DAAD

Scientific Progress Report

BY

Charles Chukwuemeka William Emehel

91759575

PhD Student

RWTH Aachen University

November 2023

Ontology Learning
in
Smart Energy Grid

Table of Content

1.0 Introduction

This scientific report is on the progress of my research on semantic interoperability in the smart energy grid by building and updating a Knowledge Graph using machine learning. . in the smart energy grid. It will also highlight the presentation, conference attended and papers publish. I will also talk about some of the pending journal papers I am currently writing and my developed architecture for the Quantum Ontology Learning. I will also mentioned some the development of and Artificiall Intelligence on Deman ontology that was developed and the plan to update its metadata for use as a reference standard in ontology development for any energy use case. The significance of the research is looked into based on the need for data integration with the growing amount of data on the traditional data bases and on the semantic web. The aims and objective of the research is highlighted as a guide for the research with the conclusion of the section stating the contribution to knowledge. Section 2 is on literature review on semantic web, ontology engineering, machine learning, and smart energy grid. Section 3 is my current research on exploiting machine learning methods in ontology engineering and the development of a python-based web application and service tool for the implementation of the algorithms and result validation for semantic interoperability and data integration of the data from devices on the smart grid which will enable automation and autonomous operations of the grid components. Section 4 is the conclusion with a short statement of future work.

Figure 2 – Ontology Engineering Work flow [1]

2.0 Knowledge Graph

Knowledge Graph is a graph that contains interpretable context information with the aid of concept and relations description in hierarchies. Ontology engineering is the different activities that builds, maintains or carries out analytics or inference on ontology data. Ontology engineering helps man and machine to model data in a formal explicit and specified way for data sharing and conceptualization in a domain of interest. [2]. Since knowledge changes over time and domain always evolve, ontology engineering activities has to keep up with this change over the life cycle of the building of the ontology and provides the ontology developer an intelligent change model for expressing the local changes of this ontology in a descriptive way. One of the ontology engineering activities is alignment of links between RDF graphs as illustrated in the figure 2 below.

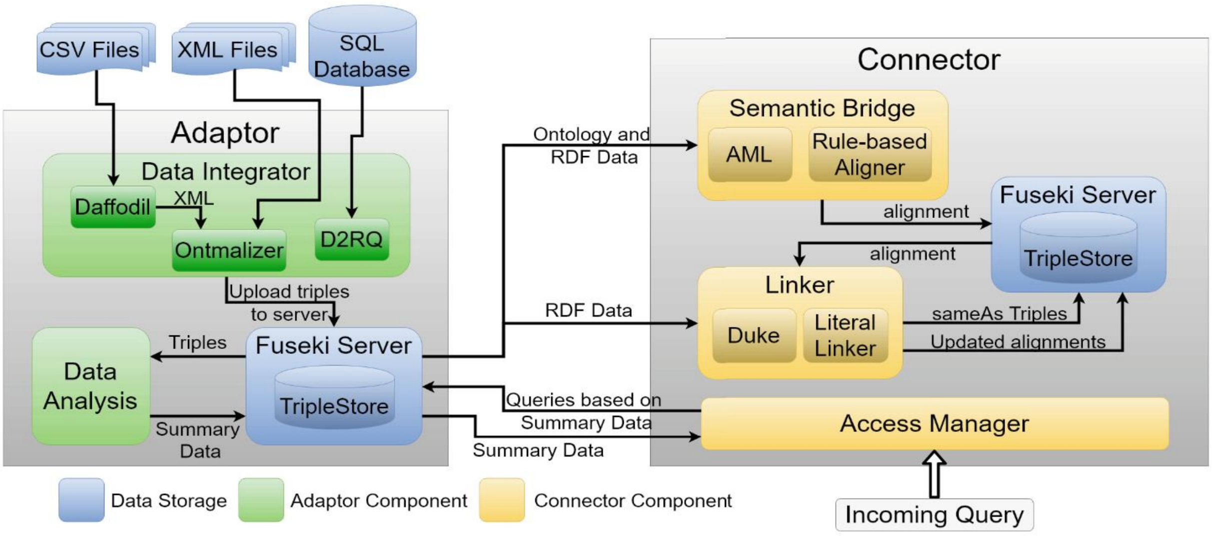


Figure 2 – Ontology Engineering Work flow [1]

Building Knowledge graph with ontology standard and schema is a way for ensuring or guaranteeing a realistic and convergeable multi domain knowledge graph. This will reduce noise and haluciantion in making decision and automating a critical infrastructure like the electric grid.

3.0 Ontology Development

During the course of my research and work at RWTH, I developed and ontology for Artificial intelligence on Demand (AIoD) and also validated and ontology which I reviewed as a deliverable and proposed a paper on it.

3.1 Artificial Intelligence on Demand Ontology (AIoDO)

The Artificial Intelligence on Demand Ontology (AIoD) was developed as a proposal for the Artificial Intelligence 4 European Union (AI4EU). The ontology development is based on a hybrid approach involving Unified Process for ontology (UPON) and Ontology Design pattern (ODP). The UML part of the UPON approach was developed based on the CIM profile and requirements of each use case as a pre ontology modelling to achieve a progressive evaluation in the I-NERGY ontology as already reported in D3.2 *I-NERGY Data Interoperability and Processing Management Services (2nd Tech rel.)*. This increases ontology correctness and ontology quality as well as serve as an aid for the AIoD platform application development.

The reasons for the meta modelling and hybrid development are due to the uniqueness of the requirements in the use cases, the payloads from pilot sites, the reuse of existing ontologies like CIM, SARGON, etc and to capture terms that could be related to terms in data analytics and objection function optimization. These forms the meta ontology design pattern in **Error! Reference source not found.**. This approach will help in any future recurring ontology modelling.

The meta model or meta ontology modelling framework in **Error! Reference source not found.** is used to design all the sub ontologies for the 15-use case of the I-NERGY project. The I-NERGY ontology development process also followed a bottom-up approach since existing standard ontologies have been reused. Use case concepts and properties are classified to their super classes or inherited concepts according to the terms from pilot datasets or user requirements. The ontology modelling for the 15 use cases is on the conceptual heterogeneity level to show the uniqueness of the terms or vocabulary be it new or already defined in other standard namespaces. This ensures the richness of the ontology for both the Smart grid and the smart building application for the AIoD platform. Ontology Alignment is done on the use case sub ontology and on the pilot level to achievement a common ontology for the I-NERGY project.

The I-NERGY Information System Model or Ontology was developed with Open World Assumption (OWA) on all the 15 use cases across 9 pilots for easy reuse and extensibility. It is developed with standard vocabularies and newly defined vocabularies for easy information sharing and data integration in the building and Grid domains. It will also enable having a common knowledge representation and an aid in performant data analytics among the machine learning developers as well as all stakeholders in each pilot as regards data integration and semantic interoperability as well as in semantic annotation during the definition of new data model. This will help in feature labelling during the use of classification or clustering algorithm or other statistical data analysis in the development of machine learning models.

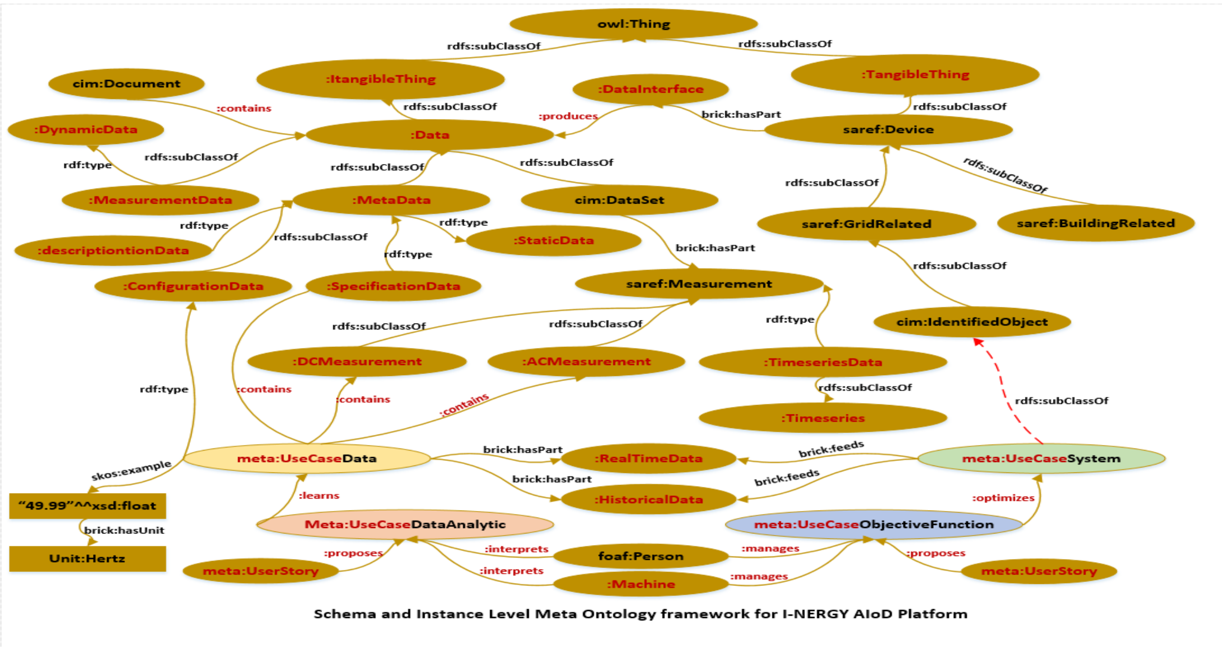
The competence question that triggered the ODP methodology for the I-NERGY ontology development are:

- Does the dataset or metadata from the pilot site match the attributes of the concept or device producing the data?
- What are the classes, properties and instances of the concepts, data and the relation which are classified as intangible and tangible things?

- Which feature or target variables for data analytics and which objective or constraint function for optimization is of interest from the user requirements and how do they relate to the data and concept that produced it.

Error! Reference source not found. below is the meta model or meta ontology modelling framework for all use case with vocabularies from SAREF, SARGON, BRICK, FIBO and CIM. Standard prefixes with the local name or vocabularies are in black while those that are defined for the I-NERGY project are coloured in dark shade of red (oxblood).

From **Error! Reference source not found.**, The **meta:UseCaseData**, **meta:UseCaseSystem**, **meta:UseCaseDataAnalytic** and the **meta:UseCaseObjectiveFunction** are all multi-term meta concepts and follows a logical ODP modeling that can be in any domain, e.g., or the building domain. The concepts and relations in red are those that could not be found, during mapping, as existing terms or relations in standards like **CIM**[1], **SAREF**[2], **SARGON**[3], **BRICK**[4], etc. They form the I-NERGY vocabularies that can be used on the I-NERGY Namespace. With this meta ontology framework, use case concept and use case data can be added or removed based on information extraction from user requirement in the data description documents of each use case and the pilot dataset. Some concepts are modeled as a blank node which is a standard placeholder to put a variable or unknown concept or possible concept due to lack of more information during information extraction from pilot data or use case requirement. During ontology alignment or linked data analytics, these blank nodes can be given unique Identifications for possible inferred concept with either logic, linguistic or statistical algorithms. The ontology for all the use case is implemented in Protégé[5] and these forms the common ontology for the I-NERGY project.



Schema and Instance Level Meta Ontology framework for I-NERGY AIoD Platform

Figure 3 – Ontology Engineering Work flow [1]

3.2 Cyber Physical Smart Grid Ontology (CYSGO)

[1] CIM: https://www.dmtf.org/standards/cim
[2] SAREF: https://saref.etsi.org/
[3] SARGON: https://sargon-n5geh.netlify.app/
[4] BRICK: https://brickschema.org/
[55] Protégé : http://protegeproject.github.io/protege/getting-started/

From my research I found out that for a realizable semantic interoperability to be achieved, the ontology development process need to follow standards. In developing Cysgo, I adopted BFO (ISO/IEC21838-4) for my TLO. This will help me achieve a correct-by-construction true converges to happen between different domains during ontology learning. One of the important process is to adopt single inheritance so as to be able to achieve: computational performance benefit, Aristotelian structure, ontology structure managemen, automated ontology alignment and the possibility of achiving multiple inheritance only after reasoning or during inferencing. This is during the running of the application involving the Ai agent making decision or in helping a decision support system or during actionable insight on the smart grid. The identified object concept is modeled as a subclass of bfo:object so as to follow the Aristotelian definitional structure for all child concepts on the cim controlled vocabulary concepts. This will enable automatric alignment for equivalent concept as well as achieving taxonomical subsumtion on the cim path for all the cim concepts.

Cysgo starts with a seed or base ontology on which the learning ontology will be anchored. The seed ontology will be the referenc e ontology which will use BFO as the TLO to acieve a realistic semantic interope=reable ontology that can easily converge. Then the ontology leanring process will generate learned ontologies as axions or fact that will ride on the seed ontology. The seeed ontology will be designed using the Content Ontology Design Patter Content-ODP or CP. The seed ntology is developed with Protégé while the ontology learning will be developed with Python from the loaded ontology eexported from Protégé.

Cysgo is an ontology developed with the above features taking it to consideration.cysgo aims to detect malicious activities in the information flow among the smart grid component like stamrt convrters with OLTC and PMUs. It uses the BFO as the TLO so as to be able to achieve convergences with other cyber security domain concepts like threats and vulnerability. Cysgo extract the domain knowledge about expected activities in the sensed and monitored system and applies reasoning to recognize and classify threats from adversarial attacks. The characteristics and properties of the smart converters and the smart meters as entities in the smart grid are matched to the payload data exchanged between the field devices, the edge devices and the backend servres to detect any anomaly in beahviours of what they are normally supposed to once theier etstae is monitored. The ontology serves as numerial and lexical pror analytics for all the ongoingss in the power and communication network.

The essence of using an ontology schema based knowledge grapgh is to detect an advanced adversial attacker that might decide to employ generative AI in creating an utopian scenario that looks so real in other to evade surveillance and detection of its malicious activities.since the cysgo ontology has ontology learning capability linked to the NVD and Att&ck Mitre data base, updates to the cysgo knowledge base will always happen when new threats is updated on the databse. Also using the Grover's algorithm searching capapvility on the semantic web information will be extarcked and updated on the cysgo ontology. Software application development following an object oriented approach always deals with state and bevaviors of entieties. The state estimation of the converter attributs of voltage, current, frequency, phase angle and power factor by the PMU will be extracted from the measurement data and these data matched with the ontology based knowledge graph.from instance, from fig 1, when a malicious attacker carries out a threat due to an exploited vulnreabity that was initiated or discovered which exposes a smart grid asset. This could have an adverse impact on the grid. The reasoner could prvide an inference that will enable the detection of this threat before it occurs. Cysgo can detect an advanced adversiary maliciuos actor that uses any of the nature inspired algorithm like Dragon Fly or Smell Agent algorithm to compromiase the smart electric grid. This can be achived by…

Quantum Ontology Learning is the use of Grover's quantum algorithm and ressources to update an ontology for scalability and efficiency. The work of [3] proves the workability of ths approach. See uploaded team presenation on the introduction of this approach. Machine learning algorithms and techniques can be applied to ontology learning as well as for the maintenance of ontology and the performance of analytics on knowledge base. The work of [4] shows the use of machine learning algorithm for searching vectorized RDF data on the semantic web. The work of [5] shows how a cluster of entity concept can be visualized through embedding learning of linked data. In demonstrating the use of machine learning for link prediction and concept classification the work of [6] and [7] shows that this can be applied in as part of ontology learning. Large corpus of data can be used to train graph neural network with an intermediate embedding being used for the training of the neural network. Distributed training is very is inevitable to handle large corpus and to shorten the training time.[8]

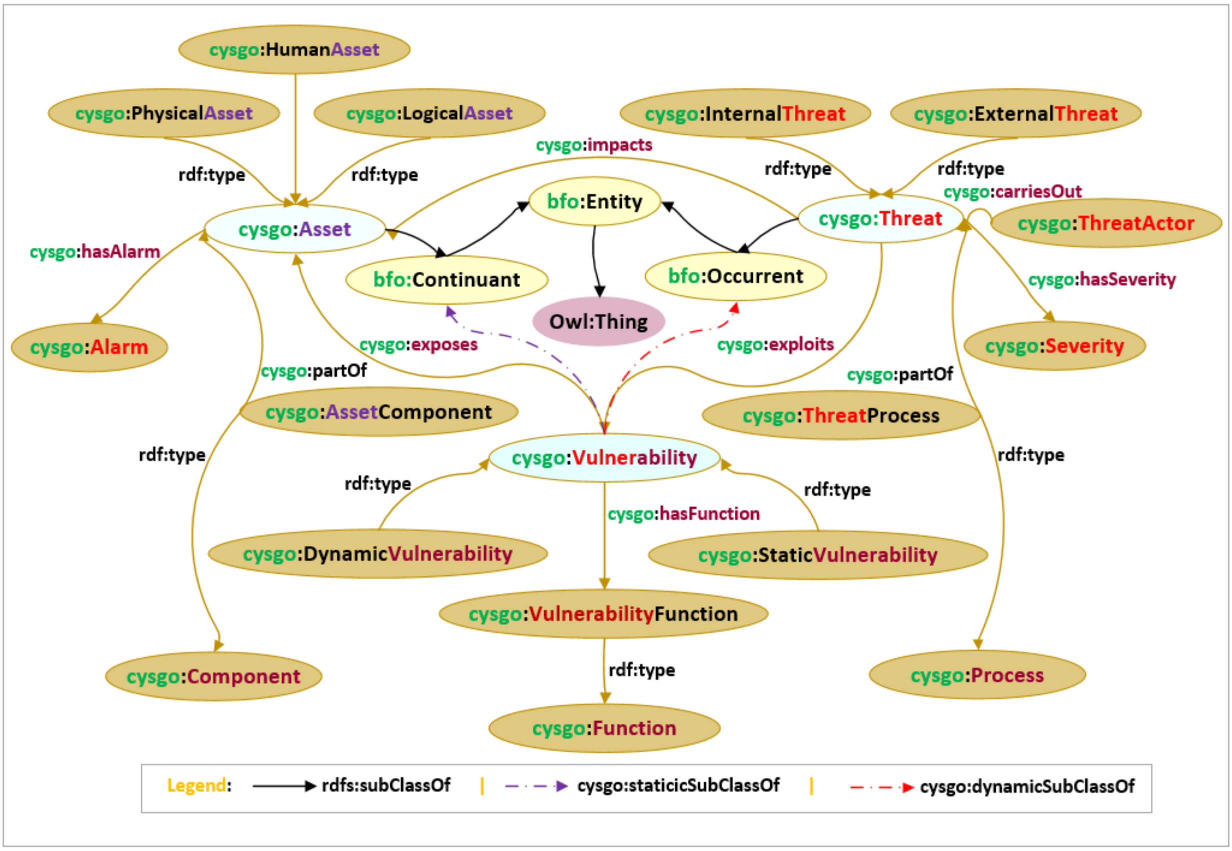


Figure 2 – Ontology Engineering Work flow [1]

3.3 Hybrid AC & DC Grid Ontology (HADGO)

The Hybrid AC & DC Grid Ontology was [9] developed by research partners from the Austrian Institute of Technology (AIT) and Engineering (ENG). I reviewed the ontology and proposed that we could develop a paper out of the ontology. The paper was written and my contribution on the paper was the validation of the ontology with a use case scenario using the HermiT Reasoner on Protégé and obtained an inference result from the Ontology after the inconsistencies where debugged. I also discussed the results and wrote the conclusion as well as the future work and updated the Abstract with the result. I also proposed the change of the name from an ***Experimental Study of a Hybrid AC/DC Grid Ontology*** to ***Development and Application of a Hybrid AC/DC Grid Ontology*** due to the validation and achievement of logical inferencing and Reasoning on the developed Ontology. The Paper was selected as one of the papers to be upgraded to a journal due to the level of work done. This is currently being worked on for the upgrade. The current conference paper has been uploaded to the DAAD portal.

4.0 Ontology Learning

My plan is to update the above three developed ontology using the Ontology Learning. Cysgo Ontology is already developed using the Aristotelian Definitional Structure with Single Ingeritance and will be easy to nparform ontology alignment, etc. AIoD and HADGO will have to be updated so that ontology learning can be performed on them. The steps involved in ontology learning are:

- Natural Language Processing (Resource Processing Stage)
- Machine Learning or Deep Learning (semantic learning stage)
- Knowledge Representation and Reasoning (Formation of Ontology Schema and Reasoning)

- User Interface Management (Knowledge Graph Management)

I started my research with using unstructured dataset to use the information on it to update the ontologies developed above. The algorithm for the implementation was based on Jaccard similairity using the coeefeicnt computed from the intersection and union of the characters in the terms from the dataset. The Jaccad index is obtained and the a threshold is used to trigeer the terms to be selected as rekated and this to be deiscarded. The result is then used to form an RDF triple and saved on a triple store as an RDF graph data.

Then another approach was the use of information extraction algorithm with the spaCy NLP package.

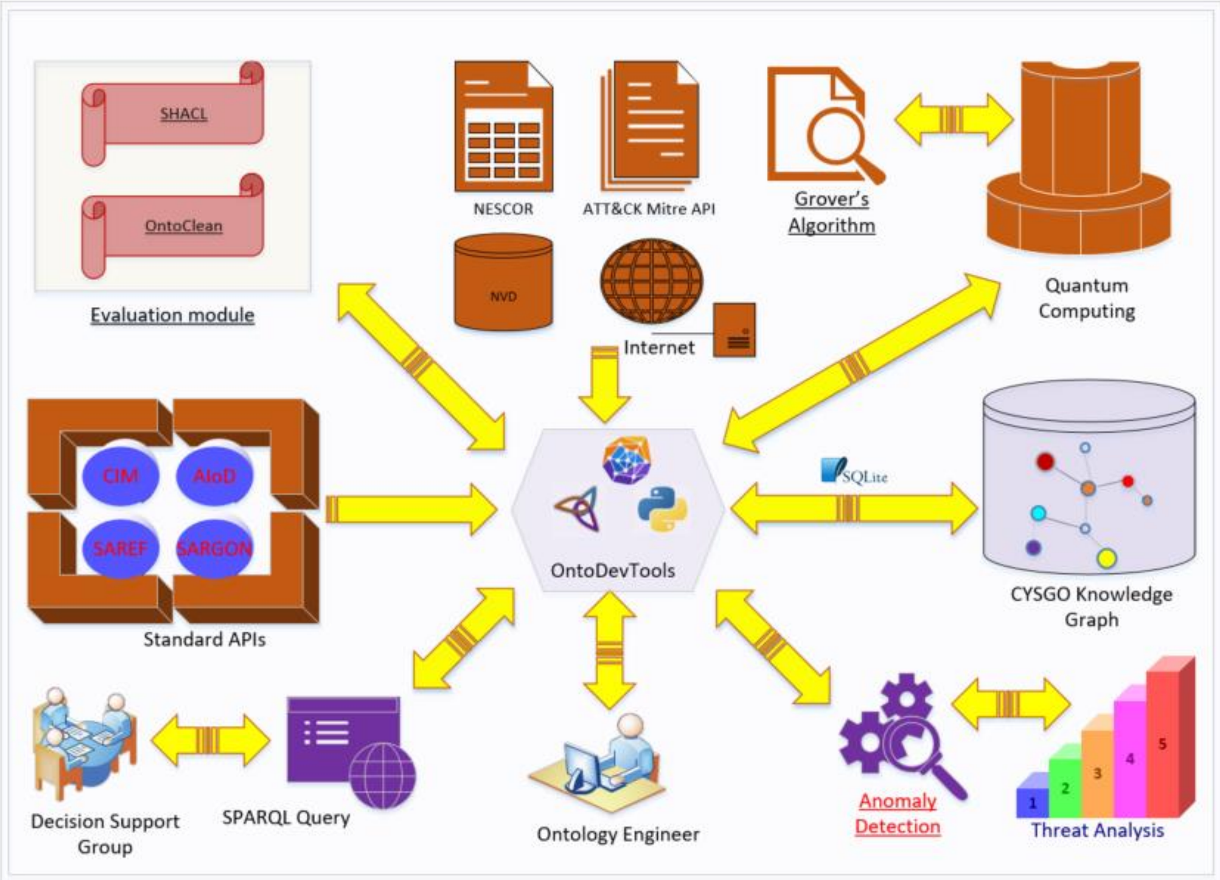
Then I used pretrained transformer models in the resource processing statge

Quantum Ontology learning in artificial intelligence is a scalable approach of efficiently scaling the ontology leaning to deal with large data and also have the capapbility of using large language model algorithms like self Attention Mechanism and Bidirectional Encoder Representations from Transfomers (BERT). The advantage of this approach can be applied to Data Space and Digital Twin research and applications. One of the complex part of ontology learning is the combination of statistics, probability and logic. The following discipline are the foundation of ontology learning:

- Philosophy
- Natural Language Processing
- Machine Learning
- Graph Theory
- Semantic Web
- Data Mining
- Cryptography
- Predicate Logic

An robust ontology learning architecture will ensure a versatile knowledge graph that will scale with the complexity of the information system in an intelligent automation infrastructure.

4.1 Ontology Learning Architecture



4.2 Ontology Learning Algorithm

5.0 Scalable Ontology Learning

The next step will be the following:

- Complete my development tool and implement the first integration of Grover's algorithm to Ontology Learning 3rd quarter of 2024
- Publish more paper on Quantum Ontology Learning.
- Publish a conference paper by 1st quarter of 2024
- Publish a journal paper by 2nd quarter of 2024

6.0 Use Case Scenarios To Be Considered In My Research

Entirety

Open ModeliCa

RTDS

7.0 Research paper and Publish TimeLine

8.0 Challenges

Some of the challenges experienced during my research are :

- Lack of dataset due to GDPR
- Lack of validation tools for ontology learning development
- Information modeling is usual different from differnet perspective of use case scenario

9.0 Reference

6.0 References

[1] F. Herrera, K. Matsui, and S. Rodríguez-González, Eds., *Distributed Computing and Artificial Intelligence, 16th International Conference*, vol. 1003. in Advances in Intelligent Systems and Computing, vol. 1003. Cham: Springer International Publishing, 2020. doi: 10.1007/978-3-030-23887-2.

[2] M. Pritoni *et al.*, 'Metadata Schemas and Ontologies for Building Energy Applications: A Critical Review and Use Case Analysis', *Energies*, vol. 14, no. 7, p. 2024, Apr. 2021, doi: 10.3390/en14072024.

[3] S. Naseem, 'Dynamic ontologies evaluation framework using quantum perceptron neural network', in *2015 International Conference on Open Source Systems & Technologies (ICOSST)*, Dec. 2015, pp. 151–157. doi: 10.1109/ICOSST.2015.7396419.

[4] A. S. Hadi, P. Fergus, C. Dobbins, and A. M. Al-Bakry, 'A Machine Learning Algorithm for Searching Vectorised RDF Data', in *2013 27th International Conference on Advanced Information Networking and Applications Workshops*, Mar. 2013, pp. 613–618. doi: 10.1109/WAINA.2013.204.

[5] Z. Wu, 'Automatic Clustering and Visualization of Linked Data through Embedding Learning', in *2019 International Conference on Intelligent Computing, Automation and Systems (ICICAS)*, Dec. 2019, pp. 778–781. doi: 10.1109/ICICAS48597.2019.00167.

[6] K. Juszczyszyn, G. Kołaczek, and D. Dudziak-Gajowiak, 'Structural Analysis and Link Prediction in Dynamic Networks of Web Services', in *2017 IEEE 26th International Conference on Enabling Technologies: Infrastructure for Collaborative Enterprises (WETICE)*, Jun. 2017, pp. 144–149. doi: 10.1109/WETICE.2017.55.

[7] P. Minervini, N. Fanizzi, C. d'Amato, and F. Esposito, 'Scalable Learning of Entity and Predicate Embeddings for Knowledge Graph Completion', in *2015 IEEE 14th International Conference on Machine Learning and Applications (ICMLA)*, Dec. 2015, pp. 162–167. doi: 10.1109/ICMLA.2015.132.

[8] S. Wang, 'DISTRIBUTED MACHINE LEARNING', Accessed: Aug. 17, 2021. [Online]. Available: https://www.academia.edu/35877759/DISTRIBUTED_MACHINE_LEARNING_STANLEY_WANG_SOLUTION_ARCHITECT_TECH_LEAD_at_SWANG68

[9] A. Rossi *et al.*, 'Development and Application of a New Ontology in the Context of Hybrid AC/DC Grids', presented at the IARIA Congress 2023, The 2023 IARIA Annual Congress on Frontiers in Science, Technology, Services, and Applications, Nov. 2023, pp. 58–64. Accessed: Nov. 27, 2023. [Online]. Available: https://www.thinkmind.org/index.php?view=article&articleid=iaria_congress_2023_1_120_50112