A hybrid artificial intelligence to support information retrieval in smart buildings

Massimo Callisto De Donato[1,*,†], Emanuele Laurenzi[2,*,†] and Daniele Porumboiu[1,†]

[1] *University of Camerino, Computer Science Division, Camerino, Italy*
[2] *FHNW - University of Applied Sciences and Arts Northwestern Switzerland, Olten, Switzerland*

Abstract

The Smart Building paradigm promises a future where buildings are intelligent, adaptive, and sustainable, offering real-time information retrieval that supports decision-making to enhance energy efficiency, occupant comfort, and security. However, achieving such a paradigm is highly complex, one major reason being the seamless integration of (a) physical and functional representations of buildings (i.e., Building Information Modeling) and (b) real-time IoT (i.e., Internet of Things) data. As a contribution to this challenge, we propose a hybrid Artificial Intelligence approach where, on the one hand, a knowledge graph retains the buildings' knowledge structures and a time-series database holds IoT data. On the other hand, a Large Language Model serves as a mediator between a facility manager and the knowledge graph and IoT data, facilitating data-driven decision-making processes. The approach has been developed through the Design Science Research methodology. The evaluation was carried out through a technical prototype that instantiates the novel approach and proves its feasibility.

Keywords

smart building, building information modeling, internet of things, hybrid ai, knowledge graphs, large language model

1. Introduction

Building Information Modeling (BIM) and the Internet of Things (IoT), while powerful individually, achieve a new level of significance when their inherent strengths are combined [1]. BIM strives to provide a comprehensive digital representation of a building's physical and functional characteristics, offering a centralized repository of information throughout the building's lifecycle, from design to demolition. IoT, on the other hand, furnishes a network of connected sensors and devices that generate real-time data on a building's operation and environment. When integrated, this confluence of rich design data and dynamic operational data enables unprecedented insights. Domain experts like facility managers can optimize energy consumption based on actual usage patterns, predict maintenance needs based on sensor readings, and enhance occupant comfort through automated adjustments, leading to substantial cost savings and improved building performance. However, integrating BIM and IoT remains challenging. Because they evolved separately, differences in data formats, communication systems, and interoperability hinder data sharing and collaboration between the fields [1]. Overcoming these issues is crucial to realizing the full potential of a connected, data-driven approach in construction and building management. This challenge is often exacerbated by the fact that most IoT devices and sensors are deployed during the operational phase, while they are not typically integrated into the initial BIM models created during the design phase [2]. Consequently, real-time data produced by IoT systems (such as energy usage, environmental conditions, or equipment status) frequently remains excluded from the BIM models. This exclusion limits the ability to utilize BIM as a dynamic, evolving tool that accurately reflects the building's performance and operational conditions throughout its lifecycle.

BIR-WS 2025: BIR 2025 Workshops and Doctoral Consortium, 24th International Conference on Perspectives in Business Informatics Research (BIR 2025), September 17-19, 2025, Riga, Latvia.

*Corresponding author.
† These authors contributed equally.
massimo.callisto@unicam.it (M. Callisto De Donato); emanuele.laurenzi@fhnw.ch (E. Laurenzi); daniele.porumboiu@unicam.it (D. Porumboiu)
0000-0001-6745-4785 (M. Callisto De Donato); 0000-0001-9142-7488 (E. Laurenzi)

As we report in Section 2, while recent works strive to address the challenge of integrating BIM with IoT, they heavily rely on semantic technologies like ontologies and knowledge graphs. These have the advantage of overcoming the interoperability issue among data models, but their use for effective information retrieval still necessitates substantial engineering to develop complex interfaces that aid stakeholder decisions during building operation. In contrast, this paper proposes a hybrid AI approach that combines a structured knowledge graph for BIM with IoT-sensed data, leveraging a Large Language Model to improve information retrieval and improving decision-making for stakeholders during the building's operational phase.

The remainder of this paper is as follows. Section 2 describes the related work and motivates the research question. Section 3 introduces the Design Science Research (DSR) as the followed methodology. Section 4 describes the proposed hybrid AI architecture. The proof of concept and the discussion on the current limitations are elaborated in Sections 5 and 6, respectively. Finally, Section 7 summarizes and concludes the paper.

2. Related work

Recent literature reports significant efforts in integrating the two paradigms BIM and IoT for various purposes, with the tendency to leverage ontologies and knowledge graph technologies. In [3], the authors introduced a cross-source data management and analysis framework to support evacuation path planning and emergency response decisions in fire scenarios, supported by a specialized FireEvacuation ontology. The work in [4] proposed a service-oriented architecture for data-driven smart buildings, utilising semantic technologies as an integral part of the architecture, essential for adding context to operational data and creating links between diverse systems. The approach in [5] proposed a flexible energy modeling framework based on the SAREF ontology and its SAREF4BLDG extension. It offers models for typical systems and devices, and a method for linking and simulating components using also the SAREF4SYST extension. Researchers in [6] presented a framework for integrating BIM and IoT data using an ontology-based mediation mechanism. It enables integrated access to local BIM and IoT data through query-rewriting processes. The paper [7] introduced the Building Topology Ontology (BOT), which provides a high-level description of the topology of buildings including storeys and spaces, the building elements they contain, and their web-friendly 3D models. They also describe how existing applications produce and consume datasets combining BOT with other ontologies that describe product catalogues, sensor observations, or Internet of Things (IoT) devices. [8] demonstrates how the integration of BIM and IoT data can be used to monitor the indoor environmental quality of a building. With their approach, they were able to query the topology, static, and dynamic properties from a graph database and then query the corresponding sensor data from a time-series database. The work in [9] conducted a review of the main ontologies and applications that support the development of Decision Support Systems and decision making in the different phases of a building's life cycle. This study also highlighted that most ontologies lack real-life applications and some applications are focused mostly on the design phase of a building or its early operation, indicating their early development stage. Researchers in [10] designed an ontology, called Building Performance Ontology (BOP), that integrates topological building information with static and dynamic properties for improving the monitoring of indoor environments. Authors in [11], introduced a novel multi-layer architecture and a comprehensive framework for smart-building digital twins, with a primary focus on enabling semantic interoperability among smart-building digital twin applications. The approach provides a semantic static (BIM) and dynamic (IoT) building data that satisfy the real-time data requirements of smart-building digital twins while preserving IoT data in its optimal time-series data storage. Similarly, in [12] the authors showcased the integration of construction documentation, facility management records, and real-time data obtained from building automation systems within a Cognitive Digital Twin. A W3C-compatible approach was created, drawing from the BOT ontology and integrating it with the Brick Ontology. [13] developed a Digital Twin using a micro-service architecture, which facilitates cloud deployment and enables modularly defined functionalities. The knowledge graph acts as the contextual interface that provides a comprehensive view of all data and all models. The semantic information is stored in the Neo4j database and structured as a property graph, following the concepts and relations defined in the IFC schema. [14] constructed a general City Information Model ontology to integrate heterogeneous building information modeling (BIM), geographic information system (GIS) and IoT data. A new ontology has been developed (BIM-GIS Integration Ontology) and mapped with the Brick and SSN ontologies. In [15], the researchers described a realization of a Semantic Digital Twin through the use of modular knowledge graphs instead of using monolithic graph architectures. The advantage of the approach lies in the possibility to merge independently developed knowledge graphs into a single one that is easier-to-understand, better to reason with, and also reusable. In addition, when integrated with real-life systems, modular graphs improve performance by loading only the needed segments, eliminating problems with querying and reasoning in large stores. Although semantic-based approaches have advantages, information retrieval presents a common challenge, demanding either substantial expertise in ontology design for complex SPARQL queries or the development of integrated, user-friendly interfaces.

Recent advances in Information Systems suggest combining the strengths of both Knowledge Graphs and Large Language Models to address these limitations [16, 17]. Such a combination falls into the research realm of Hybrid Artificial Intelligence, because it considers two approaches from the two sides of AI: the Symbolic AI (i.e., KG) and Sub-symbolic AI (i.e., LLM). Therefore, the research question we investigated in this work is the following: *How effective is a hybrid artificial intelligence approach that combines knowledge graphs and LLMs to integrate BIM and IoT data for supporting information retrieval by domain experts?* As mentioned in the introduction, facility managers are examples of domain experts in Smart Buildings.

3. Methodology

The methodology followed to answer the research question of this work is the Design Science Research (DSR), which proposes five main phases [18]: problem awareness, suggestion, development, evaluation and conclusion.

During the *problem awareness phase*, we aim to deepen the understanding of the problem from both research and application perspectives. Based on our literature review, we found the architecture proposed in Donkers et al. [8] (for real-time building performance monitoring using semantic digital twins) to be the closest to our problem, therefore we decided to build our approach from it. Their architecture integrates knowledge structures of BIM and IoT data to monitor the indoor environmental quality (e.g. air quality index) of a building. Specifically, they used an Open Smart Home (OSH) dataset [19] conformed to the BOT [7] and BOP ontologies [10] and a built a custom-built Python component to navigate the knowledge graph, pick a property (e.g., temperature, humidity, and illuminance), and return the corresponding value from a time-series database for an overall calculation.

Differently from [8], we replaced the static Python component with an LLM-based component, to promote scalability of the approach.

From an application point of view, we focused on the same OSH dataset and analysed it using the developed prototype.

This dataset was developed by the Fraunhofer Institute for Building Physics in Nuremberg, Germany [20] and provides both static and dynamic aspects of a real smart home environment and is intended to support investigations into energy efficiency, control strategies, and building performance analysis. The OSH scenario represents a two-story building; the ground floor is shown in Fig. 1. This floor has a bathroom, kitchen, lobby, and toilet, each enclosed by four walls and a ceiling. Walls include doors, windows, and sensors. A gas boiler provides hot water to radiators in all rooms except the lobby. Windows have manual shutters for shading.

Static data is available in IFC and Revit formats, and also in RDF and makes use of concepts from the BOT and PROPS ontologies. Dynamic data, representing sensor readings, is provided in CSV and RDF formats, employing the SSN/SOSA ontologies. Measurements from the sensors span the period from March 9, 2017, to June 6, 2017, with a variable sampling rate up to 15 minutes. The smart home is

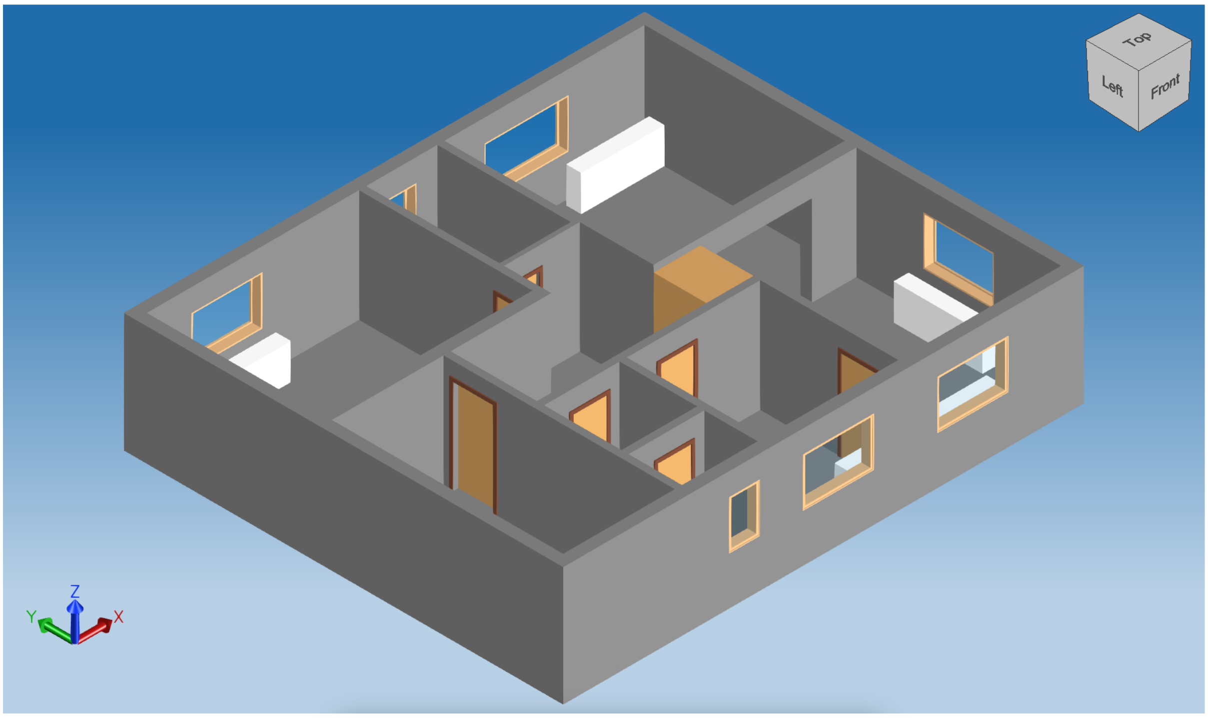

Figure 1: Rendering of first building level from Open Smart Home Data [19].

equipped with a system offering the following capabilities:

- Wall-mounted sensors in rooms with space heaters (not in the staircase or lobby), measuring air temperature, illuminance, and humidity.
- Remote-controlled thermostat valves on each heater, logging the setpoint and local air temperature.
- A base station that communicates wirelessly with sensors and actuators, connects to the internet for weather forecasts, and provides a virtual outdoor temperature per room.
- A smartphone application for controlling setpoints, scheduling, and monitoring real-time measurements.

In the *suggestion phase*, we proposed a novel hybrid AI architecture where an LLM enables the integration of concepts from a knowledge graph and respective IoT values from a time-series database. Therefore, we leveraged the architecture in [8] with he same dataset to inform our suggested artifact.

In the *implementation phase*, we implemented the proposed architecture in a technical prototype, where the user interface takes the form of a virtual assistant or bot.

Finally, in the *evaluation phase*, we integrated the OSH RDF-based dataset in the technical prototype and then answered competency questions that a facility manager would find helpful. The strategy of answering competency questions is a well-known evaluation technique in ontology engineering [21]. The competency questions have been derived based on the presented dataset and the work of Donkers et al. [8]. An example for such a competency question is the following: *For thermal performance and maintenance purposes, it is important to know the material composition, the structure and the dimensions of specific walls.*

4. The hybrid AI architecture

In this section, we describe the proposed hybrid AI architecture that combines knowledge graph and LLM capabilities to support information retrieval for domain experts. The resulting architecture is reported in Fig. 2.

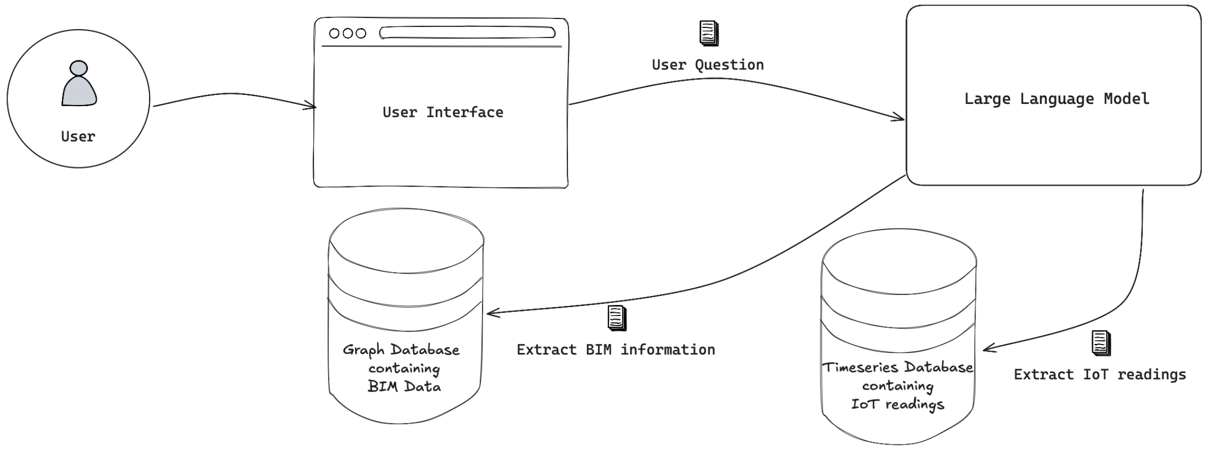

Figure 2: Overview of the proposed hybrid AI architecture, integrating an LLM with a knowledge graph and a time-series database to support users in retrieving information from BIM and IoT data.

The hybrid AI architecture relies on a **knowledge graph** to semantically represent the structure of both the BIM model and the IoT infrastructure. Accordingly, the selection of the appropriate ontologies is essential to enable knowledge-driven smart buildings. Following the findings from Donkers et al. [8], we selected the BOT and the BPO ontologies. The BOT ontology is used to describe the physical structure of the building, including elements such as spaces, walls, and their topological relationships, making it suitable for modeling the knowledge structure of the building. The BPO ontology, on the other hand, is employed to represent knowledge information about sensor and actuator data. This combination allows us to create a more comprehensive model that can represent the integration of BIM and IoT information effectively. The combined semantic information including building structure and IoT infrastructure information are stored into RDF format using the knowledge graph.

While a graph database model is well suited for storing entities and their semantic relationships in RDF triple stores, it does not scale efficiently when dealing with large volumes of historical IoT data [22]. To handle the actual IoT sensor data, a **time-series database** is used as a complementary component to the knowledge graph [23]. Its role is to efficiently store and manage the sensor readings over time, such as temperature, humidity, and light levels. The database enables fast retrieval of the most recent sensor values, supports aggregation queries over historical data, and allows for horizontal scalability when managing large volumes of IoT measurements. The linkage between the two layers is established via *unique sensor identifiers* that are consistently represented in both the knowledge graph and the time-series database.

The combination of the knowledge graph and the time-series database results in two distinct data storage systems, typically each with its own query such as SPARQL for semantic information retrieval from the knowledge graph, and dedicated NoSQL-like query languages for accessing the time-series database [24]. Moreover, having two separate data stores requires the user to define additional aggregation mechanisms, potentially increasing the complexity of retrieving meaningful insights about the building.

To overcome this limitation, we introduce the **LLM** component with a retrieval-augmented generation capability [25], which provides uniform access to both data sources. The LLM acts as a mediator in the information retrieval process, being enhanced with external knowledge retrieved from both the knowledge graph and the time-series database. We supplemented the LLM with a dedicated user interface, which acts as a **virtual assistant** capable of understanding and responding to natural language queries from domain experts. This assistant enables users to ask complex questions about building components, their properties, and sensor readings, thus providing a more intuitive and accessible approach to information retrieval tasks related to the knowledge structure of the building and IoT data.

5. Proof of concept

In this section, we describe both (1) the technical prototype that implements the suggested hybrid AI architecture, and (2) the evaluation of the approach's effectiveness. For the latter, we used the competency questions that, in natural language, have been added as prompts to the virtual assistant. The results have been compared qualitatively with the OSH dataset, which is the ground truth. The source code for the implementation is available on GitHub[1].

5.1. Artifact development

For storing the BIM and IoT knowledge structures, we used **GraphDB** by Ontotext,[2] a semantic graph database compliant with W3C standards and designed to store and manage RDF data. GraphDB is fully compatible with widely used ontology standards and supports the SPARQL query language, which enables users to retrieve and manipulate data stored in the database.

For storing the IoT data, we used the time-series database **InfluxDB** [3] by InfluxData. InfluxDB is optimized for fast ingestion, querying, and aggregation of time-series data, making it well-suited for monitoring, IoT, and real-time analytics applications. InfluxDB provides *Flux*, a powerful functional query language designed explicitly for time-series workloads. It supports data manipulation, statistical analysis, joins, and integration with external systems.

For the LLM component, we integrated **Claude 3.5 Sonnet**[4], developed by Anthropic, using the standard cloud API. We selected Claude over other LLMs because, at the time of implementation, it demonstrated adequate results in graph query generation tasks [26]. Claude allows us to interpret natural language queries from users by identifying the underlying intent behind questions related to the knowledge structure of the building and IoT sensor data. It then generates appropriate queries for both the knowledge graph and the time-series database, processes the raw results, and produces human-readable explanations.

To integrate LLM and data sources, we employed **LangChain**[5], a popular Python-written framework designed to facilitate the development of LLM-driven applications. LangChain is based on three core components: *Chains* which represent deterministic sequences of steps to handle user inputs such as prompting, parsing, and transformation; *Tools* which are custom Python modules used to integrate the external systems through APIs and allow the LLM to obtain augmented information used to enrich the responses; *Agents* which use the LLM to reason over tasks and dynamically decide which Tools or Chains to invoke based on the available context.

We used the Chain *OntotextGraphQAChain* from LangChain to interact with GraphDB. We configured the Chain to automatically generate SPARQL queries using Claude based on the input and retrieve relevant building information from GraphDB. This is done by passing to the Chain the ontologies as the input schema of the database and a set of prompts to instruct the LLM. Listing 1 shows an excerpt from the prompt we defined to query the knowledge graph. The full set of prompts used in the evaluation is available on the project's GitHub repository.

```
You are an expert GraphDB Developer translating user questions into SPARQL to answer questions about a building and the elements contained in it. Use only the provided relationship types and properties in the schema.
Do not use any other relationship types or properties that are not provided.

Your answers should be concise and to the point. Do not include any additional information that is not requested. Answer with only the generated SPARQL statement.
Try to use meaningful aliases for the nodes and relationships in the query. Here there are some examples of how to respond to the user's question:
```

[1] BIM-IoT-Assistant: https://github.com/PROSLab/BIM-IoT-Assistant
[2] GraphDB: https://www.ontotext.com/products/graphdb/
[3] InfluxDB: https://www.influxdata.com/
[4] Claude: https://www.anthropic.com/claude
[5] LangChain: https://www.langchain.com/

```
<example>
Tell me about the bathroom in the building
PREFIX bot: <https://w3id.org/bot#>
PREFIX props: <https://w3id.org/props#>
SELECT ?room ?relationship ?value
WHERE {{
  ?room props:longNameIfcSpatialStructureElement_attribute_simple  ?name.
  FILTER(CONTAINS(?name, "Bathroom"))
  ?room ?relationship ?value
}}

What can be found in the kitchen?
PREFIX default1: <https://w3id.org/bot#>
PREFIX default2: <https://w3id.org/props#>
SELECT ?space ?element
WHERE {{
  ?space default2:longNameIfcSpatialStructureElement_attribute_simple ?name.
  FILTER(CONTAINS(?name, "Kitchen"))
  ?space default1:containsElement ?element.
}}
....
```

Listing 1: An excerpt from the prompt used to generate SPARQL queries over GraphDB.

To retrieve the data from the IoT sensors, we developed a Tool that executes a custom Chain used by the LLM to retrieve the sensor identifier from the knowledge graph. The LLM then, informed with the identifier, queries InfluxDB to retrieve the actual time-series values. An Agent orchestrates these two components by driving user queries based on the inputted context.

Finally, the web user interface has been developed using the Streamlit[6], a Python library that supports the user in the interaction with the virtual assistant. The web application implements a chat-like interface through which users can submit queries in natural language and receive responses.

5.2. Artifact evaluation

To evaluate the effectiveness of the proposed approach, we used the prototype. Namely, we first imported the OSH dataset into GraphDB using its RDF representation, which conforms to the BOT and BOP ontologies. To import the IoT sensor data into InfluxDB, we adopted a strategy similar to that described in [8]. That is, we first converted the input CSV files into the InfluxDB line protocol format and then imported them into the time-series database.

We configured the Chain to generate the SPARQL automatically and we passed both the BOT and BOP ontologies, in TTL format. However, this configuration caused the input to exceed the 200,000-token limit of Claude 3.5 Sonnet during the query generation. To reduce the number of input tokens, we then passed only the instances and respective relationships of the dataset, thus omitting the schema. This reduced the total token count to under 60,000 and allowed us to define a set of prompt examples to guide the LLM in generating the appropriate SPARQL queries.

We applied a similar strategy in the Chain we developed to retrieve the information about IoT data. However, Claude consistently failed to generate SPARQL queries to extract the sensor identifiers from the GraphDB, which was required to then query InfluxDB. More in detail, the LLM attempted to retrieve sensor measurements directly from GraphDB rather than the sensor identifiers. We believe that this behavior may be attributed to the excessive amount of contextual information provided, which likely caused the model to misinterpret the user's intent.

To address this issue, we modified the custom chain to use only the schema, i.e., classes, object properties, and data properties. The corresponding SPARQL query is shown in Listing 2. This solution significantly reduced the amount of contextual information passed to the LLM, lowering the input to

[6]Streamlit: https://streamlit.io/

fewer than 5,000 tokens and enabling the generation of SPARQL queries that correctly extracted the sensor identifiers.

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
CONSTRUCT {
    ?class a rdfs:Class .
    ?class ?objectProperty ?relatedClass .
    ?class ?dataProperty "" .
}
WHERE {
    {
        # Extract classes
        SELECT DISTINCT ?class
        WHERE {
            ?instance a ?class .
        }
    }
    {
        # Extract object properties for each class
        SELECT DISTINCT ?class ?objectProperty ?relatedClass
        WHERE {
            ?instance1 a ?class.
            ?instance2 a ?relatedClass.
            ?instance1 ?objectProperty ?instance2.
        }
    }
    {
        # Extract data properties for each class
        SELECT DISTINCT ?class ?dataProperty
        WHERE {
            ?instance a ?class.
            ?instance ?dataProperty ?dataValue.
            FILTER NOT EXISTS { ?dataValue a ?moreclass.}
        }
    }
}
```

Listing 2: SPARQL query used as the schema in the custom Chain to extract sensor identifiers.

As a result of the adjustments described above, the virtual assistant was able to respond to both building-related and sensor-related queries successfully.

In accordance with the research methodology, we present a set of competency questions used to evaluate the effectiveness of the approach. Specifically, we engaged with the virtual assistant by asking the competency questions. In doing this, we simulated the interaction of a facility manager seeking information about the building structure and the data generated by the IoT infrastructure. Most of the competency questions have been answered correctly. In the following we described a few of the competency questions that have been correctly answered. In section 6, we show the only competency question that the virtual assistant failed to answer. The list of the competency questions is available on the project's GitHub repository.

For example, when we asked for information about the rooms located on the first floor, the assistant correctly listed all the rooms present in the dataset (see Fig. 3). A similar result was obtained when we requested information about a specific room, such as the kitchen. The virtual assistant accurately retrieved the elements contained within it, including the instances of the IoT sensors (see Fig. 4).

Another example involves retrieving information about the material composition of the building structure. In this case, when we asked for details about a specific wall, the virtual assistant successfully returned comprehensive information. The virtual assistant successfully retrieved information about the

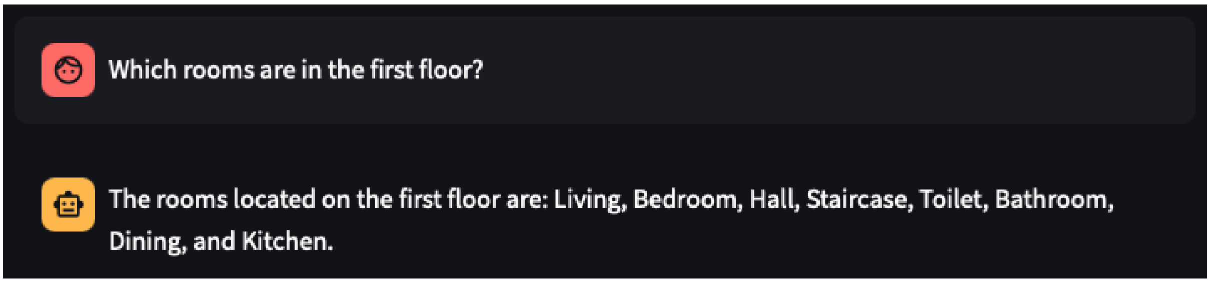

Figure 3: Answer to the user query "List the rooms located on the first floor".

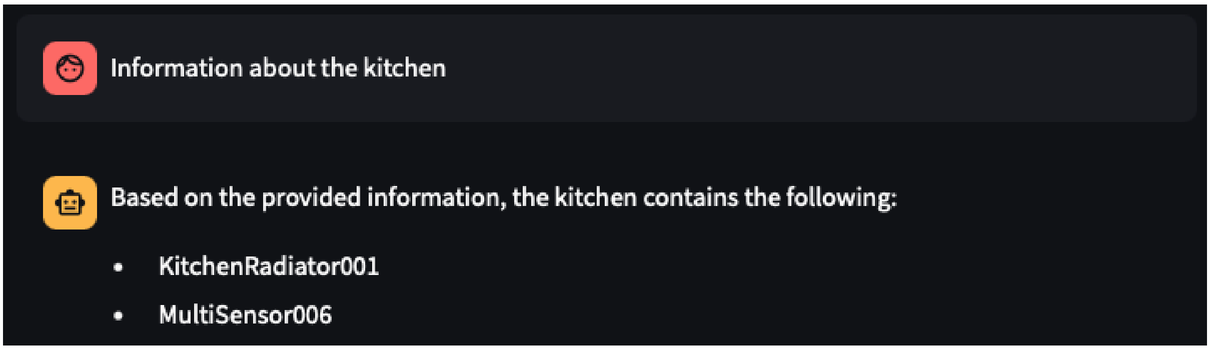

Figure 4: Answer to the user query "Information about the kitchen room".

IoT infrastructure, providing the user with the sensors installed on the wall (see Fig. 5).

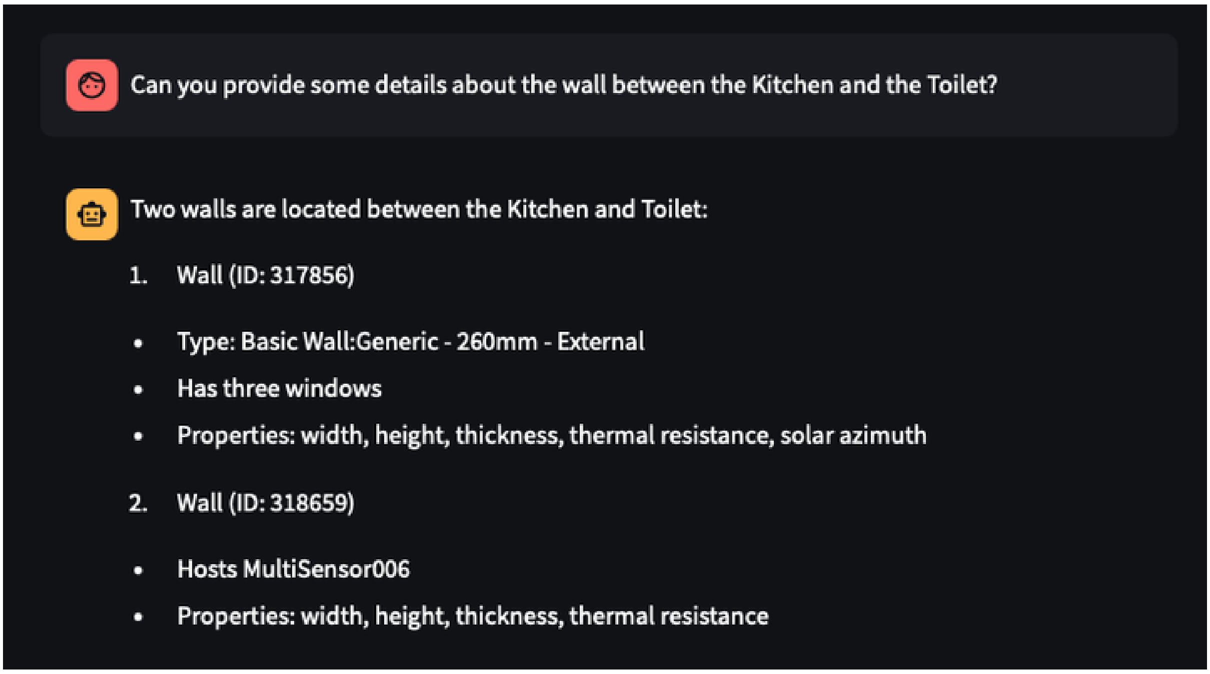

Figure 5: Answer to the user query "Details of a specific element of the building structure".

We then asked the assistant to provide information about the values recorded by a specific sensor installed on a wall. As the OSH dataset does not include real-time data, we specified a time range in 2017 for which measurements were available. Once again, the virtual assistant successfully responded to the query, providing the user with the temperature measurement alongside the time stamp of the observation (see Fig. 6).

Figure 6: Answer to the user query "Which is the lowest temperature registered in the kitchen?"

6. Discussion and limitations

Given the positive output in the evaluation, we regard the proposed hybrid AI architecture as a promising approach to overcome the integration issues between BIM and IoT in a scalable way. In fact, the virtual assistant is able to interact with the user, interpret their questions, generate appropriate queries for the databases, retrieve the data, and present the results to the user in a natural language.

The evaluation proved that the virtual assistant is capable of answering the majority of the queries. It showed some difficulties to retrieve information that involve relationships and hierarchies. For instance, when we asked questions involving building structures such as identifying rooms adjacent to another room, the assistant failed to answer meaningfully (see Fig. 7). We observed that this behavior stems from the way the building structure is represented: rooms are not directly connected to one another, but are instead linked through intermediary elements such as walls and slabs. However, in all the generated queries, the LLM attempted to retrieve elements directly adjacent to the room, thereby overlooking the actual structural hierarchy.

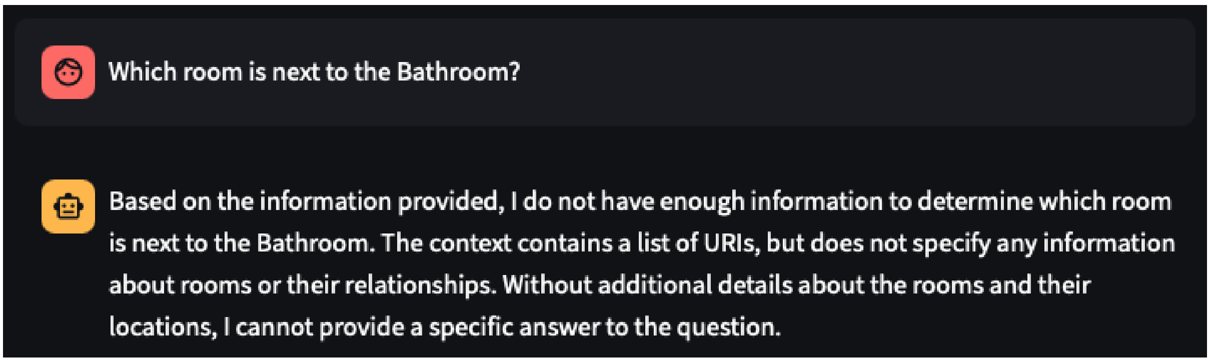

Figure 7: Answer to the user query "Rooms adjacent to another room".

To address these issue, we are currently investigating the design of additional prompts to help the LLM reason over such fine-grained representations of building topology.

Furthermore, we observed that accessing the knowledge structure of the building stored in GraphDB often requires the LLM to execute multiple SPARQL queries to retrieve certain properties. This is due to the granular representation of the knowledge structure in GraphDB, which makes the information retrieval process more complex. As a consequence, the LLM must be instructed to generate more complex SPARQL queries to retrieve these properties, which involves providing additional examples in the prompt and increases the overall complexity of the query generation process.

7. Conclusion

In this paper, we proposed a novel hybrid artificial intelligence approach that combines a structured knowledge graph for BIM with IoT-sensed data, leveraging both semantic technologies and Large Language Models (LLMs).

We adopted a Design Science Research strategy to address the research question of how such a combination can support information retrieval for domain experts. In result, the proposed architecture has been implemented in a technical prototype, which takes the form of a virtual assistant.

The open OSH RDF-based dataset has been selected as a suitable real-world case for a BIM and IoT integration. This dataset served as a ground truth in the evaluation of the proposed architecture. In result, the approach proved to be effective as most of the competency questions have been correctly answered.

Future works are multiple. Firstly, it would be interesting to replicate the evaluation by considering other datasets as ground truth. We also plan to integrate and compare the performance of other LLM models to assess the effectiveness of our prototype. Next, given the evolution of LLMs towards reasoning capabilities, as future work, we aim to go beyond information retrieval and leverage the LLM to make calculations that can be regarded as useful by facility managers such as the indoor environmental quality of the building.

Another interesting future direction concerns the use of Labeled Property Graphs (LPGs), which have already been applied in similar contexts [27]. The goal is to assess whether LLMs can provide comparable support when used with LPGs instead of RDF-based knowledge graphs. This investigation will also raise questions regarding the effectiveness of LPGs in addressing some of the limitations discussed in this paper.

Acknowledgments

This work has been partially supported by the European Union – NextGenerationEU - National Recovery and Resilience Plan, Mission 4 Education and Research - Component 2 From research to business - Investment 1.5, ECS_00000041-VITALITY - Innovation, digitalisation and sustainability for the diffused economy in Central Italy - CUP J13C22000430001.

Declaration on Generative AI

The author(s) have not employed any Generative AI tools.

References

[1] X. Huang, Y. Liu, L. Huang, E. Onstein, C. Merschbrock, BIM and IoT data fusion: The data process model perspective, 2023. doi:10.1016/j.autcon.2023.104792.

[2] L. Chamari, E. Petrova, P. Pauwels, A web-based approach to BMS, BIM and IoT integration: a case study, CLIMA 2022 conference (2022). doi:10.34641/clima.2022.228.

[3] B. Pang, J. Shi, L. Jiang, Z. Pan, A semantic approach to dynamic path planning for fire evacuation through BIM and IoT data integration, Advances in Civil Engineering 2024 (2024). doi:10.1155/2024/8839865.

[4] L. Chamari, E. Petrova, P. Pauwels, An end-to-end implementation of a service-oriented architecture for data-driven smart buildings, IEEE Access 11 (2023) 117261–117281. doi:10.1109/ACCESS.2023.3325767.

[5] J. Bjørnskov, M. Jradi, An ontology-based innovative energy modeling framework for scalable and adaptable building digital twins, Energy and Buildings 292 (2023) 113146. doi:10.1016/j.enbuild.2023.113146.

[6] M. Shahinmoghadam, A. Motamedi, An Ontology-Based Mediation Framework for Integrating Federated Sources of BIM and IoT Data, volume 98, Springer International Publishing, 2021. doi:10.1007/978-3-030-51295-8_63.
[7] K. Janowicz, M. H. Rasmussen, M. Lefrançois, G. F. Schneider, P. Pauwels, BOT: The building topology ontology of the w3c linked building data group, Semantic Web 12 (2020) 143–161. doi:10.3233/SW-200385.
[8] A. J. Donkers, D. Yang, B. de Vries, N. H. Baken, Real-time building performance monitoring using semantic digital twins, in: 9th Linked Data in Architecture and Construction Workshop, LDAC 2021, CEUR-WS. org, 2021, pp. 55–66.
[9] F. Lygerakis, N. Kampelis, D. Kolokotsa, Knowledge graphs' ontologies and applications for energy efficiency in buildings: A review, Energies 15 (2022). doi:10.3390/en15207520.
[10] A. Donkers, D. Yang, B. de Vries, N. Baken, Semantic web technologies for indoor environmental quality: A review and ontology design, Buildings 12 (2022). doi:10.3390/buildings12101522.
[11] D. D. Eneyew, M. A. Capretz, G. T. Bitsuamlak, Toward smart-building digital twins: BIM and IoT data integration, IEEE Access 10 (2022) 130487–130506. doi:10.1109/ACCESS.2022.3229370.
[12] K. El Mokhtari, I. Panushev, J. J. McArthur, Development of a cognitive digital twin for building management and operations, Frontiers in Built Environment 8 (2022) 1–18. doi:10.3389/fbuil.2022.856873.
[13] C. Ramonell, R. Chacón, H. Posada, Knowledge graph-based data integration system for digital twins of built assets, Automation in Construction 156 (2023). doi:10.1016/j.autcon.2023.105109.
[14] J. Shi, Z. Pan, L. Jiang, X. Zhai, An ontology-based methodology to establish city information model of digital twin city by merging BIM, GIS and IoT, Advanced Engineering Informatics 57 (2023) 102114. doi:10.1016/j.aei.2023.102114.
[15] I. Fatokun, A. R. N. Sheela, T. Mecharnia, M. Lefrançois, V. Charpenay, F. Badeig, A. Zimmermann, Modular knowledge integration for smart building digital twins, CEUR Workshop Proceedings 3633 (2023) 123–138.
[16] A. d'Avila Garcez, L. C. Lamb, Neurosymbolic AI: the 3rd wave, Artificial Intelligence Review 56 (2023) 12387–12406.
[17] R. Buchmann, J. Eder, H.-G. Fill, U. Frank, D. Karagiannis, E. Laurenzi, J. Mylopoulos, D. Plexousakis, M. Y. Santos, Large language models: Expectations for semantics-driven systems engineering, Data & Knowledge Engineering 152 (2024) 102324. doi:10.1016/j.datak.2024.102324.
[18] V. Vaishnavi, W. Kuechler, S. Petter, Design science research in information systems, Journal of Management Information Systems 28 (2004) 75–105.
[19] G. F. Schneider, M. H. Rasmussen, Technicalbuildingsystems/opensmarthomedata: First release of open smart home data set, 2018. doi:10.5281/zenodo.1244602.
[20] G. Schneider, M. Rasmussen, P. Bonsma, J. Oraskari, P. Pauwels, Linked building data for modular building information modelling of a smart home, in: J. Karlshøj, R. Scherer (Eds.), eWork and eBusiness in Architecture, Engineering and Construction, CRC Press, 2018, pp. 407–414. doi:10.1201/9780429506215-51.
[21] N. Noy, D. Mcguinness, Ontology development 101: A guide to creating your first ontology, Knowledge Systems Laboratory 32 (2001).
[22] A. Gupta, S. Tyagi, N. Panwar, S. Sachdeva, U. Saxena, Nosql databases: Critical analysis and comparison, in: 2017 International conference on computing and communication technologies for smart nation (IC3TSN), IEEE, 2017, pp. 293–299.
[23] S. Tang, D. R. Shelden, C. M. Eastman, P. Pishdad-Bozorgi, X. Gao, A review of building information modeling (BIM) and the internet of things (IoT) devices integration: Present status and future trends, 2019. doi:10.1016/j.autcon.2019.01.020.
[24] B. Shah, P. M. Jat, K. Sashidhar, Performance study of time series databases, arXiv preprint arXiv:2208.13982 (2022).
[25] P. Lewis, E. Perez, A. Piktus, F. Petroni, V. Karpukhin, N. Goyal, H. Küttler, M. Lewis, W.-t. Yih, T. Rocktäschel, et al., Retrieval-augmented generation for knowledge-intensive nlp tasks, Advances

in neural information processing systems 33 (2020) 9459–9474.
[26] S. Munir, A. Aldini, Towards evaluating large language models for graph query generation, arXiv preprint arXiv:2411.08449 (2024).
[27] N. Baken, Linked data for smart homes: Comparing RDF and labeled property graphs, in: LDAC2020—8th Linked Data in Architecture and Construction Workshop, 2020, pp. 23–36.