

SARGON2 Ontology: Updating SARGON Ontology for Semantic Data Sharing in the Energy Domain

Su Mon Tun

*Institute for Automation of Complex Power Systems
RWTH Aachen University
Aachen, Germany
su.tun@eonerc.rwth-aachen.de*

Zhiyu Pan

*Institute for Automation of Complex Power Systems
RWTH Aachen University
Aachen, Germany
zhiyu.pan@eonerc.rwth-aachen.de*

Fengshuo Hao

*Institute for Automation of Complex Power Systems
RWTH Aachen University
Aachen, Germany
fengshuo.hao@eonerc.rwth-aachen.de*

Ferdinanda Ponci

*Institute for Automation of Complex Power Systems
RWTH Aachen University
Aachen, Germany
fponci@eonerc.rwth-aachen.de*

Antonello Monti

*Institute for Automation of Complex Power Systems
RWTH Aachen University and Fraunhofer FIT
Aachen, Germany
amonti@eonerc.rwth-aachen.de*

Abstract—The power systems are becoming a decentralized cyber-physical system with new energy demands, grid infrastructure, and renewable energy sources, introducing challenges in managing heterogeneous data across various domains within the energy domain. As a result, the harmonization of heterogeneous data becomes one of the possible solutions for the development and analysis of novel solutions in the energy domain. Semantic Web technologies serve as effective solutions to tackle the challenges of heterogeneous data integration across various sectors, including smart energy systems. By providing machine-readable information and facilitating a shared understanding of data for both humans and machines, these technologies, especially ontology, enable the linking of domain-specific information. To address this issue, we propose the development of SARGON2 ontology by using a semi-automatic ontology development framework that extends SARGON to enhance interoperability within the energy domain. SARGON2 extends SARGON ontology to cover more renewable energy resources, weather information, and power system components. By integrating the ontology with the existing standards, such as IEC61970 CIM, SARGON2 incorporates entities related to RES, building domains, and non-energy domains like weather information.

Index Terms—SARGON, ontology, semantic interoperability, CIM.

I. INTRODUCTION

The transition of power systems has introduced unprecedented changes, including in energy demand, grid infrastructure, low-carbon landscape, and more renewable energy resources (RES). Due to the fact that these power systems highly rely on the digital architecture and also include multiple domains, the interoperability of heterogeneous data across various domains within the energy sector has become a major challenge. To address this issue, ontologies are considered a crucial element of the semantic web, offering knowledge about entities and their relationships in formats that are readable by both humans and machines.

On the other hand, ontology plays a significant role in data spaces by providing a structured data model for representing knowledge and facilitating interoperability among diverse data

sources. Data spaces are essential to promote the digitalization of power systems, leading to a new era characterized by enhancing efficiency, reliability, and sustainability [1]. In a data space, ontologies can provide data discovery, sharing, and integration by allowing systems to understand the meaning of the data they interact with. In addition, various stakeholders such as utilities, data providers, and data consumers with different levels of knowledge are interconnected with the data space. As a consequence, this kind of data management system can introduce scalability challenges and complexity in maintenance for handling unstructured or rapidly changing data. Ontologies can help address these challenges by providing standardized vocabularies and relationships that facilitate the interpretation of data across stakeholders, hence simplifying integration processes and enhancing interoperability.

SARGON [2] ontology covers the smart grid and building information by linking with CIM and IEC standards. However, the entities of SARGON related to the power systems and renewable energy resources are not sufficient, and the extension of the ontology is necessary by considering these points. This work develops SARGON2, which is an extension of SARGON using a semi-automatic ontology development framework in [3] by integrating the existing standards in the energy domain like IEC61970 Common Information Model (CIM) [4]. SARGON2 aims to include entities such as renewable energy resources (biomass, biogas, geothermal, solar, and wind), the energy domain, the building domain, and weather information. In this paper, we focus on solving the problem of heterogeneous data by developing a SARGON2 ontology. More specifically, the paper provides the following contributions:

- the extension of SARGON ontology by adding new entities in RES, weather information, and more power system components based on the specified use case;
- the step-by-step task-specific updating workflow of the ontology by following the semi-automatic ontology de-

velopment framework.

II. LITERATURE REVIEW

This section discusses existing ontologies in the smart energy domain, state-of-the-art ontology development tools, the CIM standard, and ontology.

A. Existing Ontology in Smart Energy Domain

SARGON [2] ontology is developed to cover several ontologies in the smart grid and building automation. Additionally, the requirement specifications of the ontology were surveyed in conjunction with the CIM and IEC standards. SARGON includes classes covering medium voltage distribution grid automation, phasor measurement unit (PMU) interaction for better grid monitoring, and building automation, focusing on energy demand management through smart meters and actuators, with active user participation.

SAREF4BLDG [5] is designed to bridge the interoperability gap between different stakeholders (such as architects, engineers, consultants, contractors, and product manufacturers) and the applications used to manage building information throughout the phases of the building life cycle (Planning and Design, Construction, Commissioning, Operation, Retrofitting/ Refurbishment/ Reconfiguration and Demolition/ Recycling). By adopting SAREF4BLDG, smart devices from manufacturers supporting the Industry Foundation Classes (IFC) data model can communicate seamlessly with each other.

SAREF4ENER [6] is designed to focus on demand response scenarios, where consumers offer energy flexibility to the smart grid. Use cases include optimizing appliance operation through power profiles to enhance energy efficiency, monitoring appliance status, and responding to grid requests like consuming more or less energy based on availability, or reducing power usage during emergencies.

The Smart Building Evacuation Ontology (SBEO) [7] is a reusable framework for indoor spaces that integrates three key data models: user, building, and context. Its structure includes a user model to represent the characteristics and relationships of the occupants, a building model to describe the layout and infrastructure, and a context model to capture dynamic changes in both the building and its occupants. Through SBEO, building evacuation processes can be enhanced by providing a comprehensive understanding of occupants, infrastructure, and real-time conditions.

B. Ontology Development Tools

In the matter of ontology development methodologies, several tools have been developed, starting from the 1990s [8]. Skeletal Methodology [9], also known as the Enterprise Method, is used primarily to build enterprise ontologies. This methodology emphasizes the importance of defining the objectives and understanding user requirements before beginning the development of the ontology.

Ontology development 101 [8] is designated for developing a domain ontology. The approach relies on an iterative design process, enabling developers to create an ontology with the

use of Protégé [10]. This also considers reusing existing ontologies.

The On-To-Knowledge (OTK) methodology [11] is designed for the application-driven ontology development, particularly during the implementation of ontology-based knowledge management (KM) systems. This methodology guides the entire process, from the initial stages of a knowledge management project to the final deployment of the ontology-based application. On-To-Knowledge consists of six activities, i.e., feasibility study, kickoff, refinement, evolution, and maintenance. Therefore, it involves the early stage of starting a KM project to the final version of the ontology-based KM application.

The methodology outlined in [12] is a scenario-based approach that identifies nine specific scenarios to develop ontologies and ontology networks. Additionally, this framework supports the creation of ontologies and ontology networks through the reuse of existing ontologies, non-ontological resources, and the application of ontology design patterns.

Linked Open Terms (LOT) [13] is a lightweight framework designed for developing ontologies and vocabularies, particularly focusing on industrial applications and the Semantic Web. It integrates various activities such as requirements specification, ontology implementation, publication, and maintenance while emphasizing collaboration between domain experts and software engineers. LOT encourages the reuse of existing ontologies and promotes best practices for publishing ontologies in compliance with Linked Data principles.

The comparison of the above-mentioned methodologies for developing ontology has been carried out depending on the ontology reuse, data model transformation, ontology integration, and maintenance factors. It is discovered that the mentioned ontology development methods, except the Skeletal Methodology, define the ontology reuse. Nonetheless, how to carry out the knowledge search for the existing ontology has not been mentioned.

C. CIM standard and CIM ontology

Common Information Model (CIM) is an open standard developed by the Electric Power Research Institute (EPRI) in North America to represent power system components [14]. Additionally, CIM offers a standardized vocabulary used to establish the interfaces for integrating services and applications, as well as defining internal data structures within application database schemas and message formats for information exchange in RDF or XML schema [15]. Typically, CIM is represented with Unified Modeling Language (UML), which outlines the classes, properties, and relationships that can be utilized to describe an object.

In contrast, ontologies provide a much richer vocabulary for expressing these relationships. CIM ontology defines the relationships between various entities and concepts within the CIM, allowing for better representation of knowledge and facilitating applications like information management and systems integration [15].

III. ONTOLOGY DEVELOPMENT

In the ontology development process, the development of ontology (SARGON2) is performed by following the semi-automatic ontology development framework in [3]. The procedure and components utilized in this process are illustrated in Figure 1. The procedure is discussed in detail as follows:

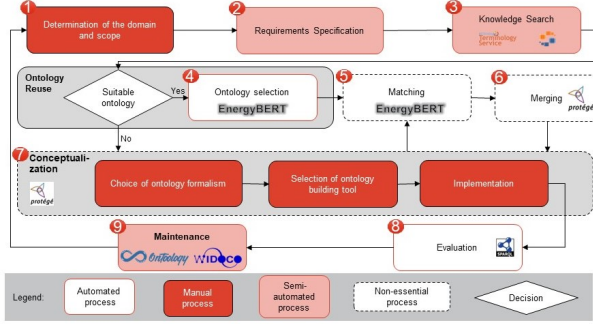


Fig. 1. Task-specific ontology development workflow

A. Determination of the domain and scope

Before starting the development of the ontology, the process of defining the purpose of the ontology, end users, use cases, and requirements specification was performed. By following the guidelines mentioned in [16], the Ontology Requirement Specification Document (ORSD) has been formulated based on the desired use cases. The scope of the ontology is to focus on the energy, building, and weather domains.

B. Requirement Specification

In this paper, the ontology requirements are divided into non-functional requirements and functional requirements. While non-functional requirements are the general requirements that should be met by the ontology, the functional requirements define more specific content-based requirements in the form of competency questions (CQ) and the answers to them. The last section of ORSD has mentioned the frequencies of the terminology (names, adjectives, and verbs) that appeared in the CQs and their answers that are formally represented in the ontology by means of entities, classes, attributes, relations, and instances. The most frequent terms are utilized as keywords for the knowledge search section.

C. Knowledge Search

This section conducted the survey on the existing ontologies and data models that are relevant to the domain and scope of the ontology by using the terms from the ORSD. The keywords participated in the discovery section include *Energy, building, weather, renewable energy resources*. This knowledge search is performed mainly on the Terminology Service NFDI4ING [17], LOV, and FIWARE [18]. The related ontology results can be seen in Table I.

TABLE I
KNOWLEDGE SEARCH RESULTS

Ontology	Domain
SARGON	Energy, Building, Device
SAREF4BLDG	Energy, Building, Device
SAREF4ENER	Energy, Device
SAREF4GRID	Energy, Smart Grid
EM-KPI ontology	Energy, Building, Device, Weather
Semantic Sensor Network	Device
Flow Systems Ontology	Energy, Device
CIM Ontology	Energy
Smart Building Evacuation Ontology	Building

D. Ontology Selection

In order to select the suitable ontologies to be reused based on the requirements of ORSD, the ontology selection is performed. Since SARGON2 is an extension of SARGON with the integration of CIM ontology, both SARGON and CIM ontology are selected by default. However, there are still entities required to meet the requirements that can be adopted from the other ontologies. Therefore, the selection of suitable ontologies from the knowledge search is performed using the large language model Energy BERT [19]. Energy BERT is an extension of the BERT pre-trained model, which learns information from the left and right contexts of the dedicated word and captures the different meanings for the same word in different contexts. The Energy BERT is trained by using dynamic masking and a larger batch size to make full use of the semantic information contained in the energy domain corpus. Masked Language Model (MLM) is used in training the Energy BERT without using Next Sentence Prediction (NSP), since it is more applied to quality assurance (QA) tasks, which are not so helpful with semantic matching.

For the ontology selection algorithm, the lexical chain in the relevant domain is defined based on the knowledge search and is used as the first entry in the workflow. Then, the Energy BERT semantic similarity score is calculated for each ontology in the knowledge search with the word chain.

E. Ontology Matching

The selected ontologies from the previous step are used as input for the matching step. The purpose of this step is to find the semantic correspondence between the classes, properties, or axioms of two ontologies that are chosen to be merged. The reason behind this process is to avoid duplication in the entities after the merge. To perform this, the semantic similarity of the entities within two selected ontologies is calculated using the pre-trained Energy BERT model. During the process, the threshold for the similarity score is set to only consider entities with higher scores to participate in the merging process. The results of the classes, object properties, and data properties that are higher than the threshold value are checked to decide whether they are identical or a child class of the other counterpart manually.

F. Ontology Merging

The ontology merging is the process of combining two or more ontologies into one ontology. During the merging process, the decisions made in the ontology matching process are performed. Protégé [10] is used for merging two whole ontologies: SARGON and CIM, and the merging of a cluster of classes, seen in Figure 2, from the EM-KPI ontology. EM-KPI ontology [20] is designed to facilitate the exchange of master data and key performance indicators (KPIs) for energy management at the district and building levels, focusing solely on electricity usage. It integrates KPIs and their calculations, along with relevant master data domains, to provide a framework for tracking and improving energy performance. The ontology includes various components, such as KPI, Infrastructure, Weather, Occupancy, Observation, and Domain parameters.

G. Conceptualization

The merge between SARGON, CIM ontology, and EM-KPI ontology fulfills the requirement specifications of the use case to some extent. However, some of the concepts and relationships need to be defined more to answer as many CQs as possible. The modification and adaptation are performed on the merged ontology depending on the requirements. In this step, new concepts like *Hydrogen*, *Biomass*, *Biogas*, *Geothermal*, *Indoor Environment*, *indoorSpaceCooling*, *indoorSpaceHeating* are added as classes to fulfill the requirements in the ORSD.

H. Evaluation

Among the existing ontology evaluation tools, like OOPS [21], SARGON2 is tested and evaluated using the plug-in reasoner available in Protégé, Pellet [22], an open-source OWL-DL reasoner available which is faster in this case of a huge ontology, reliable, and easier to run the evaluation in Protégé. This reasoner is based on the tableau algorithm and supports expressive description logics. Furthermore, the ontology is tested with the OntoDebug tool [23] in Protégé, which gives the final statement of the ontology as a prompt.

I. Maintenance

The final step of the iteration is the maintenance, which includes the ontology publication and considering further requirements as needed. In this step, the ontology publication is carried out by OnToology [24]. OnToology is a web-based application that works with Git-based environments and other semantic web technologies. It provides the documentation, evaluation, publication, and versioning of the ontology. Firstly, the ontology is pushed into the git repository, and OnToology gets access to it. Subsequently, it runs the process automatically and is able to publish the ontology with the specific Internationalized Resource Identifier (IRI). The ontology is then checked with the Syntax checker on OnToology, and it gives the result that the syntax is correct. The published SARGON2 ontology can be accessed in [25]. In this step, according to the use case and user requirements, the editing

of the ontology can be done and publish the ontology as a version.

IV. SARGON2 ONTOLOGY

As mentioned in III-E, the ontology matching step is performed before the ontology merge to avoid duplication of the entities after the merging of two ontologies. For this reason, CIM and SARGON ontologies are taken into the ontology matching process under the threshold value of 0.75 Energy BERT similarity score. The concept behind it is to decide whether entities should be considered as either *EquivalentTo* or *SubClassOf*, or *SubPropertyOf*. In this work, the merging of the classes are emphasized rather than the relationships since they define the main entities and concepts. After running the matching process, 518 entities that are above the similarity score of 0.75 were detected. Following this, the manual decision on whether the entities should be put in the relevant relations is performed. As a result, 13 *EquivalentTo*, 17 *SubclassOf*, and 128 *SubPropertyOf* relationships between the entities of two ontologies are performed. Finally, at the step of ontology evaluation, the ontology debugger gives the result as "The plug-in ontology is consistent and coherent."

SARGON2 ontology has been developed based on the requirements of the Crete Valley European project. This project is envisioned as a decentralized renewable energy system that combines leading-edge ICT technologies, interoperable, and open digital solutions (including data sovereignty). Considering the requirements in the project, the ORSD for SARGON2 has been developed, having 22 CQs such as *What are the types of renewable energy resources? Hydrogen, Biomass, Biogas, Geothermal, Solar, Wind*, and 88 keywords both from CQs and answers to CQs.

By following the procedures in Section III, the SARGON2 ontology has been developed. As a result, the SARGON2 ontology consists of a total of 78,889 axioms, including 24,974 classified as logical and 14,332 declaration axioms count. Within its structure, there are 2,109 classes, 6,840 object properties, 2,277 data properties, and 3,093 individuals, as well as 14 annotation properties. Moreover, it features 2,380 *SubClassOf* axioms, with 14 *EquivalentClasses* and 5 *DisjointClasses*. The overview structure of this final ontology is shown in Figure 2.

The resulting ontology represents the energy domain, including building, and weather information based on the specified use case. Additionally, this ontology has entities on RES, consisting of hydrogen, biomass, biogas, geothermal, solar, and wind. In the evaluation phase, where SARGON2 is tested for the answerability, this ontology can answer more than 80% of CQs defined in the ORSD. Therefore, the SARGON2 ontology not only covers the necessary components of the power systems but also fulfills the requirements.

A. Application

As SARGON2 mainly focuses on the energy domain, including RES, the ontology serves as a shared vocabulary that facilitates the integration of data from different sources,

allowing disparate systems to communicate effectively. For example, by providing data interoperability, SARGON2 can significantly benefit [26], a tool which supports real-time simulation operating in the dynamic phasor (DP) approach. As the terminologies used for the specific data might be different for each stakeholder, for example, *Voltage* is used in *Company A* but *V* is used in *Company B*. In this case, DPsim may encounter difficulties in processing data if it is presented with inconsistent terminology. SARGON2 addresses this challenge by providing a standardized vocabulary and relationships, ensuring that DPsim can accurately interpret and utilize the data it receives.

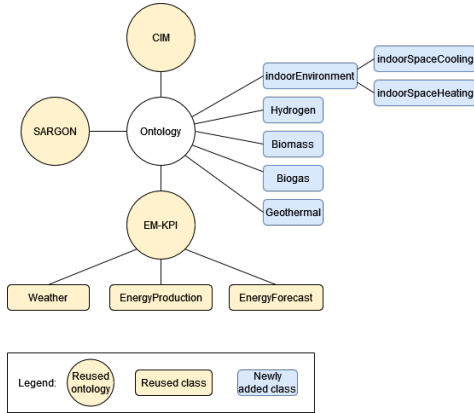


Fig. 2. General visualization of SARGON2 ontology

V. CONCLUSION

In this work, the step-by-step process of developing SARGON2 ontology by using a semi-automatic ontology development framework has been discussed. Based on the requirements of the Crete Valley project, SARGON2 has been formulated by reusing SARGON, CIM, and EM-KPI ontologies by focusing on the extension of the SARGON ontology. The Energy BERT language model has been utilized to select the related ontologies, and the result is helpful in choosing which ontology should be reused. Moreover, the process of ontology matching with Energy BERT efficiently gives the entities that have a high similarity score between ontologies in a short amount of time in order to avoid duplication. However, the extension of the SARGON ontology with the involvement of the CIM ontology and weather information classes from EM-KPI fulfilled the requirements with high answerability for CQs by presenting the required power system components according to the specified use case. As a result, SARGON2 ontology achieves high coverage in real-world datasets. SARGON2 focuses on the entities of power system components, weather-related, and building-related.

ACKNOWLEDGMENT

The work presented is based on research conducted within the framework of the Horizon Europe European Commission project Crete Valley (Grant Agreement No. 101136139).

REFERENCES

- [1] A. Dognini, A. Monti, A. Kung, A. Medela, C. Joglekar, C. Schaffer, D. Stampatori, D. Jimenez, E. Maqueda, F. Coelho, F. Mancel, F. Sedighi, G. Hartner, G. Lipari, J. Valino, J. Jimeno Huarte, L. Guitart, L. Schmitt, L. Carreras, L. Karg, M. Kollenstart, M. Santos Mugica, M. Kurz, M. Arles, M. Stroot, M. Baka, M. Galluccio, M. Bertoncini, M. Fantino, O. Hödl, O. Genest, R. Bessa, R. Dornmair, S. Jimenez, T. Tsitsanis, T. Strasser, and V. Berkhout, "Blueprint of the common european energy data space," Publisher: Zenodo.
- [2] M. Haghighi, I. Sychev, A. Monti, and F. H. Fitzek, "Sargon-smart energy domain ontology," *IET Smart Cities*, vol. 2, no. 4, pp. 191–198, 2020.
- [3] Z. Pan, Y. Gao, F. Ponci, and A. Monti, "Semi-automatic ontology development framework for building energy data management," *IEEE Access*, vol. 11, pp. 111991–112003, 2023.
- [4] International Electrotechnical Commission (IEC), "IEC 61968 Application integration at electric utilities - System interfaces for distribution management - Part 11: Common Information Model (CIM)," standard (draft), IEC. IEC Draft.
- [5] M. Poveda-Villalón and R. García-Castro, "Extending the saref ontology for building devices and topology," in *LDAC*, pp. 16–23, 2018.
- [6] L. Daniele, M. Solanki, F. den Hartog, and J. Roes, "Interoperability for smart appliances in the iot world," in *The Semantic Web-ISWC 2016: 15th International Semantic Web Conference, Kobe, Japan, October 17–21, 2016, Proceedings, Part II 15*, pp. 21–29, Springer, 2016.
- [7] Q. Khalid, A. Fernandez, M. Lujak, and A. Doniec, "SBEO: Smart building evacuation ontology," vol. 20, no. 1, pp. 51–76.
- [8] N. F. Noy, D. L. McGuinness, et al., "Ontology development 101: A guide to creating your first ontology," 2001.
- [9] M. Uschold and M. King, *Towards a methodology for building ontologies*. Citeseer, 1995.
- [10] M. A. Musen and t. P. Team, "The protégé project: A look back and a look forward," vol. 1, no. 4, p. 4.
- [11] Y. Sure, S. Staab, and R. Studer, "On-to-knowledge methodology (otkm)," *Handbook on ontologies*, pp. 117–132, 2004.
- [12] M. C. Suárez-Figueroa, A. Gómez-Pérez, and M. Fernandez-Lopez, "The neon methodology framework: A scenario-based methodology for ontology development," *Applied ontology*, vol. 10, no. 2, pp. 107–145, 2015.
- [13] M. Poveda-Villalón, A. Fernández-Izquierdo, M. Fernández-López, and R. García-Castro, "LOT: An industrial oriented ontology engineering framework," vol. 111, p. 104755.
- [14] International Electrotechnical Commission (IEC), "IEC 61970 Energy management system application program interface (EMS-API) - Part 301: Common Information Model (CIM) Base," standard, IEC, Nov. 2003. IEC 61970-301.
- [15] S. B. Neumann, J. Devos, A. N., and S. E. Widergren, "Use of the CIM Ontology, article, February 8, 2006; Tulsa," *Okla-homa*. (<https://digital.library.unt.edu/ark:/67531/metadc891396/>: accessed April, vol. 2, 2025).
- [16] M. C. Suárez-Figueroa, A. Gómez-Pérez, and B. Villazón-Terrazas, "How to write and use the ontology requirements specification document," in *On the Move to Meaningful Internet Systems: OTM 2009*, vol. 5871, pp. 966–982, Springer Berlin Heidelberg. Series Title: Lecture Notes in Computer Science.
- [17] A. Kraft, F. Engel, and A. Klinger, "Terminologies in RDM for engineering - a service approach: NFDI4ing terminology service," vol. 1.
- [18] "Smart data models - FIWARE."
- [19] Z. Pan, M. Yang, and A. Monti, "Schema matching based on energy domain pre-trained language model," vol. 6, no. 1, p. 22.
- [20] Y. Li, R. García-Castro, N. Mihindukulasooriya, J. O'Donnell, and S. Vega-Sánchez, "Enhancing energy management at district and building levels via an em-kpi ontology," *Automation in Construction*, vol. 99, pp. 152–167, 2019.
- [21] M. Poveda-Villalón, A. Gómez-Pérez, and M. C. Suárez-Figueroa, "OOPS! (Ontology Pitfall Scanner!): An On-line Tool for Ontology Evaluation," *International Journal on Semantic Web and Information Systems (IJSWIS)*, vol. 10, no. 2, pp. 7–34, 2014.
- [22] E. Sirin, B. Parsia, B. C. Grau, A. Kalyanpur, and Y. Katz, "Pellet: A practical owl-dl reasoner," *Journal of Web Semantics*, vol. 5, no. 2, pp. 51–53, 2007. Software Engineering and the Semantic Web.

- [23] K. Schekotihin, P. Rodler, and W. Schmid, “OntoDebug: Interactive ontology debugging plug-in for protégé,” in *Foundations of Information and Knowledge Systems* (F. Ferrarotti and S. Woltran, eds.), pp. 340–359, Springer International Publishing.
- [24] A. Alobaid, D. Garijo, M. Poveda-Villalón, I. Santana-Perez, A. Fernández-Izquierdo, and O. Corcho, “Automating ontology engineering support activities with ontoology,” *Journal of Web Semantics*, vol. 57, p. 100472, 2019.
- [25] “Sargon2 ontology.” <https://w3id.org/def/sargon2>, 2025. Accessed: (2025-04-08).
- [26] M. Mirz, S. Vogel, G. Reinke, and A. Monti, “Dpsim—a dynamic phasor real-time simulator for power systems,” *SoftwareX*, vol. 10, p. 100253, 2019.