Leveraging AI for Enhanced Semantic Interoperability in IoT: Insights from NER Models

Mohammad Ali Nemer, Joseph Azar, Abdallah Makhoul, and Julien Bourgeois

Femto-St Institute, UMR 6174 CNRS, Université de Franche-Comté, France
Emails: mohammad_ali.nemer@univ-fcomte.fr; joseph.azar@univ-fcomte.fr; abdallah.makhoul@univ-fcomte.fr; julien.bourgeois@univ-fcomte.fr

***Abstract*—In Industry 4.0, achieving semantic interoperability is a significant problem due to the complexities of current automation systems and the numerous standards involved. The study explores how Artificial Intelligence (AI) and semantic interoperability connect within the Internet of Things (IoT) framework to overcome barriers to technology adoption. The main goal is to analyze how AI's adaptive and predictive abilities might transform semantic interoperability by studying AI-driven methodologies to provide a flexible and efficient solution. The main objective of the paper is to leverage Named Entity Recognition (NER) AI models to streamline the identification of entities within the Internet of Things (IoT) for achieving semantic interoperability. It tests a Natural Language Processing (NLP) translator on data representations not seen during training, and the outcome highlights the efficiency of NLP in correctly understanding and processing these representations.**

***Index Terms*—Internet of Things, Semantic interoperability, Natural Language Processing, Cyber Physical Systems, Artificial Intelligence**

I. Introduction

The IoT ecosystem is growing rapidly and includes devices ranging from household appliances to industrial sensors [1] . As of 2023, the number of connected IoT devices worldwide has reached billions, which is expected to grow exponentially. This growth, while promising, also creates complexities and challenges in maintaining interoperability between disparate systems [2].

Interoperability plays a pivotal role in ensuring seamless integration and communication between diverse devices and platforms. Interoperability is achieved generally at five distinct levels:

1) **Protocol-level standardisation:** This standardisation refers to the process of establishing uniform technical standards or protocols for data exchange between systems. It ensures that different systems can communicate with each other at a basic level. An example of protocol-level standardization is the Internet Protocol Suite, commonly known as TCP/IP (Transmission Control Protocol/Internet Protocol).
2) **Platform Interoperability:** This level of interoperability ensures that different software components can work in conjunction within a unified framework. Containerization is one of the key methods to achieve this, as it allows for the encapsulation of software in a way that is portable and consistent across different computing environments.
3) **Syntactic Interoperability:** Which focuses on achieving a standard structure or format for data exchange to avoid vendor lock-in scenarios. By standardizing the way in which data is formatted and exchanged, organizations can ensure that their systems are not overly dependent on a single vendor's technology, thus promoting flexibility and choice in their tech stack. An illustrative example of this approach in action is the use of Zigbee2MQTT middleware [3]. Zigbee2MQTT acts as a bridge between Zigbee devices and a MQTT server, translating Zigbee communication into a standard MQTT format.
4) **Semantic Interoperability:** Beyond syntactic consistency, understanding the underlying meaning or context of each data packet becomes crucial at the edge node. This level is concerned with ensuring that the semantics of exchanged information are understood by both the sender and the receiver.
5) **Organizational interoperability:** Organizational interoperability is the exchange of important data between multiple organizations, regardless of their geographical locations, infrastructure, and information systems [4].

The emergence and advancement of Artificial Intelligence (AI) has introduced promising pathways to address interoperability challenges. AI's ability to understand, process, and respond to natural language, patterns, and contexts makes it uniquely positioned to bridge the semantic gaps that often impede interoperability within IoT [5]. This paper addresses the challenges of technological adoption in Industry 4.0 by exploring AI-driven techniques that provide a flexible and efficient solution to achieve semantic interoperability in modern automation systems with diverse standards. It tackles semantic interoperability and translation beyond just exchanging data. The objective is to ensure the data is meaningful and usable across different systems. This involves translating information models and standards that define the semantics of data and services so they can be correctly interpreted by various industry sectors, products, vendors, etc., even if these entities use different representational systems.

A. Motivation

Expanding on Nilsson's prior work on interoperability with the Machine-to-Machine (M2M) translator [6], we were motivated to advance our research by creating a Named Entity Recognition (NER) model. The NER model is designed to recognize and retrieve items from the extensive data within the IoT communication ecosystem. We aim to leverage NER technology to enhance data comprehension and facilitate improved translation and communication among interconnected devices and systems. This motivation stems from the understanding that precise entity extraction is pivotal for enabling seamless and meaningful translations within the complex network that converts data to a unified format at the gateway or the edge, ultimately enhancing interoperability and functionality within the Internet of Things (IoT) environment.

B. Contribution

In the context of managing multiple interconnected systems (A, B, C, etc.), a significant challenge arises in efficiently collecting and storing entities extracted from various data sources. The problem involves developing a robust data management solution for the semantic extraction, organization, and storage of entities into a unified storage. This article discusses our work in connecting Natural Language Entity Recognition (NER) with the Internet of Things (IoT) to improve semantic interoperability in the IoT environment. We are developing a NER middleware translator that can function with many IoT devices and their many data formats, including JSON, XML, and YAML. This translator is trained on data from different real sensing devices with unique data representations. It is tested on diverse datasets with different formats and structures. Central to our solution is a generic data processing and preparation middleware. The middleware is intended to clean and standardize incoming data, regardless of its initial format or structure, to provide a uniform and consistent input for subsequent processing. After completing the first data preparation phase, we implemented an advanced (NER) model. This model aims to precisely identify and categorize entities in the data (temperature, humidity, GPS location, pressure, etc) and efficiently convert them into a unified data structure.

The organization of this paper is as follows: Section II provides a comprehensive overview of AI and semantic interoperability. Section III presents some of the related work on semantic interoperability. Sections IV and V presents the detailed contributions of this paper and the implementation methodology of the proposed technique respectively. Section VI presents the experimental results, and finally, Section VII concludes the paper.

II. Background

Industry 4.0 represents the fourth industrial revolution, characterized by the integration of digital technologies, automation, and data exchange in manufacturing environments. Semantic interoperability is a cornerstone of Industry 4.0, enabling seamless communication between various components such as machines, sensors, and human operators.

Several solutions have been developed to address semantic interoperability challenges in Industry 4.0, including ontology, standardization initiatives, middleware solutions, semantic web, and machine learning.

1) Ontology-based approaches: Utilizing ontologies to define common vocabularies and relationships, enabling machines and systems to understand and interpret data consistently. For example, check the work of Kelly et al. [7].

2) Standardization initiatives: Adopting industry standards and protocols to ensure data is exchanged in a universally understood format. For example, check this work for reference [8].

3) Middleware solutions: Implement middleware layers that act as translators between different systems, ensuring data is interpreted correctly across diverse platforms. For example, check this work for reference [9].

4) Semantic web: Employs ontology-based approaches, leveraging ontologies to establish shared vocabularies and associations, facilitating uniform data comprehension and interpretation across machines and systems. Thus achieving seamless integration and smarter interactions across the IoT ecosystem. For reference check [4].

5) Artificial intelligence: Artificial Intelligence (AI) has significantly enhanced interoperability across various domains. AI's capabilities in machine learning, natural language processing, and real-time analysis enable the development of intelligent systems that can autonomously handle inconsistencies and changes within interconnected devices. For example, Nilson et al. [10] combined interoperability and M2M translation in this work.

III. Recent Work on Semantic Interoperability

Recent advancements in Neural Machine Translation (NMT) and Semantic Interoperability have significantly contributed to addressing the challenges of data heterogeneity and ensuring seamless communication across various domains, including the IoT and healthcare. This section synthesizes key findings from recent literature, highlighting innovative approaches and methodologies.

In the domain of IoT, semantic interoperability mechanisms have been a focal point, with researchers proposing various solutions to enhance communication and integration. Rahman et al. [4] and Santo et al. [11] both highlight the need for semantic models and interoperability solutions in IoT, with the latter specifically discussing the integration of NMT in IoT devices. Lakka et al. [12] and Venceslau et al. [13] propose semantic interoperability mechanisms and review the state-of-the-art in this area, while Pliatsios et al. [14] introduces the concept of Semantic Social Network of Things (SSNT) to support open, interoperable IoT environments. Balakrishna et al. [15] and Novo et al. [16] further explore semantic approaches for IoT data integration and the extension of the Web of Things architecture to enhance semantic interoperability. Lastly, Gui et al. [17] presents a data-driven natural language interface for Industrial IoT use cases, which could potentially benefit from NMT advancements.

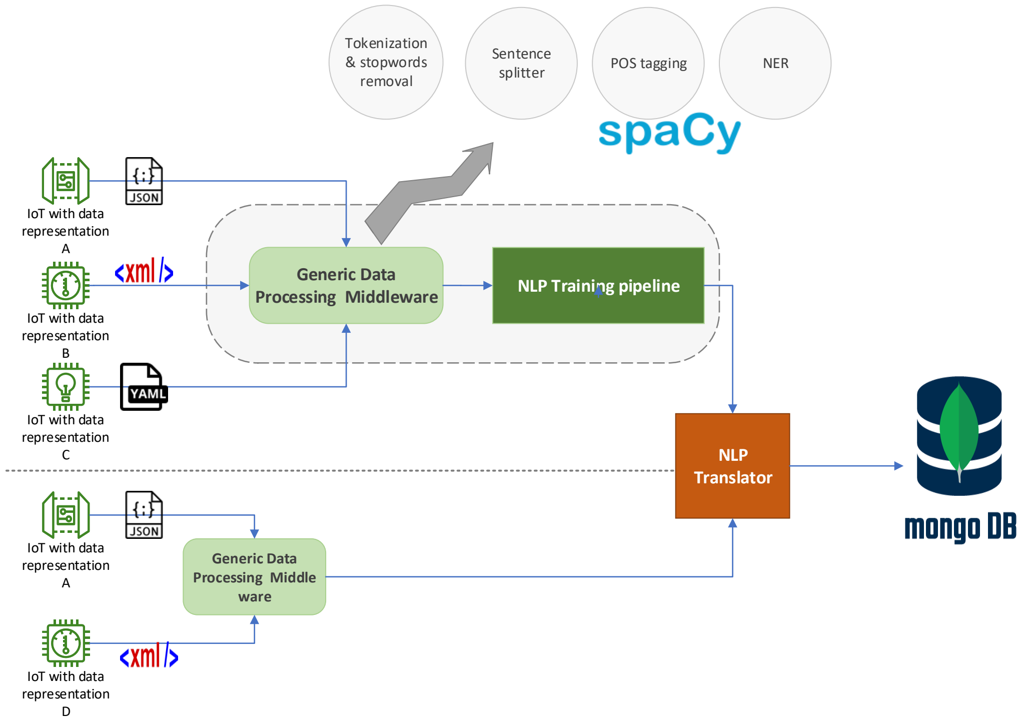

Fig. 1: The proposed framework for semantic interoperability.

Parallelly, the field of NMT has seen remarkable progress with the integration of deep learning models and semantic knowledge. Tian et al. [18] and Sharma et al. [19] both highlight the potential of deep learning models in NMT, with [18] specifically focusing on the use of ontologies to enhance translation speed and quality. Nguyen et al. [20] and Rapp et al. [21] explore the integration of semantic information into NMT models, with [20] demonstrating the effectiveness of abstract meaning representation (AMR) semantic graphs and [21] using semantic role labeling (SRL) to improve translation performance. Sulem et al. [22] and Moussallem et al. [23] propose methods for enhancing NMT through semantic structural decomposition and knowledge graph augmentation, respectively. Finally, Yin et al. [24] and Li et al. [25] address the issue of compositional generalization in NMT, with [24] introducing categorization to source contextualized representations and [25] proposing a neuron interaction-based representation composition approach. These studies collectively underscore the potential of deep learning and semantic knowledge integration in advancing NMT.

Despite these advancements, several limitations and challenges persist. Key among them is the need for more robust models that can handle the vast diversity of IoT devices and data formats. Achieving high levels of semantic interoperability requires models that can adapt to new, unseen data structures and languages. Additionally, the integration of semantic information into NMT models remains a complex task, necessitating further research to optimize these processes for real-world applications.

IV. Proposed Approach

We define the Neural Machine Translation (NML) interoperability problem in terms of Named Entity Recognition (NER), which identifies readings from each device. For instance, consider Device A outputting in JSON format, Device B in XML, and Device C in YAML, among others. The goal is to convert these outputs into a unified format. Generally, the translated messages from Devices A, B, and C may not be semantically and structurally identical to the original messages communicated. However, we can enhance the translator to recognize entities and transform them into a unified format. As illustrated in Figure 1, Devices A, B, and C, each with different output formats, send their data to an NLP middleware. The middleware then detects entities and generates a standardized format, which is then saved on the edge/cloud.

V. Implementation Methodology

This section describes the methodology used in this paper and covers model selection, dataset preprocessing, and labeling.

A. Model Used:

In the context of Natural Language Processing (NLP), spaCy [26], an advanced open-source library, offers a particularly powerful tool for Named Entity Recognition (NER). NER in spaCy is designed to identify and categorize key information in text, such as names of people, places, organizations, as well as expressions of times, quantities, monetary values, percentages, and more. This functionality is critical in a wide range of applications, from information extraction to content classification and entity linking.

We used this model configuration which based on "bert-base-uncased" and is designed for token classification tasks, utilizing "BertForTokenClassification." It consists of 12 transformer layers with hidden size 768 and 12 attention heads. The model employs GELU activation and 0.1 dropout probabilities for both attention and hidden layers. Absolute position embeddings are used, with a maximum position embedding of 512 and a vocabulary size of 30522. The configuration supports token classification labels and enables caching during computation.

B. Dataset and pre-processing

In this work the data was publicly available data on Data.Gov [1]. We chose some files with Iot data with several formats json, xml which are automated sensors readings with multiple formats, the data is under the name (The Array of Things). (AoT) is an experimental urban measurement system comprising programmable, modular "nodes" with sensors and computing capability so that they can analyze data internally, for instance counting the number of vehicles at an intersection (and then deleting the image data rather than sending it to a data center). AoT nodes are installed in Chicago and a growing number of partner cities to collect real-time data on the city's environment, infrastructure, and activity for research and public use. The concept of AoT is analogous to a "fitness tracker" for the city, measuring factors that impact livability in the urban environment, such as climate, air quality, and noise.

Pre-processing serves as a crucial preliminary step in the development of NER models. This pivotal preprocessing stage is designed to refine and optimize text data, ensuring its suitability for subsequent NER tasks. We implemented various preprocessing techniques, including the removal of unnecessary characters, tokenization to segment text into meaningful units, normalization of decapitalization, and potentially applying advanced linguistic processes like lemmatization or stemming. Additionally, we tailored specific preprocessing steps to address the nuances inherent in NER tasks, such as managing entity mentions, resolving coreferences, or standardizing entity representations. The output is illustrated in Figure 2.

C. Data labeling

It involves annotating words or phrases in a corpus to indicate their respective entity types such as device Id, locations, organizations, dates, and more. This process requires human interference to assign appropriate labels to each entity instance, ensuring consistency and accuracy throughout the dataset. Effective data labeling not only enhances the performance of NER models but also contributes to the development of robust natural language processing applications across various domains. Entity annotation involves manually or semi-automatically labeling the text data with named entity tags, indicating the boundaries and types of entities present. for example *(17, 32, 'TEMPERATURE')* which indicated the start and the end of the temperature entity in the input text. Annotated data serves as the training set for supervised NER models, guiding them to recognize named entities accurately during training and inference.

D. Data preparation

In our data preparation process, we begin with text cleaning as the first step. Text cleaning serves as the initial stage of preprocessing where we aim to remove noise and irrelevant information from our text data. Our common text cleaning techniques encompass lowercasing, removing special characters, and eliminating stopwords. Lowercasing ensures consistency in casing, while removing special characters and stopwords aids in reducing noise within the text.

Moving on to the next step, we engage in Tokenization. Tokenization involves the segmentation of text into individual words or tokens. Here, spaCy offers efficient tokenization methods, encompassing word and sentence tokenization. Our tokenizer adeptly manages complex tokenization scenarios such as hyphenated words and contractions, thus contributing to accurate Named Entity Recognition (NER) outcomes.

Proceeding to Step 3, we delve into Part-of-Speech (POS) Tagging. This stage leverages POS tagging capabilities to furnish contextual information. POS tagging proves crucial for comprehending context within our data. For instance, in the sentence "The temperature outside is 29 degrees," POS tagging aids in capturing the relationship between the words "temperature" and "29 degrees" as they are contextually linked. Moreover, POS tagging may encounter influences from factors like word ambiguity, context dependency, and syntactic variations.

Our final step encompasses linguistic annotations aimed at normalizing words, sentence segmentation for dividing text into sentences, and word vectors for numerically representing word meanings. These annotations encompass tokenization for segmenting text into individual units, part-of-speech tagging for labeling each token's grammatical category, dependency parsing to uncover syntactic relationships between words, and named entity recognition for identifying entities such as location and temperature.

VI. Results and Analysis

This work uses the Keras library and the Tensorflow backend to train the Custom spaCy model. We used an NVIDIA Tesla Titan X GPU to train and assess the models on the dataset.

For training we took 70% of the data, 15% for testing and 15% for evaluation, we utilized **LOSS_TOK2VEC** which pertains to the loss function associated with the token-to-vector (TOK2VEC) component of the model, and **LOSS NER**, which is the loss function associated with the named entity

[1]https://catalog.data.gov/dataset/?tags=iot

		Loss Tok2Vec	Loss NER	F1 score	ENT_P	ENT_R	Accuracy
With Data Preparation	Training	0.00	0.00	100.00	100.00	100.00	1.00
	Validation	7.85	8.76	100.00	100.00	100.00	1.00
	Benchmark	21.86	36.45	**94.35**	**96.15**	**92.59**	**0.96**
Without Data Preparation	Training	9.78	46.77	99.87	99.82	99.63	0.99
	Validation	87.31	110.79	92.56	92.42	92.71	0.92
	Benchmark	214.90	612.09	0.73	0.64	0.7	0.72

TABLE I: Results of training, evaluation, and benchmarking of the model using prepared and unprepared data.

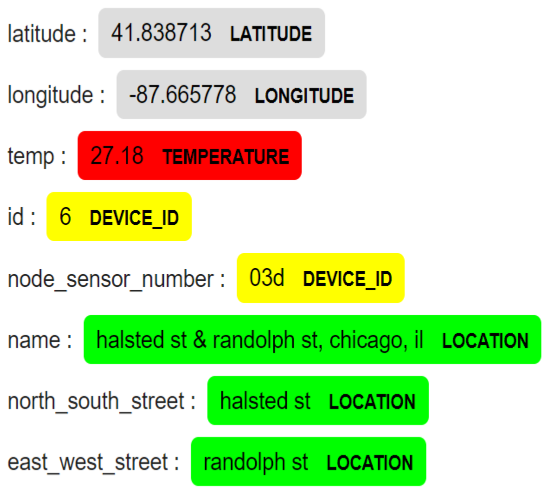

Fig. 2: Training data sample with annotations.

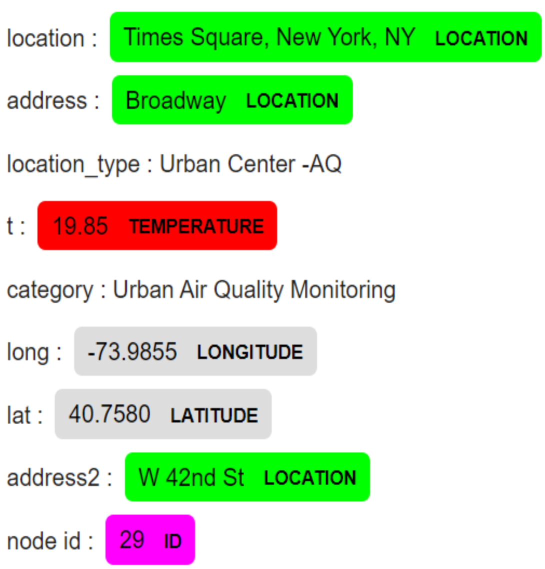

Fig. 3: Benchmark data sample with different keys and values unseen in training.

recognition (NER) component of the model. The F1 score, the harmonic mean of precision and recall, serving as a measure of the model's accuracy in identifying named entities. The precision score, measuring the ratio of true positive predictions to the total number of positive predictions made by the model. The recall score, measuring the ratio of true positive predictions to the total number of true entities present in the data. Lastly, The score which is the accuracy of the model.

The table 1 presented in this analysis provides a comprehensive examination of a model's performance across various phases, with and without data preparation, in the context of named entity recognition (NER). In the scenario with data preparation, the model exhibited exceptional performance during the training phase, achieving a perfect fit to the data with Tok2Vec loss and NER loss registering at 0.00. Furthermore, it attained perfect scores for F1, entity precision, recall, and overall accuracy, indicating robust performance across different evaluation metrics. This consistency was maintained during the validation phase, reaffirming the model's ability to generalize well to unseen data. Even in the benchmark phase, where randomly generated data was introduced, the model showcased commendable performance with high F1 scores, precision, recall, and accuracy, underscoring its capacity for effective generalization. Conversely, in the scenario without data preparation, the model's performance was notably lower across all phases, with higher losses and slightly diminished scores for evaluation metrics. Particularly, in the benchmark phase, the model exhibited significant degradation in performance, highlighting the critical role of data preparation in enhancing the model's ability to generalize to unseen datasets. This analysis underscores the importance of meticulous data preparation in NER tasks, as it substantially influences the model's overall performance and generalization capabilities. Additionally, it's worth noting that the model's accuracy decreased when focusing too much on the structural aspects and the opening and closing of each entity, indicating the need for further refinement in its approach.

Analyzing Figure 3 in detail, it becomes apparent that despite the JSON file containing a variety of keys and values, the model consistently identifies the entities it was trained to detect. This resilience is particularly notable when the order of the data is shuffled. The model's ability to maintain accuracy across different permutations underscores its robustness and adaptability to diverse input structures. This suggests a high degree of generalization in its learning, enabling it to effectively process varied data formats and arrangements without compromising performance.

VII. Conclusion

In conclusion, our exploratory study shows that AI improves semantic interoperability in the IoT ecosystem. The technique relies on Named Entity Recognition (NER) models to detect and classify textual data from heterogeneous IoT devices with different data representations. The framework uses generic data processing middleware to normalize heterogeneous IoT data. This middleware integrates data into a uniform NLP training pipeline for tokenization, stopword removal, sentence splitting, POS tagging, and spaCy-powered NER. This careful approach ensures that the NLP Translator accurately translates refined and organized data. The translated entities are stored in NoSQL storage like MongoDB, demonstrating the system's scalability and data retrieval. Our NER model, trained with excellent tokenization and entity extraction precision, scored well in key performance criteria. Despite the complexity of validation across varied datasets, the model has shown precision and adaptation, suggesting its potential for wider IoT use. The architecture supports current IoT infrastructures and Industry 4.0 principles, demonstrating how AI can transform semantic interoperability and a smarter, more connected environment.

ACKNOWLEDGMENT

This work has been supported by the EIPHI Graduate School (contract "ANR-17-EURE-0002").
Computations have been performed on the supercomputer facilities of the "Mesocentre de calcul de Franche-Comté"

REFERENCES

[1] E. Siow, T. Tiropanis, and W. Hall, "Analytics for the internet of things: A survey," *ACM computing surveys (CSUR)*, vol. 51, no. 4, pp. 1–36, 2018.

[2] F. Burzlaff, N. Wilken, C. Bartelt, and H. Stuckenschmidt, "Semantic interoperability methods for smart service systems: A survey," *IEEE Transactions on engineering management*, vol. 69, no. 6, pp. 4052–4066, 2019.

[3] J. Azar, T. Khoury, A. Makhoul, and R. Couturier, "Home automation system with iot stack and chatgpt for people with reduced mobility," in *2023 IEEE 4th International Multidisciplinary Conference on Engineering Technology (IMCET)*. IEEE, 2023, pp. 44–49.

[4] H. Rahman and M. I. Hussain, "A comprehensive survey on semantic interoperability for internet of things: State-of-the-art and research challenges," *Transactions on Emerging Telecommunications Technologies*, vol. 31, no. 12, p. e3902, 2020.

[5] E. Alreshidi, "Smart sustainable agriculture (ssa) solution underpinned by internet of things (iot) and artificial intelligence (ai)," *arXiv preprint arXiv:1906.03106*, 2019.

[6] J. Nilsson and F. Sandin, "Semantic interoperability in industry 4.0: Survey of recent developments and outlook," in *2018 IEEE 16th international conference on industrial informatics (INDIN)*. IEEE, 2018, pp. 127–132.

[7] M. Kelly, J. Greenberg, C. B. Rauch, S. Grabus, J. P. Boone, J. A. Kunze, and P. M. Logan, "A computational approach to historical ontologies," in *2020 IEEE International Conference on Big Data (Big Data)*. IEEE, 2020, pp. 1878–1883.

[11] A. Santo, A. Santos, and H. S. Mamede, "Integrated Language Translation IoT Devices : A Systematic Literature Review," in *2022 17th Iberian Conference on Information Systems and Technologies (CISTI)*. IEEE, jun 22 2022.

[8] Y. Yi, G. Gong, and K. Mandal, "Implementation of three lwc schemes in the wifi 4-way handshake with software defined radio," *arXiv preprint arXiv:1909.11707*, 2019.

[9] Q. Bai, S. Alsudais, C. Li, and S. Zhao, "Maliva: Using machine learning to rewrite visualization queries under time constraints," *arXiv preprint arXiv:2112.00182*, 2021.

[10] J. Nilsson, F. Sandin, and J. Delsing, "Interoperability and machine-to-machine translation model with mappings to machine learning tasks," in *2019 IEEE 17th International Conference on Industrial Informatics (INDIN)*, vol. 1. IEEE, 2019, pp. 284–289.

[12] E. Lakka, N. E. Petroulakis, G. Hatzivasilis, O. Soultatos, M. Michalodimitrakis, U. Rak, K. Waledzik, D. Anicic, and V. Kulkarni, "End-to-End Semantic Interoperability Mechanisms for IoT," in *2019 IEEE 24th International Workshop on Computer Aided Modeling and Design of Communication Links and Networks (CAMAD)*. IEEE, 9 2019.

[13] A. Venceslau, R. Andrade, V. Vidal, T. Nogueira, and V. Pequeno, "Iot Semantic Interoperability: A Systematic Mapping Study," in *Proceedings of the 21st International Conference on Enterprise Information Systems*. SCITEPRESS - Science and Technology Publications, 2019.

[14] Antonios Pliatsios and C. Goumopoulos, "A Review on IoT Frameworks Supporting Multi-Level Interoperability – The Semantic Social Network of Things Framework," 2020.

[15] S. Balakrishna, V. K. Solanki, V. K. Gunjan, and M. Thirumaran, *A Survey on Semantic Approaches for IoT Data Integration in Smart Cities*. Springer Singapore, jun 28 2019, pp. 827–835.

[16] O. Novo and M. D. Francesco, "Semantic Interoperability in the IoT," *ACM Transactions on Internet of Things*, vol. 1, no. 1, pp. 1–25, feb 29 2020.

[17] Z. Gui and A. Harth, "Towards a Data Driven Natural Language Interface for Industrial IoT Use Cases," in *2021 IEEE 2nd International Conference on Human-Machine Systems (ICHMS)*. IEEE, sep 8 2021.

[18] Y. Tian, S. Khanna, and A. Pljonkin, "Research on Machine Translation of Deep Neural Network Learning Model Based on Ontology," *Informatica*, vol. 45, no. 5, jul 20 2021.

[19] S. Sharma, M. Diwakar, P. Singh, A. Tripathi, C. Arya, and S. Singh, "A Review of Neural Machine Translation based on Deep learning techniques," in *2021 IEEE 8th Uttar Pradesh Section International Conference on Electrical, Electronics and Computer Engineering (UPCON)*. IEEE, nov 11 2021.

[20] L. H. B. Nguyen, V. H. Pham, and D. Dinh, "Improving Neural Machine Translation with AMR Semantic Graphs," *Mathematical Problems in Engineering*, vol. 2021, pp. 1–12, jul 8 2021.

[21] R. Rapp, "Using Semantic Role Labeling to Improve Neural Machine Translation," *International Conference on Language Resources and Evaluation*, 2022.

[22] Elior Sulem, Omri Abend, and A. Rappoport, "Semantic Structural Decomposition for Neural Machine Translation," *STARSEM*, 2020.

[23] Diego Moussallem, Mihael Arcan, A. N. Ngomo, and P. Buitelaar, "Augmenting Neural Machine Translation with Knowledge Graphs," *arXiv.org*, 2019.

[24] Y. Yin, Y. Li, F. Meng, J. Zhou, and Y. Zhang, "Categorizing Semantic Representations for Neural Machine Translation," 2022.

[25] J. Li, X. Wang, B. Yang, S. Shi, M. R. Lyu, and Z. Tu, "Neuron Interaction Based Representation Composition for Neural Machine Translation," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 34, no. 05, pp. 8204–8211, apr 3 2020.

[26] M. Honnibal and I. Montani, "spaCy 2: Natural language understanding with Bloom embeddings, convolutional neural networks and incremental parsing," 2017, to appear.