, each of which plays an important part in the construction of the KG. To begin, "knowledge sources" include a wide range of information repositories, such as organizational records, archives, and ontologies. Second, the "knowledge acquisition" step entails pulling relevant data from these sources while assuring accurate and domain-specific information retrieval and that, using a deep learning model and machine learning algorithms. "Fusion and entity alignment," the third stage, merges the retrieved data from the different sources while resolving semantic heterogeneity to assure consistency in portraying entities and their connections. Following that, the fourth step, "KG construction," arranges the combined data into a complete and linked KG, forming a network of interconnected entities and characteristics. Finally, the built KG becomes a significant resource for applications such as domain relevance analysis and knowledge reasoning, providing insights and support for decision-making processes in the surveying and remote sensing domains. The cohesive framework connects these various elements, making it easier to build and use the KG to improve knowledge management and exploration in the field.

3.2 Knowledge Sources

There are 5 main knowledge sources: the organizational documents available in the database, 3 domain ontologies that have already been constructed, implemented, tested, and validated by domain experts, and the organizational archives. Organizational documents contain valuable information about the organization's operations, processes, and policies. Furthermore, ontologies provide a standardized vocabulary and conceptual framework for organizing the knowledge, making it easier to integrate and interconnect data from different sources. Lastly, archives are repositories of historical organizational data and records that provide a rich source of information for constructing a KG.

The schema provided in figure 3 lays at the core of the KG.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

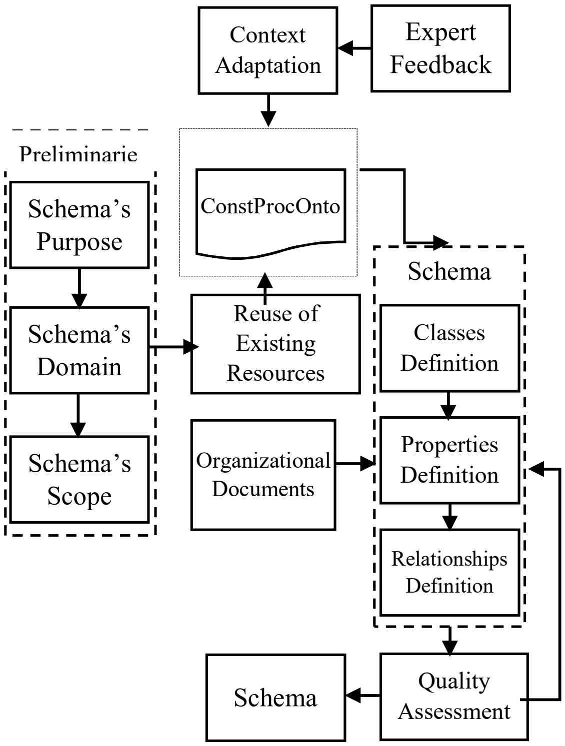

Figure 3: Schema Construction

The process of schema construction commenced with a preliminary phase, wherein the purpose, domain, and scope of the ontology were delineated. This step served to establish the ontology's coverage, intended applications, and content, among other essential aspects. Additionally, an existing construction procurement ontology was adapted to suit the specific context of the research, thoroughly verified, tested, and approved by domain experts, and judiciously repurposed. The following step required methodically and sequentially creating and implementing class definitions, properties, and semantic connections.

Quality tests were undertaken before merging the ontologies into the KG to assure their correctness, coherence, and suitability for the intended applications. This schema creation technique strengthens the KG's dependability and efficacy as a complete knowledge store for the construction area.

3.3 ONTOLOGIES

Two ontologies provide a formal framework for organizing and combining data pertaining to two critical areas of BIM: procurement and building processes.

The procurement ontology establishes the connections between procurement entities such as suppliers, contracts, and purchase orders. It also includes a standardized language for defining procurement-related information including pricing, delivery timelines, and terms and conditions. In contrast, the construction ontology specifies the interactions between construction items such as materials, equipment, and labor. It also includes a standardized language for describing construction data such as building rules, safety standards, and development timelines.

Combining and mapping the procurement and the construction ontologies into a single domain ontology created the base for the BIM ontology. This latter provides common vocabularies and rules for displaying data from different sources. The mapping approach recognizes common concepts and relationships between the ontologies and generates connections between them. As such, a comprehensive and integrated BIM ontology has been developed that incorporates important information from the procurement and construction processes.

The process of ontology validation was a vital stage in the creation of the BIM ontology. A group of automation engineers, computer science specialists, and construction professionals who were familiar with the usage of BIM provided constructive criticism, which helped identify missing and/or incorrect concepts, connections, or limitations. The domain experts also ensured that the ontology was consistent with industry standards and best practices. Interviews were used to gather further expert opinion, which was then integrated into the ontology creation process, resulting in an updated and validated BIM ontology. The validation step was especially important in ensuring that the ontology properly represented the domain and was helpful to BIM stakeholders.

The following figures provide some insights into the developed ontologies.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

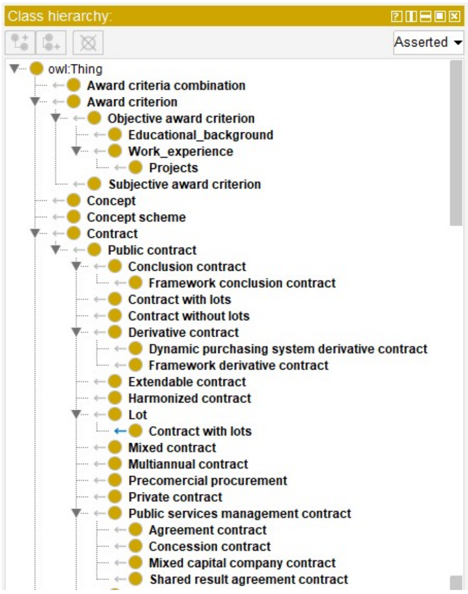

Figure 4: Class Hierarchy Example - Ontology 1

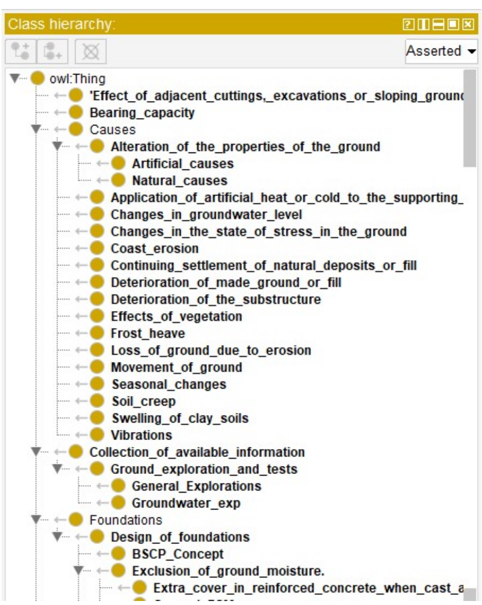

Figure 5: Class Hierarchy Example - Ontology 2

Each class in the class hierarchy represents a set of entities where subclasses inherit the features and characteristics of their parent classes. The class hierarchy is necessary for effective reasoning, querying, and representation.

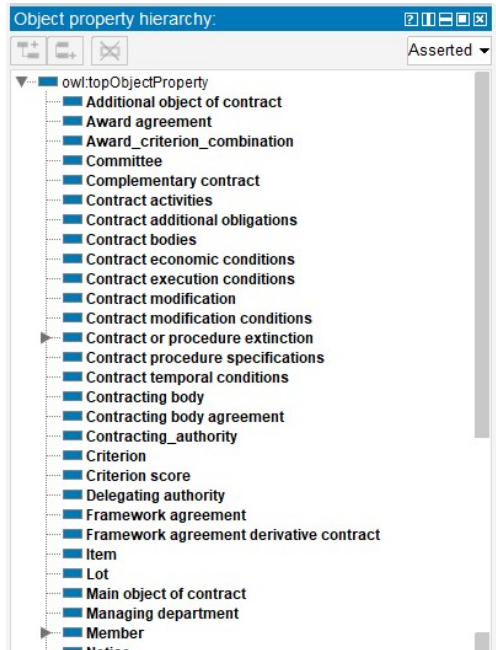

Figure 6: Object Property Hierarchy Example - Ontology 1

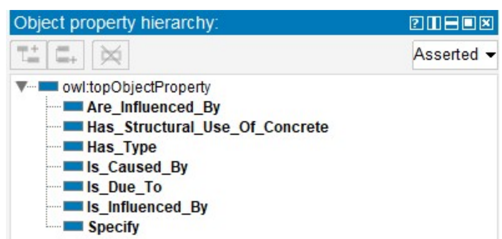

Figure 7: Object Property Hierarchy Example - Ontology 2

The object property hierarchy allows for the structuring of object properties based on their relationships and attributes. Each object property in the hierarchy is connected to one or more other object properties; this allows relationship structuring based on attributes and traits. Moreover, this hierarchy is important because it allows for automated reasoning tasks like as consistency checking, categorization, and inference.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

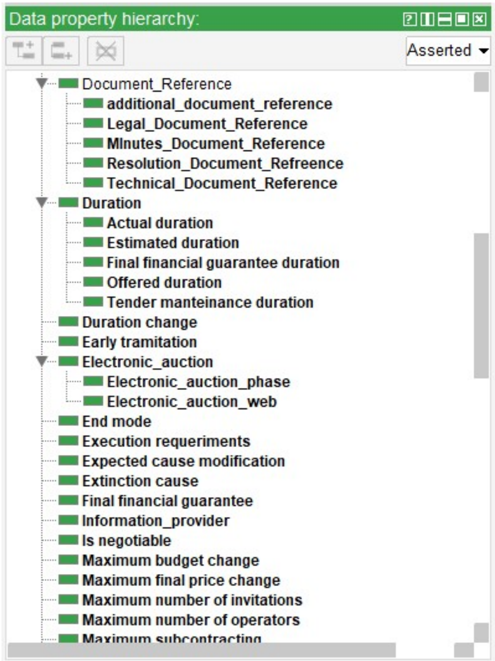

Figure 8: Data Property Hierarchy Example

The data property hierarchy allows for the structuring of data properties based on their relationships and attributes. Each data property in a data property hierarchy is connected to one or more other data properties, establishing a hierarchy. This enables attribute organization depending on their qualities and traits.

The procurement ontology and the construction ontology were merged to create the BIM ontology. This latter was divided into several groups including construction activities, suppliers, and building components. To explain further, the procurement ontology specified the linkages between building components and suppliers by utilizing classes such as "Supplier" and "Product" as well as relationships like "Supplies_Product" and "Has_Component." On the other hand, the construction ontology established the links between building components and construction tasks using classes like "Task," "Building_Component," and "Equipment," as well as relationships like "Uses_Component" and "Performs_Task."

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

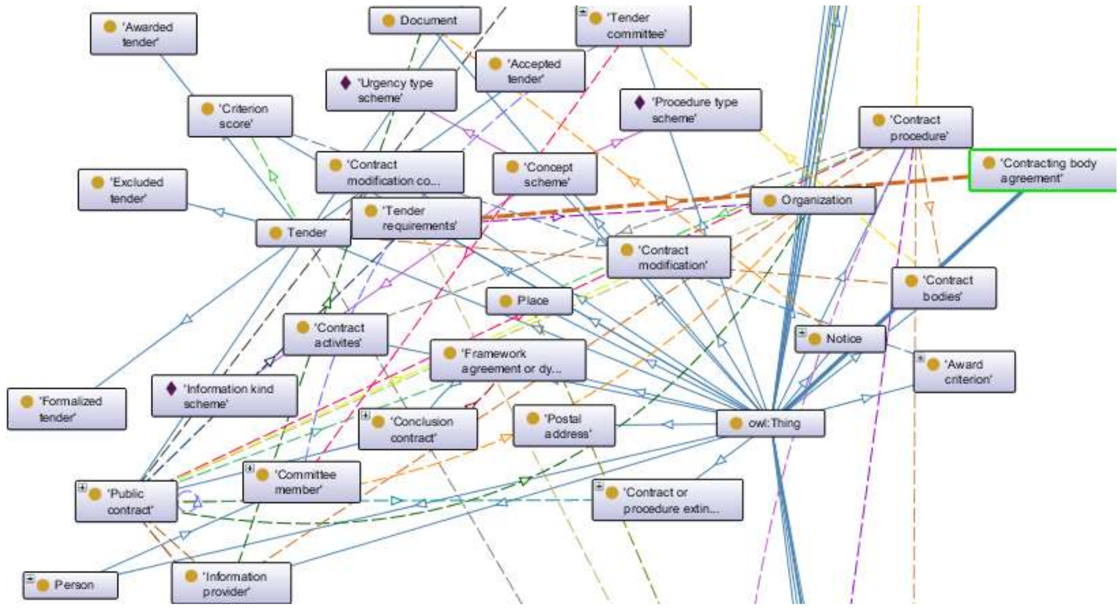

Figure 9: BIM Ontology

Several advantages can be associated with the developed BIM ontology. First, it enabled the seamless integration of procurement and construction procedures, simplifying and making the process more efficient. Second, automated some processes which helped reduce mistakes and delays. Third, it improved cooperation and communication among the AEC stakeholders, including architects, engineers, contractors, and suppliers. Finally, it established a standardized method of storing and accessing organisational data and knowledge that will be utilized in the future.

3.4 KNOWLEDGE EXTRACTION AND ACQUISITION

For knowledge extraction and acquisition, a long short-term memory (LSTM) network and a conditional random field (CRF) model were combined. The BiLSTM-CRF is a natural language processing (NLP) method that may be used for several applications such as named entity recognition, part-of-speech tagging, and intent classification. It also proved to be accurate, making it an excellent candidate for a wide range of NLP tasks.

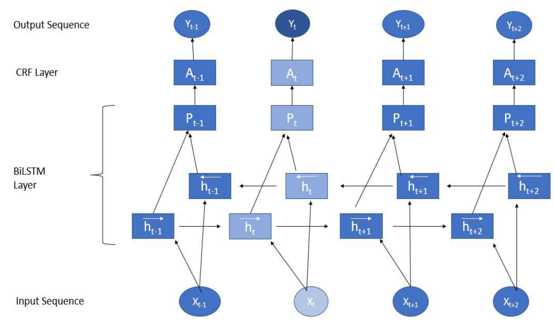

Figure 10: BiLSTM-CRF Algorithm Structure

The LSTM uses a single direction input flow while the Bi-LSTM uses the input in both forward and reverse direction. In 1st graph input sequence as (X_{t-1}, X_t..X_{t+2}). The output of this BiLSTM layer is a sequence of the hidden states for each input word vectors, denoted as (h_{t-1}, h_t .. h_{t+2}). Each final hidden state is the concatenation of the forward hi and backward hi hidden states. We know that:

$\overleftarrow{h_i} = lstm\ (x_i, \overleftarrow{h_{t-1}})$ Forward LSTM

$\overrightarrow{h_i} = lstm\ (x_i, \overrightarrow{h_{t-1}})$ Reverse LSTM

$h_i = [\overleftarrow{h_i}\ ; \overrightarrow{h_i}]$ BiLSTM

This is used to describe the sequential relationships between words and sentences in both directions.
The CRF model is an undirected graph model containing a node's conditional probability at supplied nodes.

A graph can be shown as: G = (V, E), where v is the set of nodes and E is the set of undirected boundaries.

$$Y = \{Y_v \mid v \in V\}$$

Random variables are represented by Y_v corresponding to the node v∈V. Showing the Markov property in Y_v. (X, Y) is a CRF, where X denotes an observed sequence, w v ~ 108 indicates all the neighbor nodes of w connected with node v in G graph.

3.5 Dataset

The BiLSTM-CRF algorithm requires a large-scale training dataset. Nonetheless, there are no available datasets that are ready to use and focused on procurement or construction. Therefore, the researchers have built the dataset and manually annotated it. The dataset was based on the British Standards Codes of Practice for construction in addition to Procurement Practices.

Regarding the annotation scheme, the BIO2 tagging scheme was adopted. The BIO2 tagging scheme uses three tags: B (for beginning), I (for inside), and O (for Outside). Also, several entities were identified, as shown in table 1.

Table 1: Entities

SEC	Section
MAT	Material
PROP	Property
PROC	Process
STRT	Structure
LOC	Location
WTH	Weather
FET	Feature
TEMP	Temperature
METH	Method
STD	Standard
COMP	Competence
SPEC	Specification
MES	Measurement

Additionally, the dataset was increased by seven times through the application of data augmentation techniques. Data augmentation is a method for making modified versions of already-existing data to enhance the size of a dataset. The following approaches were employed in this research:

- Synonym replacement: replacing synonyms for certain terms in a dataset. This can help to improve the model's resilience by making it more tolerant to changes in the input data.
- Random word swapping: sometimes known as word shuffling, this technique incorporates ad hoc word shifting inside a sentence.
- Random phrase swapping: rearranging phrases inside a text or document at random.
- Random insertion entails randomly inserting new data into a dataset.

The capacity to artificially increase the size of a dataset by adding more sentences or words to a text can be useful for a variety of machine learning applications, particularly in NLP.

3.6 KNOWLEDGE GRAPH CONSTRUCTION
3.6.1 KNOWLEDGE FUSION

Knowledge fusion is an effective method for merging data from many sources to make more accurate and insightful predictions [11]. It blends the strengths of several algorithms and data sources to provide more precise and dependable forecasts. For example, knowledge fusion may be used to combine a machine learning model's prediction with insights from a domain expert to provide more accurate forecasts. Knowledge fusion may also be used to combine the capabilities of multiple algorithms, such as merging deep learning algorithm predictions with simpler linear model predictions [12]. There are several knowledge sources in this project. Therefore, to construct the KG, knowledge fusion techniques had to be used. Figure 10 presents the general knowledge fusion process.

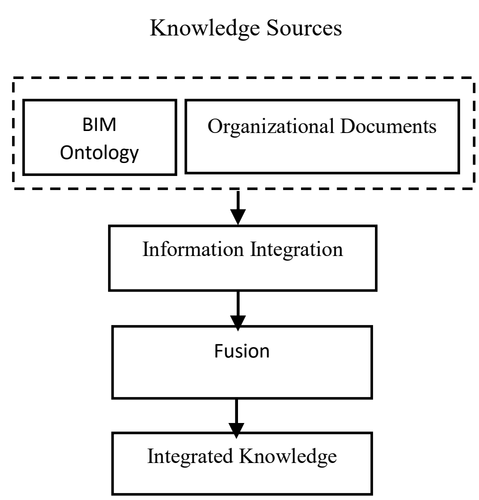

Figure 11: Knowledge Fusion

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

There are two basic sources of knowledge: data from organizational papers and the BIM ontology. Before being fused, the information from these sources is first merged. Domain fusion, attribute fusion, and relationship fusion are all carried out throughout the fusion process. The knowledge integration and fusion algorithm is presented in the table below.

Table 2: Knowledge Fusion algorithm

Knowledge Fusion Algorithm

```
function updateKnowledgeGraph(KG, keywords,
definitions):
  for i = 0 to length(keywords) do
    keyword = keywords[i]
    definition = definitions[i]
    if not keywordExistsInGraph(KG, keyword) then
      addNodeToGraph(KG, keyword, definition)
    else if not informationExistsForKeyword(KG,
keyword) then
      updateDefinitionInGraph(KG, keyword,
definition)
  return KG
function keywordExistsInGraph(KG, keyword):
  return keyword in KG["Keyword"]
function informationExistsForKeyword(KG,
keyword):
  keywordIndex = findIndexForKeyword(KG,
keyword)
  return KG["Definition"][keywordIndex] != ""
function addNodeToGraph(KG, keyword, definition):
  KG["TreeID"].append(generateUniqueID())
  KG["Primary Keyword"].append(keyword)
  KG["Keyword"].append(keyword)
  KG["Definition"].append(definition)
function updateDefinitionInGraph(KG, keyword,
definition):
  keywordIndex = findIndexForKeyword(KG,
keyword)
  KG["Definition"][keywordIndex] = definition
function findIndexForKeyword(KG, keyword):
  return index in KG["Keyword"] where
KG["Keyword"][index] == keyword
```

3.6.2 KNOWLEDGE STORAGE

The storage format of Neo4j is based on the property graph concept where data is represented as nodes, relationships, and attributes. Relationships are node connections, whereas attributes are key-value pairs that provide additional information about nodes and relationships. In Neo4j, each node and connection have a unique identity (ID) that allows for quick and efficient data retrieval.

4. RESULTS AND DISCUSSION

The knowledge graph is built following a methodical process of data integration and knowledge fusion. Initially, 5 data sources including organizational documents and existing ontologies are gathered to represent various aspects of the AEC industry. These sources include material information, legislation, and project data, among others. Semantic heterogeneity is also addressed by mapping and alignment to ensure consistency in expressing entities and their connections. The knowledge integration approaches then combine this disparate information to build a cohesive graph that connects various entities.

The KG constructed is shown in figure 12.

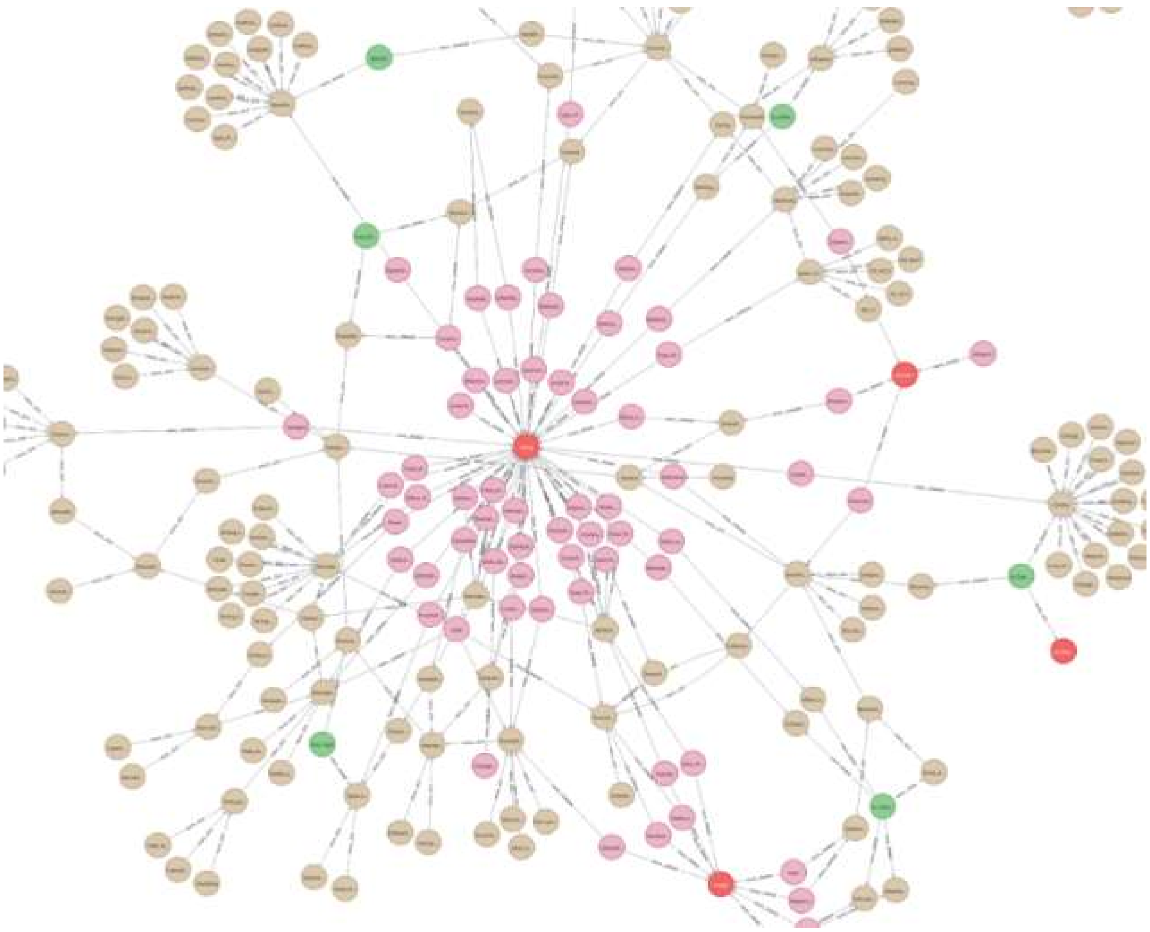

Figure 12: The Knowledge Graph

A KG tailored for the AEC industry has a lot of potential for multifaceted applications as it offers a comprehensive and unified overview of the industry. The integration of diverse knowledge sources empowers stakeholders and supports them in making informed strategic decisions. Through the used algorithms, patterns and trends within the industry can be identified, which can be very beneficial in shaping future strategies. In this sense, predictive analysis can gauge the potential repercussions of regulatory changes or shifts in material availability, for instance, and uncover opportunities or threats that might otherwise remain obscured. Moreover, the KG proves instrumental in managing project data efficiently. By acting as a repository, it adeptly organizes essential project information. This centralized storage and accessibility streamline project execution and contributes to heightened efficiency and reduced potential for costly errors. Lastly, the KG fosters collaboration and communication within the industry. By interlinking relevant stakeholders, it expedites information sharing and decision-making processes. Consequently, this collaborative environment ensures effective and efficient project completion, minimizing delays and enhancing project outcomes.

In steering through the complexities of this research, where ontologies, deep learning models, and KGs converge, the critical significance of construct validity is very important. It's imperative to ensure that the used ontologies authentically capture the subtleties of the domain, while the deep learning models adeptly encompass the specified features. Recognizing the vital role of external validity, the study conscientiously evaluates the transferability of its findings beyond the immediate context and assessing their applicability in broader domains or real-world scenarios. The careful consideration of these validity concerns collectively affirms the robustness and reliability of the research outcomes.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

The construction of the KG presented certain limitations, particularly concerning semantic heterogeneity across diverse data sources and ontologies. To achieve consistency in the representation of entities and relationships required meticulous mapping and alignment efforts, in addition to expert feedback. While data integration techniques contributed to enhancing data quality, inherent limitations persisted due to the presence of noisy and incomplete data in the source documents.

Addressing these limitations necessitates the development of more sophisticated data cleaning and validation approaches. Moreover, as the KG expands over time and incorporates additional data sources, maintaining its integrity and coherence becomes increasingly complex. This will also necessitate regular updates and revisions to ensure alignment with the most current state of knowledge. Additionally, building and managing a large-scale KG demand substantial computational resource, making it a resource-intensive endeavor. Considering this, future research should explore optimizations aimed at reducing computational overhead to enhance the efficiency of the KG construction and maintenance processes.

5. CONCLUSION AND FUTURE WORKS

This research article presents the construction and implementation of a KG by integrating diverse data sources, with a specific focus on the construction industry, and more specifically BIM. The goal is to pioneer a new method of utilizing IT technologies like ontologies and KGs in the construction sectors. A complete framework for building the KG is outlined and explored, explaining the step-by-step approach followed. A validation of the ontologies by domain experts was conducted prior to their validation. Moreover, a test implementation of the KG was carried out using Neo4j to confirm the feasibility and functionality of the proposed method. This test implementation enabled the KG to be evaluated and fine-tuned before deploying it.

This paper provides a foundation for furthering the integration of KGs in the AEC sector. Future work might concentrate on a variety of issues to improve the usability and breadth of the KG. To begin, broadening the data sources to incorporate a wider variety of construction-related information will improve the knowledge representation of the KG. Second, investigating machine learning approaches for entity recognition and relationship extraction might automate and scale up the KG creation process. Incorporating semantic reasoning skills into the KG can also enable it to execute more complex inference and analytics, allowing it to get deeper insights from the interrelated data. Moreover, efforts should be dedicated to promoting collaboration and data sharing among stakeholders to continuously enrich the KG and keep it up to date with the latest industry developments. Lastly, evaluating the KG's practicality and effectiveness in real-world construction and procurement scenarios will be pivotal in validating its potential for transformative impact in these industries.

REFERENCES

[1] IBM. (n.d.). What is a knowledge graph? IBM. https://www.ibm.com/topics/knowledge-graph
[2] Zou, X., 2020. A Survey on Application of Knowledge Graph. Journal of Physics: Conference Series, 1487(1), p.012016.
[3] Sarshar, M., Memari, A. M., and Amor, R. (2018). "Graph-Based Semantic Mapping for Information Exchange in the AEC/FM Industry." Journal of Computing in Civil Engineering, 32(4), 04018014. DOI: 10.1061/(ASCE)CP.1943-5487.0000753
[4] Wang, C., Ma, X., Chen, J., & Chen, J. (2018). Information extraction and knowledge graph construction from geoscience literature. Computers & geosciences, 112, 112-120.
[5] Zhang, Y., Chen, P., Zhang, M., and Zhang, X. (2021). "A Review of Building Information Modeling and Knowledge Graph Applications in Construction Industry." Frontiers in Built Environment, 7, 678985. DOI: 10.3389/fbuil.2021.678985
[6] Mahdavinejad, M., Zarli, A., and Janssen, J. (2021). "An Ontology-Based Building Information Model to Facilitate Knowledge Management in the Architecture, Engineering, and Construction Industry." Automation in Construction, 124, 103508. DOI: 10.1016/j.autcon.2020.103508
[7] Jiang, Y., Zhao, X., Huang, Y., & Xie, X. (2021). An ontology-based approach for integrating heterogeneous BIM data sources using knowledge graphs. Automation in Construction, 127, 103803.
[8] Li, J., Fan, Y., Huang, T., & Zhang, L. (2021). A BIM-based knowledge graph framework for building performance analysis. Automation in Construction, 126, 103742.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

[9] Zhang, L., Luo, H., Lu, Y., & Shen, W. (2019). A knowledge graph-based approach to facilitate BIM data integration and retrieval. Automation in Construction, 102, 109-119.

[10] Zhao, X., Zhang, Z., Ma, S., & Wang, J. (2021). A knowledge graph-based approach for risk identification and decision support in BIM-based construction project management. Automation in Construction, 129, 103873.

[11] Dong, X.L. et al. (2015) 'From Data Fusion to Knowledge Fusion', Proceedings of the VLDB Endowment, 7(10), pp. 881–892. doi:10.14778/2732951.2732962.

[12] Zhen, M. et al. (2023) 'Application of a fusion model based on machine learning in Visibility Prediction', Remote Sensing, 15(5), p. 1450. doi:10.3390/rs15051450.