Large Language Models for the Creation and Use of Semantic Ontologies in Buildings: Requirements and Challenges

Ozan Baris Mulayim*
omulayim@andrew.cmu.edu
Carnegie Mellon University / Lawrence Berkeley National Laboratory
Pittsburgh, PA, United States

Lazlo Paul
LPaul@lbl.gov
Lawrence Berkeley National Laboratory
Berkeley, CA, United States

Marco Pritoni
mpritoni@lbl.gov
Lawrence Berkeley National Laboratory
Berkeley, CA, United States

Anand Krishnan Prakash
akprakash@lbl.gov
Lawrence Berkeley National Laboratory / Carnegie Mellon University
Berkeley, CA, United States

Malavikha Sudarshan
msudarshan@lbl.gov
Lawrence Berkeley National Laboratory / University of California, Berkeley
Berkeley, CA, United States

Gabe Fierro
gtfierro@mines.edu
Colorado School of Mines
Golden, CO, United States

ABSTRACT

Semantic ontologies offer a formalized, machine-readable framework for representing knowledge, enabling the structured description of complex systems. In the building domain, the adoption of ontologies like the `Brick` schema has transformed how buildings and their systems are modeled by providing a standardized, interoperable language. However, the complexity and the steep learning curve involved in developing and querying semantic models present substantial challenges, often requiring a workforce with specialized expertise. This paper builds on our experience in investigating how Large Language Models (LLMs) can help address these challenges, focusing on their role in constructing and querying of semantic models, particularly using the `Brick` Schema. Our study outlines the requirements and metrics for evaluating the scalability and effectiveness of LLM-based tools, while also discussing the current challenges and limitations in developing such tools. Ultimately, this paper aims to orient research efforts as various groups experiment with diverse techniques, while enabling more effective comparison of emerging solutions and fostering collaboration across the field.

CCS CONCEPTS

• **Computing methodologies → Machine learning**; • **Information systems → Ontologies**.

KEYWORDS

Semantic Ontology, Large Language Models, Knowledge Graphs

ACM Reference Format:
Ozan Baris Mulayim, Lazlo Paul, Marco Pritoni, Anand Krishnan Prakash, Malavikha Sudarshan, and Gabe Fierro. 2024. Large Language Models for the Creation and Use of Semantic Ontologies in Buildings: Requirements and Challenges. In *The 11th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation (BUILDSYS '24), November 7–8, 2024, Hangzhou, China.* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3671127.3698792

BUILDSYS '24, November 7–8, 2024, Hangzhou, China

ACM ISBN 979-8-4007-0706-3/24/11.
https://doi.org/10.1145/3671127.3698792

1 INTRODUCTION

Semantic ontologies have become a cornerstone in the building domain, fundamentally transforming the way physical systems are described. By offering a structured, machine-readable representation, these ontologies facilitate interoperability, data integration, and automation across various building management tasks [21]. However, despite their immense potential, the complexity involved in developing and querying models based on these ontologies poses significant challenges. While efforts have been made to capture the intricacies of building systems, tools that enable building managers and application developers to effectively create and query these models have not been adequately developed [5]. The lack of necessary easy-to-use tools restricts adoption to users with advanced programming and information systems knowledge, who must also possess a deep understanding of building systems, their components, and modeling choices.

We explore how Large Language Models (LLMs) can address the challenges that previous methods have faced (e.g., need for building-specific training) in building model construction and query generation. Trained on a vast corpus of data, LLMs have the potential to generalize across domains without requiring the extensive domain-specific training that many earlier approaches needed. Furthermore, their human-interactive interface makes them a more adaptable and accessible tool, as they are not restricted to predefined rules or models, allowing broader applicability in diverse building contexts. By leveraging LLMs, these processes can be democratized, reducing reliance on specialized knowledge and making ontology-based models more widely usable.

In this paper, we make the following contributions: (1) we specify a list of features and performance indicators that can be used to evaluate the scalability and effectiveness of future LLM-based tools in real-world building-related applications, (2) we discuss the current implementation challenges and future research directions for these application areas, providing directions for the development of LLM-enhanced semantic ontology workflows for building applications. Though we use Brick Schema as a case study, this work is equally applicable to other building ontologies.

2 BACKGROUND

The workflow and steps involved in both the construction of a model for a target building and the generation of a query to develop applications are closely interconnected, as illustrated in Figure 1. In the following sections, we discuss the specific challenges and previous efforts that have shaped these tasks.

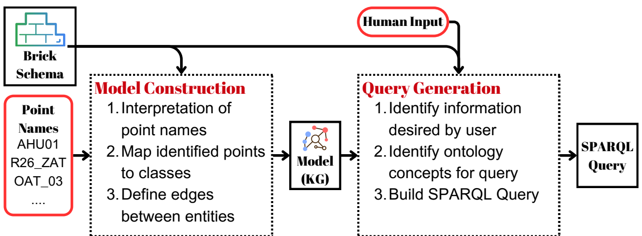

Figure 1: Steps involved in Model Construction and Query Generation Workflow

2.1 Model Construction

Semantic models streamline application development and deployment by providing a standardized framework to understand spatial and functional relationships between equipment, enabling generalization of applications such as fault detection and diagnostics (FDD) and control optimization. However, manually constructing these models is a complex task, requiring expertise in building systems and ontologies such as `Brick` [4] and SAREF [8]. This process often involves manually mapping data points (e.g., sensors, setpoints) to standardized semantic classes, defining relationships, and ensuring that the model follows the ontology constraints, often interpreting cryptic Building Automation System (BAS) point names. The task is labor-intensive and costly, particularly for older buildings, which may have inconsistent point naming conventions and require synthesis of metadata from various sources [21].

Automation of model construction has been studied since the early days of ontology use in buildings [6]. A significant portion of the literature focuses on identifying semantic tags or classes from building automation system (BAS) point names. This has been achieved using machine learning (ML) approaches based on labeled BAS point names [2, 15, 28, 29]. Some studies have used time series measurements to more accurately classify certain sensor data (e.g. zone temperature sensors)[13, 19, 22]. These approaches perform well for identifying the class of each point and provide information on the relationships between a point and related equipment. Fewer studies investigate data-driven approaches to determine the spatial and functional relationships between equipment [20, 28], which provide contextual information for controls and analytics. While promising, previous ML-based methods require representative data sets for training and may not perform well on unseen building point naming conventions or time series data streams. They also do not use the full wealth of information traditionally used in building semantic modeling, including as-built diagrams and BAS graphics, which may be used by multi-modal LLMs.

Recent research has explored the potential of LLMs for the creation of ontologies. Studies have shown that LLMs can be used to create ontologies from unstructured text [11], expand existing ontologies to represent new concepts [26, 32], and validate ontologies [27]. One work utilized an LLM to create an ontology and also instantiate it with test data to create a Knowledge Graph (KG), one type of semantic model [16]. The majority of published works to date have focused on the development of new ontologies, rather than the use of existing ontologies to create semantic models of real data. The ability of LLMs to format unstructured text into an ontological structure is applicable to the construction of `Brick` models. However, limitations of LLMs, including self-consistency and hallucination [33], present greater challenges when instantiating an ontology than creating one, because Universal Resource Identifiers (URIs) from an ontology must be recalled and used verbatim.

Retrieval-Augmented Generation (RAG) has emerged as a promising technique to boost LLM performance in domain-specific tasks. RAG combines the broad knowledge of pre-trained language models with the ability to retrieve and integrate relevant external information [18]. In constructing semantic models, RAG offers benefits like improved accuracy, reduced hallucination, domain adaptation, and the inclusion of up-to-date knowledge. However, its success depends on the relevance of retrieved information, and challenges remain in maintaining consistency and contextual understanding within ontologies. Other efforts to simplify model construction, such as [10], use templates to reduce the effort by reusing common patterns across buildings. SHACL-based validation ensures models are built correctly and contain necessary information to support applications. While this approach streamlines model creation, users must still understand BAS point names and building topology, then align their data with existing templates or create new ones.

2.2 Query Generation

Query generation enables users to retrieve structured information from a semantic model, which is crucial for scaling up applications such as FDD and analytics. SPARQL is a query language used to interrogate graph-based data [1]. Translating natural language questions into SPARQL queries, often referred to as Text-to-SPARQL, is critical for increasing usability of KGs, as users often lack the technical expertise to write SPARQL directly.

This translation process forms the foundation of Knowledge Graph Question Answering (KGQA), where semantic parsing methods convert natural language questions into formal query languages like SPARQL. Early research in semantic parsing-based KGQA primarily involved manually crafted SPARQL queries designed to test and evaluate ontologies, as demonstrated in several foundational studies [7, 12, 24]. These approaches, while effective in controlled settings, were labor-intensive and lacked scalability, making them impractical for broader application. As the field progressed, ML-based methods gained traction, including models that used syntactic features to train SPARQL query ranking with Tree-Long Short Term Memory (LSTM) algorithms [31], although their performance was limited when dealing with unseen questions. Other models employed sequence-to-sequence LSTM algorithms to generate SPARQL templates [25], which were then reconstructed into final queries using rule-based heuristics. However, these methods struggled with unknown words and lacked a deeper understanding of the input questions. Traditional ML relies heavily on domain-specific training data, a challenge in the building domain where each building has a unique KG. LLMs, however, have shown promise in generating SPARQL queries with minimal training, offering greater adaptability across domains. For example, SGPT was introduced as a method that bypasses the need for manual SPARQL construction by leveraging an LLM to learn graph patterns and generate queries [23]. Building on this, SPARQLGEN utilized GPT-3 in a one-shot SPARQL generation framework, where providing relevant context in the prompt significantly improved the quality of generated queries [17]. AUTO-KGQAGPT further extended this line of research by conducting experiments with GPT-3.5, demonstrating that selectively feeding fragments of the KG's T-Box and A-Box can enhance the translation of natural language questions into SPARQL queries [3]. Additionally, an LLM-based model was developed, which integrates a Bidirectional encoder representations from transformers (BERT) encoder with a Linear-Chain Conditional Random Fields (CRF) for mention extraction, followed by entity linking and relationship selection [30]. The final queries are generated by a fine-tuned ChatGLM-6B, either by combining all entities and relationships into a single prompt or by creating multiple queries for each entity. However, each of these methods had specific limitations to work with building ontologies as they either (1) result in unmanageably large subgraphs [3, 30], (2) require fine-tuning with difficult to find knowledge bases [30], (3) require ground truth queries [17], (4) require example question-query pairs [17, 23]. Based on our experiments with `Brick` models from Mortar [9], only one domain-independent approach—Auto-KGQAGPT—among the methods listed above could be adapted to work with building ontologies with minimal modifications. However, it still struggled to accurately match input descriptions to the correct ontology nodes, largely due to the complexity and ambiguity inherent in textual descriptions within the building domain.

2.3 Early Considerations on LLMs for Buildings

While the literature on this topic is currently sparse, the authors acknowledge the rapidly evolving body of research exploring the application of LLMs to the building domain. As several groups experiment with diverse techniques to develop tools addressing complex challenges, this paper reflects on the authors' experiences and lessons learned in applying LLMs to create and query semantic models of buildings using `Brick`. For model construction, we used point names from several real buildings throughout California. During query generation, we corrected a comprehensive building model from [9] and extracted application descriptions and queries from its codebase. We tested various LLM approaches, primarily using GPT-4o.

This paper is based on our preliminary experience because the necessary building blocks for systematic and rigorous evaluations are still lacking. In the absence of such benchmarks and methods, one can either present simple test cases or lay the groundwork for a collaborative research effort by listing the requirements for future works. Our focus is on the latter rather than highlighting the details of our ad-hoc experimentation. Thus, this paper calls for community efforts to develop these foundations and aims to foster a robust discussion among researchers, encouraging the sharing of early insights to accelerate innovation.

3 FEATURES & PERFORMANCE METRICS

3.1 Features

These features provide a checklist for evaluating the scalability of model and query building tools in the building domain. Though they are not strict requirements, their presence would enhance the ability of such tools to address users needs and support the use of building ontologies at scale.

Building-Agnostic. A scalable model construction tool should ideally not require specific training or fine-tuning for each individual building or point-naming convention. Existing methods often depend on training datasets tailored to each building or individual portfolios, which require experts manually tagging point names. While this approach can be useful, it limits scalability. Building-agnostic tools are the ones which can operate across various buildings without requiring customization for each case, enabling more efficient and generalized model construction.

Similarly, one of the most critical features of effective text-to-SPARQL (i.e., Query Generation) tools in the building domain is their ability to generate queries that are independent of specific building models. Given the the heterogeneity of building configurations and the existence of alternative, yet valid, modeling choices for the same configuration [5], it is essential that query generation tools can abstract these differences and produce consistent results across various building types.

Ontology Adherence. A scalable model construction tool must ensure that the models it generates strictly adhere to the underlying ontology or ontologies, such as the `Brick` schema. While obvious, several from the literature do not validate that resulting models comply with a specified ontology, and existing tools do not safeguard against the errors that are commonly made by LLMs due to hallucination, like the use of non-existent class names.

In the context of query generation, ontology adherence enables the effective use of class hierarchies and relationships. For instance, if an application needs to query all available temperature sensors, this information may not be directly accessible in the model if the sensors were modeled using a subclass (e.g., '`Supply_Air_Temperature_Sensor`). A useful feature for query generation tools would be the ability to utilize information about class hierarchies to make data gathering more flexible. For instance, by leveraging the fact that a `Supply_Air_Temperature_Sensor` is `rdf:SubClassOf brick:Temperature_Sensor`, query tools can automatically infer and retrieve data from all relevant subclasses, making data gathering more flexible and comprehensive.

Handling Ambiguous Inputs. Real-world building data is often fraught with ambiguities, such as inconsistent point names or incomplete information, which can hinder accurate model construction. A scalable tool must effectively manage this uncertainty, recognizing and resolving ambiguities while ensuring the model's integrity. By incorporating mechanisms for uncertainty-awareness, the tool becomes more robust, capable of adapting to varying data quality levels without compromising accuracy or scalability. This ability to address ambiguous inputs is crucial for ensuring the tool's practical applicability across diverse datasets and buildings.

In addition to managing data uncertainties, query generation tools for building ontologies must process human language inputs that vary in specificity, especially regarding the terminology used by practitioners. Inputs may range from precise terms to general or ambiguous phrases, and an effective tool must account for these variations. Handling such implicit descriptions ensures the tool's flexibility, making it suitable for real-world applications where user inputs often lack clarity. This adaptability enhances the tool's practicality, ensuring accurate query results even when input descriptions are not fully explicit.

Consistency. Although LLMs are inherently probabilistic, it is crucial that model construction and query generation tools produce results that are as deterministic as possible. By carefully defining tasks and providing clear instructions, the variance in LLM outputs can be minimized, ensuring that the same input consistently yields the same output. Other approaches, such as semantic validation, may also provide this feature. This determinism is essential for building management applications, where reliable and repeatable query results are necessary for making informed decisions and maintaining operational consistency.

Model-Awareness. Entities in buildings have many interconnected relationships which previous tools for automating the construction of semantic models have struggled to capture [14]. Identifying which ontology class a point belongs to is useful, but semantic models must also contain information about how entities in a building (e.g. parts of equipment, rooms, zones) relate to each other. Modeling these relationships requires awareness of what entities have been instantiated, so that they can be connected together. Useful information for identifying these relationships is often available in images such as floor plans and as-built diagrams, in addition to point names. Thus, identification of these relationships may benefit from the multi-modal input capabilities of LLMs.

An essential feature of query generation tools in the building domain is the ability to be 'aware' of the specific building model. While queries can be written independent of a specific building model, their accuracy depends heavily on the modeling choices made. This awareness allows the tool to determine whether empty results stem from the building lacking the queried concept or from a poorly formulated query. By incorporating model-awareness, these tools can improve the reliability and relevance of query results.

3.2 Performance Metrics

To provide a more comprehensive evaluation, beyond just accuracy metrics, it is also important to assess the computational complexity required for each solution. As the authors worked on generating and querying semantic models, it became evident that conducting a generalizable evaluation of these tools is not straightforward. The performance metrics listed below, aimed at assessing efficiency, were derived from preliminary tests.

Computational Complexity (i.e., runtime). Our preliminary experiments with these tools have shown that higher accuracy can often be achieved through trial-and-error or search-based approaches. However, this can significantly impact the practicality of the tools due to the extended computational complexity. In cases where complex queries or models need to be generated, the application runtime can become a critical bottleneck, limiting the effectiveness and deployability of the tool. To allow for a fair comparison, researchers should clearly report the computational runtime of their methods. It is also important to recognize that the computational cost may not stem solely from LLM calls; frameworks developed for model construction and query generation often include other components that contribute to the overall cost, such as advanced querying techniques or embedding-based similarity checks [3, 30].

Token Size. In LLMs, computational cost is closely tied to the number of tokens processed in a prompt. For tools focused on model construction or query generation, the framework does not necessarily need to be single-shot. Thus, the cumulative number of tokens used across varying number of LLM calls should be reported for each evaluation sample. This metric helps to approximate the LLM-based computational demand, indicating how resource-intensive the method is when handling multiple tasks or queries.

Monetary Cost (in US dollars). While token usage is an indirect indicator of cost, it does not fully account for the financial expenses associated with different LLMs. Some models are commercial and require payment, while others are open-source, offering lower or no direct costs. Therefore, it becomes important to report the monetary cost for each task or query in practical terms, especially for commercial LLMs. This would provide a clearer picture of the financial implications of deploying a particular method, helping researchers judge the financial feasibility of the designed tools.

4 CHALLENGES AND FUTURE RESEARCH DIRECTIONS

The challenges and future research directions outlined in this section are based on our expert insights and practical experimentation, as the essential building blocks for empirical validation are still missing. Critical components like comprehensive evaluation datasets, standardized accuracy metrics, and benchmarks are far from being available, making it impossible to validate these tools in a meaningful and comparable way. For instance, there are no widely accepted datasets for query generation, and without proper accuracy metrics, assessing tool performance is highly subjective. This lack of foundational resources means that the challenges discussed here remain theoretical, based on our informed observations. Without a concerted effort from the research community to build these key elements, achieving reliable empirical results will be out of reach, and the advancement of LLM-based tools for model construction and query generation will remain limited.

Domain Knowledge. A key challenge in developing LLM-based tools for model construction and query generation is the need for domain-specific knowledge. Building systems, with their variety of HVAC configurations, sensors, and equipment, require tools that grasp these nuances. While LLMs help lower the expertise barrier, they still struggle with the depth of knowledge required to accurately interpret building-specific data. In model construction, this is essential when establishing functional relationships between building components. For query generation, domain knowledge is needed to infer connections between variables (e.g., linking outdoor air temperature to an AHU) when human input is implicit. Additionally, embedding domain-specific knowledge into LLMs, such as fine-tuned HVAC expert models, could create specialized agents for working with building semantics

Measurement of Accuracy. Validation is crucial to ensure the accuracy and reliability of models constructed using semantic ontologies. While SHACL can be employed to automatically validate how classes and relationships interact within the ontology, it only ensures structural correctness and adherence to the ontology's rules. However, the accurate mapping of point names to their corresponding classes and relationships remains a challenge, as semantic technologies cannot validate whether a specific point name has been correctly interpreted or classified, or if relationships modeled are truly present in the building. This aspect still requires human validation in practice.

Measuring the accuracy of query results in query generation presents a unique challenge because some queries may yield results that are nearly correct but differ slightly, such as in variable names or the order in which results are returned. Traditional exact match algorithms might fail to recognize these as valid, even though the core information is correct [23]. Approaches like variable name normalization [23] could be adapted to improve recognition of equivalent results. This approach allows for a more nuanced assessment of the tool's performance, ensuring that minor discrepancies do not obscure accuracy and providing a more realistic reflection of the tool's effectiveness in diverse querying scenarios.

Training Datasets. While some datasets exist for point names and models, they are insufficient for fine-tuning LLMs to handle the wide range of building types and configurations. The absence of complete and up-to-date datasets of labeled building data remains a significant obstacle to developing accurate tools for model construction. Future efforts should focus on creating more comprehensive datasets that include not only point names but also as-built diagrams. This would allow for the use of multimodal models, integrating both visual data, such as blueprints, with point name datasets to provide more contextually accurate model construction.

Additionally, there are no available datasets in the building domain where questions are mapped to SPARQL query outputs. The creation of such datasets would facilitate both training and enable techniques like few-shot prompting for query generation. It is important, however, to consider the inherent building dependency in such datasets. Training models on text-to-SPARQL pairs specific to a building is likely to result in much higher accuracy, as the context and structure of the queries would directly align with the building's unique configuration.

Evaluation Datasets. LLM-based applications are observed to be improving with iterative prompt engineering. However, their biggest challenge lies in understanding how to evaluate them. Thus, development of an evaluation set is inevitable for the progress of this field. Current evaluation sets for point name tagging have primarily been created by reverse-engineering available `Brick` models [2], which, while useful, are not sufficient for testing more advanced capabilities like handling ambiguous inputs. Future research should focus on developing evaluation sets that tag undefinable point names—point names that have no clear meaning or interpretation. These would be particularly valuable for testing how tools handle uncertainty and deal with ambiguous or erroneous input data. Such evaluation sets would also provide a way to measure how well LLMs can avoid hallucinating nonexistent classes for meaningless point names, ensuring that the tools remain grounded in the available ontology. It is important to note that these evaluation sets should include examples to test the whole model building capability rather than focusing on its first step (i.e., point name tagging).

While most of existing query generation techniques have been tested on specific datasets, building industry lacks such a source. Development of such an evaluation set is challenging in various ways: (1) inputs are likely to be ambiguous, (2) results and queries will depend on the specific building model, (3) effective accuracy measurements are lacking. Firstly, there is a considerable lack of human inputs that would indicate how building managers would linguistically describe the metadata requirements. Development of such a source should be systematic so it can consist sufficient diversity in two directions: variance in the application type, variance in the human language descriptions of a certain application. Secondly, another challenge lies in the way of how queries can be written. For the same metadata requirements, queries can be generated in different ways. Thus, a set of rules should be established in generating the ground truth queries or results.

Inference of Relationships. Another critical challenge for model generation is the inference of relationships between different entities in a model (e.g. points, equipment, rooms, zones). Previous work generally focuses on mapping individual points to ontological classes, and fails to create the relationships that describe how a building is composed, which are a necessary part of a semantic model. Developing methods that can infer these relationships from point names, domain knowledge, and building context would greatly enhance the utility of the generated models. Future research should explore the integration of contextual data and relationship inference mechanisms into LLM-driven tools to create more complete and accurate models.

Accurate Up-to-Date Building Models. As the way queries are written depends on the modeling choices [5], the evaluation set needs to include a variety of building models. Though there are readily available models of `Brick` schema in [4, 9], none of them are accurately modeled based on the latest modeling consensus. Thus, developing such an evaluation set inevitably requires gathering/creating representative building models of various sizes.

Subgraph Extraction. As opposed to other KGs tested by previous studies, building models created using `Brick` schema suffer from their large scale, limiting the ability to push the whole schema in the prompt. Thus, especially to provide model-awareness, tools should be able to extract the relevant subgraphs of the models within a reasonable token limit.

5 CONCLUSIONS

LLMs offer promising opportunities to perform a wide range of tasks through human-interactive interfaces, providing a novel approach to engaging with complex systems like semantic ontologies. In the building industry, where semantic ontologies have been used for over a decade, LLMs can enhance adoption by lowering the expertise barriers that have traditionally limited their widespread use. However, while LLMs can generate results that are nearly correct, they still require human refinement and verification, and achieving fully automated solutions without specialized knowledge in coding, information systems, or building operations remains a challenging but attainable goal. The development of LLM-driven tools for building ontologies will require sustained research, collaboration, and the establishment of standardized models and evaluation benchmarks to ensure consistent progress. Toward this vision, this paper provides a outline for comparing and evaluating methods in the field, identifies key challenges, and outlines future research directions to guide the effective application of LLMs in model and query development tasks.

ACKNOWLEDGMENTS

This research was supported by the Assistant Secretary for Energy Efficiency and Renewable Energy, Office of Building Technologies of the U.S. Department of Energy under contract DE-AC02-05CH11231.

REFERENCES

[1] Waqas Ali, Mohammad Saleem, Bin Yao, A. Hogan, and A. N. Ngomo. 2021. A survey of RDF stores SPARQL engines for querying knowledge graphs. *The VLDB Journal* 31 (2021), 1 – 26. https://doi.org/10.1007/s00778-021-00711-3

[2] Mahathir Almashor, Mashud Rana, John McCulloch, Ashfaqur Rahman, and Subbu Sethuvenkatraman. 2023. What's The Point: AutoEncoding Building Point Names. In *Proceedings of the 10th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation*. ACM, Istanbul Turkey, 256–260. https://doi.org/10.1145/3600100.3623748

[3] Caio Viktor S. Avila, Vânia M.P. Vidal, Wellington Franco, and Marco A. Casanova. 2024. Experiments with text-to-SPARQL based on ChatGPT. In *2024 IEEE 18th International Conference on Semantic Computing (ICSC)*. IEEE, Laguna Hills, CA, USA, 277–284. https://doi.org/10.1109/ICSC59802.2024.00050

[4] Bharathan Balaji, Arka Bhattacharya, Gabriel Fierro, Jingkun Gao, Joshua Gluck, Dezhi Hong, Aslak Johansen, Jason Koh, Joern Ploennigs, Yuvraj Agarwal, Mario Berges, David Culler, Rajesh Gupta, Mikkel Baun Kjærgaard, Mani Srivastava, and Kamin Whitehouse. 2016. Brick: Towards a Unified Metadata Schema For Buildings. In *Proceedings of the 3rd ACM International Conference on Systems for Energy-Efficient Built Environments*. ACM, Palo Alto CA USA, 41–50. https://doi.org/10.1145/2993422.2993577

[5] Imane Lahmam Bennani, Anand Krishnan Prakash, Marina Zafiris, Lazlo Paul, Carlos Duarte Roa, Paul Raftery, Marco Pritoni, and Gabe Fierro. 2021. Query relaxation for portable brick-based applications. In *Proceedings of the 8th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation*. ACM, Coimbra Portugal, 150–159. https://doi.org/10.1145/3486611.3486671

[6] Arka A Bhattacharya, Dezhi Hong, David Culler, Jorge Ortiz, Kamin Whitehouse, and Eugene Wu. 2015. Automated metadata construction to support portable building applications. In *Proceedings of the 2nd ACM International Conference on Embedded Systems for Energy-Efficient Built Environments*. 3–12.

[7] Christian Bizer and Andreas Schultz. 2008. Benchmarking the performance of storage systems that expose SPARQL endpoints. In *Proc. 4 th International Workshop on Scalable Semantic Web Knowledge Base Systems (SSWS)*. Citeseer, 39.

[8] L. Daniele, F. den Hartog, and J. Roes. 2015. *Study on Semantic Assets for Smart Appliances Interoperability: D-S4: Final Report*. Technical report. European Union.

[9] Gabe Fierro, Marco Pritoni, Moustafa Abdelbaky, Daniel Lengyel, John Leyden, Anand Prakash, Pranav Gupta, Paul Raftery, Therese Peffer, Greg Thomson, and David E. Culler. 2019. Mortar: An Open Testbed for Portable Building Analytics. *ACM Transactions on Sensor Networks* 16, 1 (Dec. 2019), 7:1–7:31. https://doi.org/10.1145/3366375

[10] Gabe Fierro, Avijit Saha, Tobias Shapinsky, Matthew Steen, and Hannah Eslinger. 2022. Application-driven creation of building metadata models with semantic sufficiency. In *Proceedings of the 9th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation*. ACM, Boston Massachusetts, 228–237. https://doi.org/10.1145/3563357.3564083

[11] Hamed Babaei Giglou, Jennifer D'Souza, and Sören Auer. 2023. LLMs4OL: Large Language Models for Ontology Learning. http://arxiv.org/abs/2307.16648 arXiv:2307.16648 [cs, math].

[12] Peter Haase, Tobias Mathäß, and Michael Ziller. 2010. An evaluation of approaches to federated query processing over linked data. In *Proceedings of the 6th international conference on semantic systems*. 1–9.

[13] Dezhi Hong, Hongning Wang, Jorge Ortiz, and Kamin Whitehouse. 2015. The Building Adapter: Towards Quickly Applying Building Analytics at Scale. In *Proceedings of the 2nd ACM International Conference on Embedded Systems for Energy-Efficient Built Environments*. ACM, Seoul South Korea, 123–132. https://doi.org/10.1145/2821650.2821657

[14] Jason Koh, Dezhi Hong, Rajesh Gupta, Kamin Whitehouse, Hongning Wang, and Yuvraj Agarwal. 2018. Plaster: an integration, benchmark, and development framework for metadata normalization methods. In *Proceedings of the 5th Conference on Systems for Built Environments* (Shenzen, China) *(BuildSys '18)*. Association for Computing Machinery, New York, NY, USA, 1–10. https://doi.org/10.1145/3276774.3276794

[15] Jason Koh, Dhiman Sengupta, Julian McAuley, Rajesh Gupta, Bharathan Balaji, and Yuvraj Agarwal. 2017. Scrabble: converting unstructured metadata into brick for many buildings. In *Proceedings of the 4th ACM International Conference on Systems for Energy-Efficient Built Environments* (Delft, Netherlands) *(BuildSys '17)*. Association for Computing Machinery, New York, NY, USA, Article 48, 2 pages. https://doi.org/10.1145/3137133.3141448

[16] Vamsi Krishna Kommineni, Birgitta König-Ries, and Sheeba Samuel. 2024. From human experts to machines: An LLM supported approach to ontology and knowledge graph construction. http://arxiv.org/abs/2403.08345 arXiv:2403.08345 [cs].

[17] Liubov Kovriguina, Roman Teucher, Daniil Radyush, and Dmitry Mouromtsev. [n. d.]. SPARQLGEN: One-Shot Prompt-based Approach for SPARQL Query Generation. ([n. d.]).

[18] Patrick Lewis, Ethan Perez, Aleksandra Piktus, Fabio Petroni, Vladimir Karpukhin, Naman Goyal, Heinrich Küttler, Mike Lewis, Wen-tau Yih, Tim Rocktäschel, Sebastian Riedel, and Douwe Kiela. 2020. Retrieval-augmented generation for knowledge-intensive NLP tasks. In *Proceedings of the 34th International Conference on Neural Information Processing Systems* (Vancouver, BC, Canada) *(NIPS '20)*. Curran Associates Inc., Red Hook, NY, USA, Article 793, 16 pages.

[19] Sakshi Mishra, Andrew Glaws, Dylan Cutler, Stephen Frank, Muhammad Azam, Farzam Mohammadi, and Jean-Simon Venne. 2020. Unified architecture for data-driven metadata tagging of building automation systems. *Automation in Construction* 120 (2020), 103411. https://doi.org/10.1016/j.autcon.2020.103411

[20] Marco Pritoni, Arka A. Bhattacharya, David Culler, and Mark Modera. 2015. Short Paper: A Method for Discovering Functional Relationships Between Air Handling Units and Variable-Air-Volume Boxes From Sensor Data. In *Proceedings of the 2nd ACM International Conference on Embedded Systems for Energy-Efficient Built Environments* (Seoul, South Korea) *(BuildSys '15)*. Association for Computing Machinery, New York, NY, USA, 133–136. https://doi.org/10.1145/2821650.2821677

[21] Marco Pritoni, Drew Paine, Gabriel Fierro, Cory Mosiman, Michael Poplawski, Avijit Saha, Joel Bender, and Jessica Granderson. 2021. Metadata schemas and ontologies for building energy applications: A critical review and use case analysis. *Energies* 14, 7 (2021), 2024.

[22] Mashud Rana, Ashfaqur Rahman, Mahathir Almashor, John McCulloch, and Subbu Sethuvenkatraman. 2024. Automatic Classification of Sensors in Buildings: Learning from Time Series Data. In *AI 2023: Advances in Artificial Intelligence*, Tongliang Liu, Geoff Webb, Lin Yue, and Dadong Wang (Eds.). Springer Nature, Singapore, 367–378. https://doi.org/10.1007/978-981-99-8388-9_30

[23] Md Rashad Al Hasan Rony, Uttam Kumar, Roman Teucher, Liubov Kovriguina, and Jens Lehmann. 2022. SGPT: A Generative Approach for SPARQL Query Generation From Natural Language Questions. *IEEE Access* 10 (2022), 70712–70723. https://doi.org/10.1109/ACCESS.2022.3188714

[24] Michael Schmidt, Thomas Hornung, Georg Lausen, and Christoph Pinkel. 2009. SPˆ 2Bench: a SPARQL performance benchmark. In *2009 IEEE 25th International Conference on Data Engineering*. IEEE, 222–233.

[25] Tommaso Soru, Edgard Marx, André Valdestilhas, Diego Esteves, Diego Moussallem, and Gustavo Publio. 2018. Neural machine translation for query construction and composition. *arXiv preprint arXiv:1806.10478* (2018).

[26] Sabrina Toro, Anna V Anagnostopoulos, Sue Bello, Kai Blumberg, Leigh Carmody, Alexander D Diehl, Damion Dooley, William Duncan, Petra Fey, Pascale Gaudet, Nomi L Harris, Marcin Joachimiak, Leila Kiani, Monica C Munoz-Torres, Shawn O'Neil, David Osumi-Sutherland, Justin P Reese, Leonore Reiser, Sofia Robb, Troy Ruemping, Eric Sid, Ray Stefancsik, Magalie Weber, Valerie Wood, and Christopher J Mungall. [n. d.]. Dynamic Retrieval Augmented Generation of Ontologies using Artificial Intelligence (DRAGON-AI). ([n. d.]).

[27] Stefani Tsaneva, Stefan Vasic, and Marta Sabou. [n. d.]. LLM-driven Ontology Evaluation: Verifying Ontology Restrictions with ChatGPT. ([n. d.]).

[28] Shanshan Wan, Mengnan Zhao, Yimin Chen, Shuyue Yang, Dongwei Qiu, and L. James Lo. 2023. A novel data-driven relationship inference approach for automatic data tagging in building heating, ventilation and air conditioning systems. *Building and Environment* 246 (2023), 110968. https://doi.org/10.1016/j.buildenv.2023.110968

[29] David Waterworth, Subbu Sethuvenkatraman, and Quan Z. Sheng. 2021. Advancing smart building readiness: Automated metadata extraction using neural language processing methods. *Advances in Applied Energy* 3 (Aug. 2021), 100041. https://doi.org/10.1016/j.adapen.2021.100041

[30] Shuangtao Yang, Mao Teng, Xiaozheng Dong, and Fu Bo. 2023. LLM-Based SPARQL Generation with Selected Schema from Large Scale Knowledge Base. In *Knowledge Graph and Semantic Computing: Knowledge Graph Empowers Artificial General Intelligence*, Haofen Wang, Xianpei Han, Ming Liu, Gong Cheng, Yongbin Liu, and Ningyu Zhang (Eds.). Springer Nature, Singapore, 304–316. https://doi.org/10.1007/978-981-99-7224-1_24

[31] Hamid Zafar, Giulio Napolitano, and Jens Lehmann. 2018. Formal query generation for question answering over knowledge bases. In *European semantic web conference*. Springer, 714–728.

[32] Antonio Zaitoun, Tomer Sagi, Szymon Wilk, and Mor Peleg. [n. d.]. Can Large Language Models Augment a Biomedical Ontology with missing Concepts and Relations? ([n. d.]).

[33] Liang Zhang and Zhelun Chen. [n. d.]. Opportunities and Challenges of Applying Large Language Models in Building Energy Efficiency and Decarbonization Studies: An Exploratory Overview. ([n. d.]).