Semantic-enhanced Digital Twin for Industrial Working Environments

Chao Yang[1][0000-0003-1902-1461], Qize Guo[2][0000-0003-3515-9727], Hao Yu[2][0000-0003-3522-3426], Yan Chen[2][0000-0002-3872-0586], Tarik Taleb[3][0000-0003-1119-1239], and Kari Tammi[1][0000-0001-9376-2386]

[1] Mechanical Engineering, Aalto University, 02150 Espoo, Finland
[2] ICTficial Oy, 02130 Espoo, Finland
[3] Electrical Engineering and Information Technology, Ruhr University Bochum, 44801 Bochum, Germany
chao.1.yang@aalto.fi

Abstract. Real-time data from diverse Internet of Things (IoT) sensors (such as cameras, temperature, light, and air quality sensors) is essential for monitoring smart manufacturing environments. However, efficiently perceiving, integrating, and interpreting this data remains a challenge, as it involves dealing with heterogeneous data formats, ensuring data accuracy, and providing real-time analytics. This paper proposes a semantic-enhanced Digital Twin (DT) to address these complexities and aims to offer a comprehensive view of industrial working environments. The paper first presents a conceptual overview of the semantic-enhanced DT architecture, followed by a detailed description of the system architecture, encompassing edge, cloud, and interface modules. Additionally, the implementation of the entire system is presented. The results demonstrate the feasibility of the proposed DT, showing its potential for deployment in real-world scenarios.

Keywords: Digital Twin · Industrial working environment · Internet of Things · Semantic model

1 Introduction

Industry 4.0 is driving a significant digital transformation in traditional manufacturing, ushering in a new era of intelligent, flexible, and adaptable production paradigms [1]. This shift necessitates efficient and comprehensive environmental data perception [2]. Accurate real-time environmental data is crucial for monitoring and optimizing industrial processes. Manufacturers can adapt swiftly to changes by gaining real-time insights into environmental conditions, ensuring optimal performance while minimizing environmental impact. This leads to improved resource utilization, reduced waste, and increased overall efficiency.

While the Internet of Things (IoT) has revolutionized data collection in factories, enabling real-time monitoring and control for optimized production [3], a critical gap remains. Current limitations include the lack of an efficient approach for capturing and translating real-world factory environmental data into a unified digital format. This challenge is further complicated by the inherent heterogeneity of environmental data at the factory floor level [4]. Sensor readings and environmental conditions are often measured in diverse formats and units, hindering seamless integration and analysis within a digital system. Additionally, inconsistent data quality and latency issues can result in inaccurate models and suboptimal decision-making [5]. Fragmented data silos across different devices and systems also impede the holistic view required for comprehensive analysis. Therefore, smart factories need an efficient approach to data perception and transmission, as well as a standardized method to represent and interpret this rich environmental data.

Digital Twins (DTs) can be a promising solution to address these issues by providing a cohesive, real-time digital representation of the physical environment. This allows for continuous monitoring and analysis of the factory floor, with real-time updates of environmental data. However, existing DTs often struggle with semantic interoperability [6], limiting their ability to fully leverage the rich data available. Diverse data sources and varying terminologies can lead to misunderstandings and misalignments in data interpretation. Semantic technologies establish a uniform interpretation of data, ensuring universal comprehension regardless of its source [7]. Integrating these technologies with DTs enables seamless communication and data exchange across different systems and platforms [8]. Semantic-enhanced DTs can achieve a higher level of interoperability, ensuring that data from various sources is accurately interpreted and applied.

This work proposes a novel semantic-enhanced Digital Twin (DT) for a holistic digital representation of industrial working environments. To facilitate knowledge representation and reasoning about edge devices and their perception, an ontological model is introduced. The proposed DT architecture consists of three key components: the edge module, the central module, and the interface module. The edge module is responsible for real-time environmental data perception and transmission. The central module utilizes a microservices-based architecture to enable scalable service deployment and distributed computing resources. The interface module provides a comprehensive suite of functionalities for end users. Additionally, this paper details the hardware design of the edge device and the microservices-based architecture within the central server. Finally, the system's performance is evaluated through practical validation, focusing on information retrieval capabilities, User Interface (UI) usability, and latency performance.

The paper is structured as follows: Section 2 provides an overview of relevant research on IoT, DTs, and semantic technologies. Section 3 introduces an ontological model and details the system architecture of the proposed DT. Section 4 covers the hardware design of the edge device, the information model it utilizes, and the microservices-based architecture employed by the central server platform. Section 5 presents the evaluation and results. Section 6 analyzes the research contributions and limitations of the proposed system. Finally, the paper concludes with a summary of the study and outlines potential avenues for future research.

2 Related Works

2.1 IoT-based Industrial Systems

The IoT with prevalent sensing capabilities in manufacturing has transformed physical entities and operators into 'cyber-ones' [9]. By leveraging IoT technologies, industrial systems enable the monitoring and optimization of various aspects of industrial processes. Through the integration of sensors, data acquisition modules, and analytical algorithms, these systems continuously collect and analyze data from critical points within industrial operations. For instance, Hossain et al. [10] presented a novel algorithm embedded in the proposed PI-based controller for real-time fault detection in power converters. In addition to monitoring machinery and equipment for operational efficiency and downtime reduction, scholars have also focused on ensuring sustainability and safety in industrial operations. For example, Palazon et al. [11] employed Wireless Sensor Networks (WSN) with mobile motes carried by workers and vehicles within the smart factory to enhance mutual perception, thereby improving health and safety conditions in the industrial environment. Moreover, Ahn et al. [12] presented an intelligent camera-based system for managing the safety of factory environments. The integration of multiple IoT devices has become indispensable for comprehensive perception and understanding within the factory field. However, the diverse data formats and heterogeneous nature of IoT devices often hinder the seamless integration and interpretation of information.

2.2 Semantic Interoperability in the IoT

To address the fragmentation of IoT ecosystems, the World Wide Web Consortium (W3C) introduced the Web of Things (WoT) architecture [13], a vendor-neutral framework enabling interoperability among diverse IoT devices. Each IoT device is required to publish a Thing Description (TD), a metadata document that formally defines the device's capabilities, properties, and interactions. Ontologies serve as formal knowledge representation models, utilizing classes, properties, and relationships to organize and describe domain-specific concepts [14]. Within TDs, ontologies are employed to provide a shared semantic model, allowing devices to consistently understand and communicate their data and functionalities. The heterogeneity of data in IoT has led to the development of multiple ontologies aimed at addressing interoperability issues among sensor data. The Semantic Sensor Network (SSN) [15] ontology published as a W3C Recommendation. The Sensor, Observation, Sample, and Actuator (SOSA) [16] ontology is a lightweight version of SSN. The Smart Applications REFerence (SAREF) [17] ontology specifies concepts in the smart appliances domain. However, this approach merely shifts the interoperability challenge to a higher level, as the lack of interoperability persists when IoT components rely on different ontologies [18]. This research aims to develop a system-level semantic layer to facilitate complex data integration, semantic search, and reasoning across heterogeneous IoT datasets.

2.3 Digital Twin and Semantic Technologies

Digital Twins (DTs) are receiving considerable focus in the industrial field due to their ability to create a virtual representation of physical assets throughout their lifecycle. This virtual counterpart facilitates advanced decision-making through data analysis, simulation, and machine learning techniques [19]. DTs achieve this by continuously integrating data from various sensors and systems, providing a comprehensive and up-to-date view of industrial operations. This holistic approach enables seamless data integration and analysis, allowing for swift adaptation to dynamic industrial environments. Furthermore, DTs leverage their accurate reflection of the physical environment to model complex interactions and predict potential outcomes, significantly enhancing decision-making processes within the industrial domain [20]. Consequently, DTs bridge the gap between fragmented, siloed solutions and the need for a cohesive and adaptive system. This adaptability ensures the system's continuous evolution alongside the industrial landscape, ultimately promoting sustained efficiency and reliability in managing industrial processes and assets. Traditionally, a DT framework consists of three primary layers: the physical layer, the information layer, and the virtual layer [21, 20]. The information layer in a DT typically handles data perception, processing, and analysis. This layer is responsible for collecting data from multiple sources, preparing it for analysis, and applying analytical techniques to extract valuable insights. However, while crucial, this layer often faces challenges in ensuring that data from diverse sources is consistently integrated and interpreted correctly, especially when dealing with heterogeneous data types and formats. A semantic layer, powered by ontologies, offers a standardized and structured approach to representing and integrating data from different sources. Ontologies ensure that data elements are understood and interpreted uniformly, leading to more reliable data analysis. Integrating DTs with semantic technologies enables effective management of heterogeneous data and the seamless merging of fragmented data silos. Therefore, this paper proposes a DT framework that leverages IoT and semantic technologies to accurately mirror industrial environments.

3 Proposed Digital Twin Framework for Industrial Working Environment

To provide a digital representation of industrial working environments, this study proposed a three-layer DT architecture, illustrated in Fig. 1. Each layer facilitates bidirectional data transfer. Within the physical layer, multiple edge devices are deployed in the factory to perceive and transmit environmental data. These edge devices serve as the foundational sensory network that captures real-time data from the physical environment. The information layer, utilizing the microservice architecture, is responsible for data processing and knowledge integration. By leveraging microservices, information processing is decomposed into smaller, independent services that manage specific data processing tasks. This approach

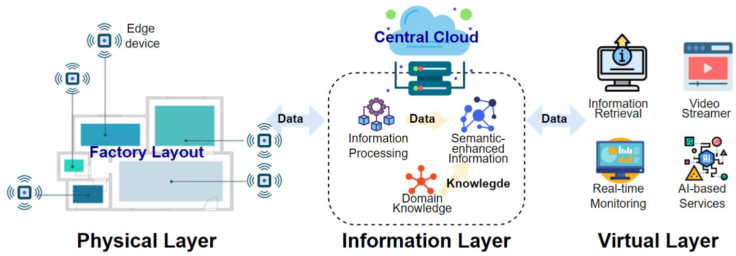

Fig. 1. The overview of the proposed Digital Twin for industrial working environments.

facilitates easier maintenance, deployment, and scaling of individual functionalities, which is advantageous in a dynamic industrial environment. Furthermore, this study introduces domain knowledge to describe the perception of edge devices, aiming to integrate heterogeneous data sources collected by these devices at the factory level and ensure harmonization and comprehensibility for further analysis. Through the fusion of semantic-enhanced data with streaming data, the system enables accurate interpretation of complex industrial environments and facilitates the cohesive integration of multiple edge devices to reflect the entire industrial working environment. In the virtual layer, end users can interact with digital services to gain a comprehensive understanding of the working environment. This interaction provides users with insightful visualizations and analytics that reflect the real-time state of the factory, enabling proactive management and optimization of industrial processes. The subsequent sections introduce the proposed domain knowledge model and system architecture of the DT.

3.1 Domain Knowledge Model within the Information Layer

This paper proposes a standardized knowledge representation using ontological models to facilitate the integration of heterogeneous environmental data from various edge devices. This approach consolidates all observed data, offering a holistic description of the working environment. Fig. 2 details a conceptual ontological model (prefix: *isien*) designed to represent the edge device's perception in the manufacturing domain. The model leverages established vocabularies and domain knowledge by reusing existing ontologies, including SOSA [16], QUDT [22], GeoSPARQL [23], InPro [14], and BOT [24]. This approach avoids overlapping and redundant modeling and fosters reliability of developing model [25].

The concepts and relationships from SOSA ontology (prefix: *sosa*) are employed to represent sensor-related information for the edge device. A *sosa:Sensor* produces a corresponding *sosa:Observation*, which acts as a procedure to calculate a *sosa:Result* for a specific *sosa:ObservedProperty* of a *FeatureOfInterest*. The *qudt:QuantityValue* class, which is part of the QUDT ontology and prefixed as *qudt*, is a subclass of *sosa:Result* that is used to express values with particular units. In GeoSPARQL, the *geo:Geometry* class, which is prefixed as *geo*, is a subclass of *sosa:Result* and is specifically utilized for representing spatial data. The

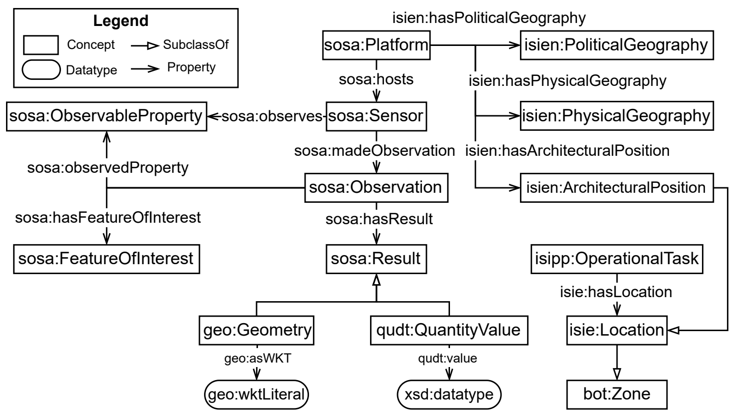

Fig. 2. A conceptual ontology for describing edge devices and their observations of industrial working environments.

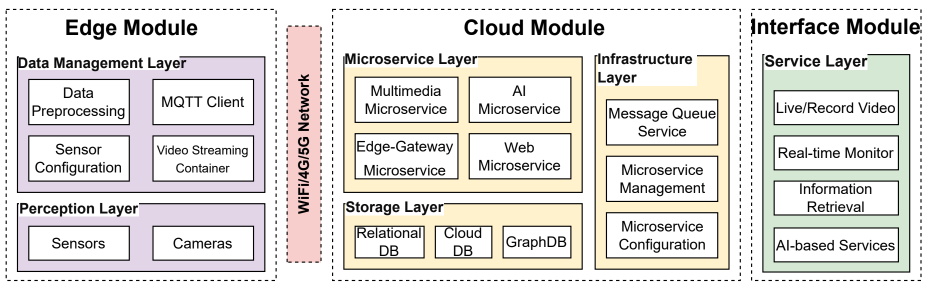

Fig. 3. The system architecture of the proposed Digital Twin.

isien:PoliticalGeography class contains information about the political aspects, such as the street, city, and country, related to the edge device. The *Physical-Geography* class represents the edge device's precise geographical position (e.g., GPS coordinates). The *ArchitecturalPosition* class, a subclass of *isie:Location* from InPro, describes the placement of the edge device within the building's topology. The *isie:Location* class is a subclass of the *bot:Zone* within BOT. The *isipp:OperationalTask* in InPro is connected to the *isie:Location* associated with each production process. Moreover, the proposed ontological model can function as the environmental module for InPro, facilitating the incorporation of environmental information and industrial production processes.

3.2 System Architecture of the Proposed Digital Twin

The system architecture of the proposed DT is illustrated in Fig. 3, consisting of three fundamental components: the edge module, the cloud module, and the interface module. The edge module serves as the system's frontline, integrating various IoT sensors for data perception and transmission. This module is structured into two layers: perception and data management. The perception layer acts as an interface between the physical environment and the digital domain, managing hardware components such as sensors and cameras. It is primarily responsible for data acquisition, enabling the retrieval of sensor data and video streams essential for monitoring diverse aspects of industrial operations. The data management layer focuses on processing and organizing collected data. Its key functionalities include data preprocessing (e.g., filtering and cleaning) and data publishing for seamless transmission of processed data to the cloud for further analysis or storage. Additionally, it facilitates sensor configuration, allowing users to customize sensor settings based on specific monitoring needs. Finally, this layer features a video stream container to support video feed transmission to multimedia servers.

To complement the limited computation and storage capacities of the edge module, the cloud module functions as a centralized hub for data processing, storage, and analysis. The cloud server architecture is structured into three distinct layers: the microservice layer, the storage layer, and the infrastructure layer, each addressing specific aspects of microservices-based functionality and data management. The microservice layer contains a set of specialized microservices. The Multimedia Service facilitates live video streaming and video-on-demand functionalities, while the Artificial Intelligence (AI) Service employs machine learning algorithms to analyze data and derive actionable insights. The Edge-Gateway Service serves as the communication interface between the edge module and the cloud module. The Web Service provides web-based visual and interactive UIs for direct user interaction. The storage layer supports diverse database technologies tailored to various data types and usage scenarios. For instance, sensory data is stored in a structured format within a relational database, while cloud databases manage large volumes of video stream data. Graph databases are used for the integration of streaming data and ontological models. Within the cloud module, the infrastructure layer supports the deployment and operation of microservices. The Message Queue Service establishes a real-time data pipeline, facilitating seamless communication and data exchange among different microservices, thereby ensuring efficient coordination and synchronization of distributed processing tasks. Additionally, microservice management and configuration streamline the centralized administration and orchestration of microservices across the entire cloud infrastructure.

The interface module provides a range of services for end users, encompassing live video monitoring, recorded video retrieval, real-time monitoring, information retrieval, AI-enhanced decision-making, and edge device configuration. These functionalities allow stakeholders to oversee the working environment of a factory remotely. Through the interface module, stakeholders obtain insights into the current state of the factory floor during the production processes. Furthermore, this layer facilitates human supervisory control, empowering stakeholders to remotely adjust parameters, initiate workflows, and respond promptly to

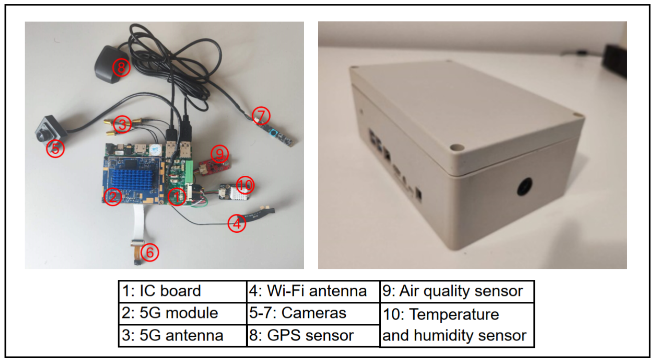

Fig. 4. The design of the edge device: left is the hardware layout of the edge device, and right is the final assembled device.

alerts or emergencies. This capability enhances operational efficiency, ensures adherence to safety regulations, and promotes proactive maintenance strategies.

4 System Implementation

4.1 Hardware Design of the Edge Device

This study developed a mobile edge device called the Digital Twin Box (DTB) for capturing environmental data, shown in Fig. 4. The DTB features an RK3399 System-on-Chip (SoC) housed on an IC (Integrated Circuit) board, operating on Ubuntu 20.04, supporting edge computing tasks. An industrial-grade Quectel RG500U IoT module integrated with the IC board enables 5G Non-Standalone (NSA) and Standalone (SA) operation, ensuring backward compatibility with 4G/3G networks. For comprehensive visual monitoring, the DTB incorporates three high-definition, wide-angle cameras capturing video streams from various perspectives. Additionally, a GPS sensor provides physical location data, while temperature, humidity, and air quality sensors collect environmental data. The compact design ($20cm$ x $15cm$ x $8cm$) allows for easy deployment on the factory floor, as depicted in Fig. 4 (right side).

4.2 The Knowledge Model of the Edge Device (DTB)

The classes and relationships that describe the DTB (prefix: *dtb*) are illustrated in Fig. 5, based on the proposed domain knowledge model. A *DTB* is a subclass of the *sosa:Platform*, comprising of various sensors, such as *dtb:TemperatureHumiditySensor*, *dtb:AirQualitySensor*, *dtb:GPS*, and *dtb:Camera*. The *dtb:TemperatureHumiditySensor* consists of two distinct observations, *dtb:TempertureObservation* and *dtb:HumidityObservation*. These observations target the *quantitykind:Temperature* and *quantitykind:RelativeHumidity* of the *isie:Location*. The results, *dtb:TemperatureResult* and *dtb:HumidityResult*, are subclasses of *qudt:QuantityValue*, representing the numerical values with

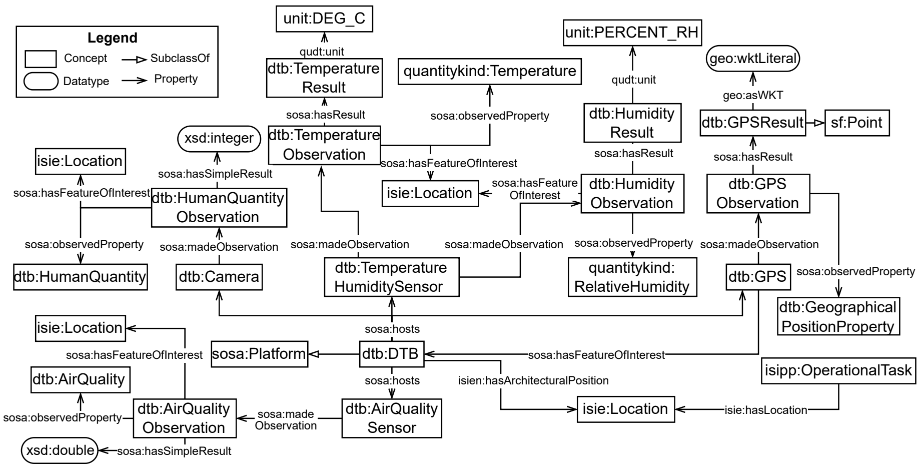

Fig. 5. The classes and relationships of the domain ontology pertaining to the designed edge device.

units specified by *unit:DEG_ C* and *unit:PERCENT_ RH* from the QUDT ontology. Similarly, the *dtb:AirQualityObservation*is designed to measure a numerical value representing the air quality (*dtb:AirQuality*) of the corresponding *isie:Location*. Additionally, the geographical position of the DTB (*dtb:DTB*) is described using the *sf:Point* class from GeoSPARQL, employing the wkTLiteral notation for spatial representation. Finally, leveraging the *isie:Location* class, observations from the DTB can be associated with a specific production process within the industrial environment. Building upon the well-defined semantic structure provided by the ontology model, an instance model is subsequently generated. This instance model facilitates the incorporation of real-time streaming data from the DTBs. The data is then stored in a graph database of the central server, enabling the creation of a comprehensive knowledge graph that furnishes a detailed description of industrial working environments.

4.3 The Microservice-based System Architecture of the Cloud Server

Fig. 6 depicts the microservices-based system implementation within the cloud server. Various technologies are utilized to construct the cloud server. The microservice layer leverages Docker for containerization to ensure microservice portability and scalability. ZLMediaKit is utilized to provide live video streaming services. The Vue.js development framework is employed within the Web Service for the development of dynamic and responsive UI. The EMQX software is utilized to realize the Edge Gateway Service for the MQTT broker. Notably, this work integrates YOLOv8 [26] for object recognition within the AI Service. In the microservice infrastructure layer, NACOS is implemented for microservice configuration and registration discovery, while Kubernetes manages

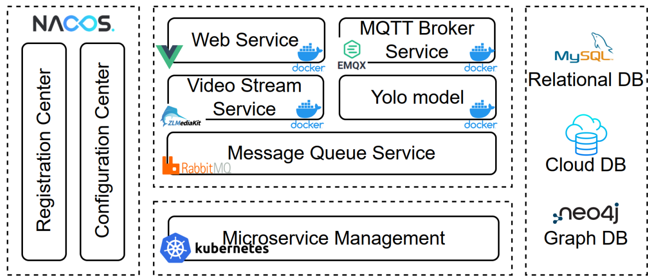

Fig. 6. The microservices-based system architecture in the cloud server.

container orchestration for microservices. Furthermore, RabbitMQ is employed as the Message Queue Service to enable asynchronous communication among various microservices within the system. The storage layer utilizes MySQL to store logical relational data, while native cloud storage solutions are leveraged for storing video data. This configuration facilitates swift retrieval and simultaneous storage/playback of extensive video datasets. Neo4j is used as a graph database to implement the instance model of the proposed ontology in the form of knowledge graphs.

5 Validation

This study implemented three DTBs within a laboratory-scale manufacturing environment at the Aalto Industrial Internet Campus (AIIC) [27]. Cloud services were provided by a German cloud infrastructure provider. The effectiveness of the proposed DT architecture was evaluated through three key aspects: information retrieval capability, UI usability, and latency performance.

5.1 Information Retrieval

In this section, we explore the utility of the proposed DT for governing heterogeneous data sources and fostering sensor interoperability to establish a high-fidelity replica of a physical environment. To assess the effectiveness of this approach, a series of evaluation questions are presented in Table 2. These questions are designed to assess the semantic DT's capacity to retrieve accurate and relevant information, thereby validating its performance within real-world applications. Three DTBs are denoted as 'DTB_1', 'DTB_2', and 'DTB_3'. Each DTB is assigned to a specific zone within the AIIC: 'DTB_1' is assigned to 'Zone_A', 'DTB_2' to 'Zone_B', and 'DTB_3' to 'Zone_C'. Table 1 shows five Cypher queries corresponding to the questions outlined in Table 2. The first query addressed Question 1 by retrieving the number of DTBs within the AIIC. Queries 2 and 3 responded to Questions 2 and 3, which focus on providing the quantitative analysis of sensor data, including the average temperature and maximum humidity. Query 4, aligned with Question 4, identified zones where there are two or more humans. Finally, Query 5 answered Question 5 by calculating the distance between two DTBs using Cypher's spatial functions.

Table 1. Cypher statements for the validation questions.

```
#Query 1
MATCH (:Space{name:"AIIC"}) - [:containsZone] → (loc:Location),
(dtb:DTB) - [:hasArchitecturalPosition] → (loc) RETURN count(dtb)
#Query 2
MATCH (:Space{name:"AIIC"}) - [:containsZone] → (loc:Location),
(tem:TemperatureObservation) - [:hasFeatureOfInterest] → (loc),
(tem) - [:hasResult] → (res:TemperatureResult) RETURN AVG(res.value)
#Query 3
MATCH (h:HumidityObservation) - [:hasFeatureOfInterest] →
(loc:Location), (h) - [:hasResult] → (res:HumidityResult)
RETURN loc.name AS location, res.value AS humidity
ORDER BY res.value DESC LIMIT 1
#Query 4
MATCH (:Space{name:"AIIC"}) - [:containsZone] → (loc:Location),
(hq:HumanQuantityObservation) - [:hasFeatureOfInterest] → (loc)
WHERE hq.hasSimpleResult ≥ 2
RETURN loc.name AS location, hq.hasSimpleResult AS value
#Query 5
MATCH (:DTB {name:"DTB_1"}) - [:hosts] → (:GPS) - [:madeObservation]
→ (:GPSObservation) - [:hasResult] → (res1:Point),
(:DTB {name:"DTB_2"}) - [:hosts] → (:GPS) - [:madeObservation]
→ (:GPSObservation) - [:hasResult] → (res2:Point)
RETURN point.distance(res1.value, res2.value) AS distance
```

Table 2. Specified questions and answers for the information retrieval case.

Specified questions	**Answers**
1. How many DTBs are in 'AIIC'?	3
2. What is the average temperature of the 'AIIC'?	20.05
3. Which location has the maximum humidity?	Zone_B, 56.00%
4. In which AIIC's location are there two people or more?	Zone_A, 2
5. What is the distance between 'DTB_1' and 'DTB_2'?	24.20

5.2 User Interface

This work realizes the interface module as a web UI for end users, providing a comprehensive suite of functionalities, as depicted in Fig. 7. Live video streaming enables the retrieval of real-time video feeds for remote monitoring, along

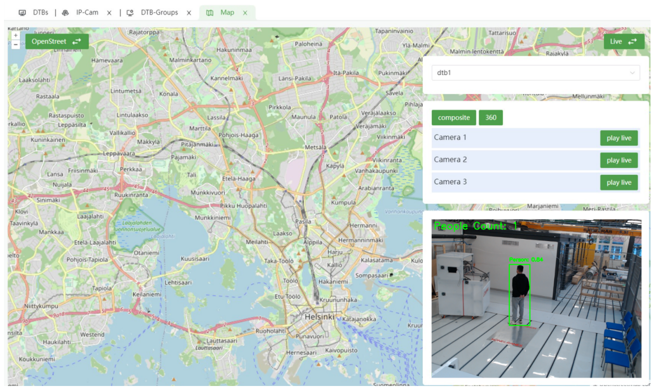

Fig. 7. The screenshot of the UI of the proposed DT.

Table 3. The latency measurement under WiFi and 4G networks.

Network	**Condition**	**Mean** (ms)	**Range**(ms)	**SD**
WiFi	Edge-UI	285.451	113.390	0.021
	Edge-Central	69.166	49.780	0.005
4G	Edge-UI	413.318	650.510	0.071
	Edge-Central	117.263	228.110	0.040

with AI-driven insights. As illustrated in Fig. 7, YOLOv8 is employed to perform real-time people counting through video stream analysis. The GPS sensor embedded into the system allows for viewing the edge device's geographical locations on a map. The video recording feature enables the review of recorded videos. The monitoring service encompasses both real-time and historical data, providing users with a holistic view of the system's operational status. Finally, the management features include device pairing, switching, and sharing, which streamline user control over device setups and operational parameters.

5.3 Latency Measurement

To assess the system's performance under varying network connectivity, latency measurements were conducted in three distinct network environments: WiFi and 4G. Within each network, a dataset of 1000 samples was collected. Each sample consisted of four timestamps capturing: the timestamp when the edge device publishes the message, the timestamp when the central platform receives the message, the timestamp when the central platform sends the message, and the timestamp when the UI receives the message. It's important to note that the two network environments (WiFi and 4G) solely influence the communication between the edge device and the central platform. The communication between the central platform and the UI likely occurs on a separate, potentially local network, and therefore remains independent of the chosen network type. The statistical outcomes of these measurements, including Mean, Range, and Standard Deviation (SD), are presented in Table 3 for each network condition. The table differentiates between "Edge-UI" and "Edge-Central" latency. "Edge-UI" represents the total latency experienced between the edge device and the UI, while "Edge-Central" refers solely to the latency between the edge device and the central platform. The average "Edge-UI" latency measured under WiFi is $285.451ms$, while under 4G it increases to $413.318ms$. Similarly, the "Edge-Central" latency follows the same trend. WiFi demonstrates a lower latency of $69.166ms$ compared to 4G's $117.263ms$. Moreover, the average latency for central platform handling was measured to be $11.259ms$, while the average latency between the central platform and the UI is $281.143ms$. In summary, the total latency of our platform is lower than $500ms$ for sensor data transmission for the edge device to the end UI.

6 Discussion and Limitation

This research aims to provide an efficient approach to perceiving, integrating, and interpreting industrial environmental data, ensuring an adaptive, interoperable, and cohesive digital representation. To achieve this objective, we introduce a semantic-enhanced DT that accurately reflects the complexities of industrial working environments. The main contributions of this work are: 1) Integrating DT with semantic technologies to enhance semantic interoperability; 2) Introducing a unified and formalized ontological model to describe edge devices and their observations of industrial working environments; 3) Applying an edge-cloud system framework for the proposed DT, where the cloud platform leverages a microservices architecture.

First, we integrate DT with a semantic layer that leverages semantic technologies, including ontologies and knowledge graphs (within the graph database Neo4j). The Thing Description (TD) provided by W3C primarily focuses on defining the semantics of individual entities (such as devices and sensors) in IoT environments, offering a framework for describing their properties and behaviors. In our work, we extend beyond individual entities by utilizing knowledge graphs to capture not only their relationships and dependencies but also the interconnected nature of the entire system. This allows for more complex domain knowledge representation and facilitates advanced reasoning and inference capabilities. In the context of DTs, knowledge graphs provide several key benefits. They allow for a unified representation of the properties of individual assets, while also capturing their relationships, dependencies, and interactions with other system components. By using an ontology-based structure, knowledge graphs enhance data integration, making it easier to incorporate and represent data from various sources. Additionally, knowledge graphs enable advanced analytics techniques, such as graph-based analysis, reasoning, and inference, which are invaluable for gaining deeper insights into complex systems and supporting informed decision-making.

Additionally, this paper proposes an ontological model designed to describe edge devices and their perception within the industrial domain. Leveraging established ontologies such as SOSA and QUDT, the model comprehensively represents sensor platforms, their observations, and data structures. It enables the semantic description of various sensors, including temperature, humidity, air quality sensors, GPS devices, and cameras. The model also integrates GeoSPARQL and supports the incorporation of AI models with cameras, enabling capabilities such as human quantity detection. Through integration with InPro, the system establishes connections between sensor data and specific production processes within industrial facilities. Moreover, the semantic model enables seamless connectivity of edge devices at the factory field level, providing a holistic view of the industrial environment. The proposed ontological model facilitates streamlined environmental data management and enhances interoperability across the industrial IoT landscape.

Lastly, this study proposes an edge-cloud system architecture designed to address the limitations of edge devices regarding computational and storage resources. The architecture utilizes a microservices framework within the cloud platform. By decomposing complex applications into four loosely coupled services, the framework promotes flexibility, scalability, and maintainability. The modular nature of microservices enhances fault isolation and resilience, ensuring robust service operation and continuity even during system failures or environmental changes. The Information Layer's microservices offer standardized third-party API interfaces. Stakeholders can customize system logic by generating an application token and accessing relevant functions in the virtual layer. This setup allows large-scale systems to interact directly with the virtual layer through API calls, tailoring functionality to specific needs. Thus, the architecture facilitates efficient utilization of edge resources while leveraging the scalability and flexibility of cloud computing paradigms in industrial IoT applications.

While our research has demonstrated the feasibility of the proposed approach, several limitations need to be addressed. One of the most critical concerns in any IoT environment is data security and privacy, particularly given the system's role in collecting data from industrial settings. The data collected often includes sensitive and proprietary information, making it essential to implement robust privacy safeguards to prevent unauthorized access and ensure compliance with regulations such as the General Data Protection Regulation (GDPR). To enhance data security, one potential solution is to enforce data sovereignty by separating data providers from data service providers. This separation can help minimize risks associated with data breaches and unauthorized access. Additionally, for video streaming, it is imperative to implement de-identification techniques to protect the privacy of human operators. These measures are crucial for maintaining the confidentiality of personal information and upholding privacy standards within industrial IoT applications.

Another limitation pertains to the ontological model developed for our proposed DTB. While the model is tailored to our specific implementation, integrating it into larger, more complex systems of systems requires ontology alignment to facilitate seamless information integration. This process involves harmonizing different ontological frameworks to ensure coherent data exchange and interoperability across diverse systems. Moreover, for systems that lack a semantic representation, additional middleware is necessary to bridge the gap between non-semantic and semantic systems. The need for such middleware introduces complexity and potential performance overhead, which must be carefully managed to prevent degrading system efficiency and responsiveness.

7 Conclusion and Future Works

The industrial working environment is becoming increasingly complex. Developing a Digital Twin (DT) of this environment is crucial for smart factories to adequately assess environmental impacts and make more informed decisions. However, existing DT encounters several limitations, such as 1) the lack of an approach to digital representation of industrial working environments, and 2) the heterogeneity of environmental data within the factory field level. Therefore, this study introduces a semantic-enhanced DT tailored for industrial working environments. The system architecture comprises three main modules: edge module, cloud module, and interface module. The edge module supports various environmental data perceptions and transmissions from the field level. The cloud module adopts a microservices approach to ensure scalability and facilitate seamless maintenance and updates. The interface module offers a range of functionalities for end-users, empowering them to effectively monitor and track the working environment. Moreover, the integration of the semantic model enables semantic interoperability and interconnectivity among all edge devices. Validation of the system has demonstrated its feasibility in terms of information retrieval, user interface usability, and performance.

To enhance the capabilities of the proposed platform, future works would contain: 1) Integration of information retrieval capabilities with the web-based UI to provide more intuitive and user-friendly operations and information visualization for end-users. 2) Support for industrial protocols like OPC Unified Architecture (OPC UA) within the system architecture to enhance interoperability and compatibility with industrial equipment and networks. 3) Validate the system's performance in a more complicated industrial working environment.

8 Acknowledgement

This work is partially supported by Business Finland for the co-innovation project through the Necoverse project (Grant 10768/31/2022), the European Union's Horizon Europe program for Research and Innovation through the aerOS project under (Grant No. 101069732), the 6G-SANDBOX project (Grant No. 101096328), and the RIGOUROUS project (Grant No. 101095933). The paper reflects only the authors' views. The Commission is not responsible for any use that may be made of the information it contains.

References

1. A. Ustundag, E. Cevikcan, Industry 4.0: managing the digital transformation, Springer, 2017.
2. M. Soori, B. Arezoo, R. Dastres, Internet of things for smart factories in industry 4.0, a review, Internet of Things and Cyber-Physical Systems (2023).
3. S. Vaidya, P. Ambad, S. Bhosle, Industry 4.0–a glimpse, Procedia manufacturing 20 (2018) 233–238.
4. J. Wang, C. Xu, J. Zhang, R. Zhong, Big data analytics for intelligent manufacturing systems: A review, Journal of Manufacturing Systems 62 (2022) 738–752.
5. H. Y. Teh, A. W. Kempa-Liehr, K. I.-K. Wang, Sensor data quality: A systematic review, Journal of Big Data 7 (1) (2020) 11.
6. X. Zheng, J. Lu, D. Kiritsis, The emergence of cognitive digital twin: vision, challenges and opportunities, International Journal of Production Research 60 (24) (2022) 7610–7632.
7. C. Fürber, C. Fürber, Semantic technologies, Springer, 2016.
8. S. Muralidharan, B. Yoo, H. Ko, Designing a semantic digital twin model for iot, in: 2020 IEEE International Conference on Consumer Electronics (ICCE), IEEE, 2020, pp. 1–2.
9. C. Yang, S. Lan, L. Wang, W. Shen, G. G. Huang, Big data driven edge-cloud collaboration architecture for cloud manufacturing: a software defined perspective, IEEE access 8 (2020) 45938–45950.
10. M. L. Hossain, A. Abu-Siada, S. Muyeen, M. M. Hasan, M. M. Rahman, Industrial iot based condition monitoring for wind energy conversion system, CSEE Journal of Power and Energy Systems 7 (3) (2020) 654–664.
11. J. A. Palazon, J. Gozalvez, J. L. Maestre, J. R. Gisbert, Wireless solutions for improving health and safety working conditions in industrial environments, in: 2013 IEEE 15th International Conference on e-Health Networking, Applications and Services (Healthcom 2013), IEEE, 2013, pp. 544–548.
12. J. Ahn, J. Park, S. S. Lee, K.-H. Lee, H. Do, J. Ko, Safefac: Video-based smart safety monitoring for preventing industrial work accidents, Expert Systems with Applications 215 (2023) 119397.
13. World Wide Web Consortium (W3C), Web of Things (WoT) Architecture, retrieved August 30, 2024 (2024).
URL https://w3c.github.io/wot-architecture/
14. C. Yang, Y. Zheng, X. Tu, R. Ala-Laurinaho, J. Autiosalo, O. Seppänen, K. Tammi, Ontology-based knowledge representation of industrial production workflow, Advanced Engineering Informatics 58 (2023) 102185.
15. M. Compton, P. Barnaghi, L. Bermudez, R. Garcia-Castro, O. Corcho, S. Cox, J. Graybeal, M. Hauswirth, C. Henson, A. Herzog, et al., The ssn ontology of the w3c semantic sensor network incubator group, Journal of Web Semantics 17 (2012) 25–32.
16. K. Janowicz, A. Haller, S. J. Cox, D. Le Phuoc, M. Lefrançois, Sosa: A lightweight ontology for sensors, observations, samples, and actuators, Journal of Web Semantics 56 (2019) 1–10.
17. L. Daniele, F. den Hartog, J. Roes, Created in close interaction with the industry: the smart appliances reference (saref) ontology, in: Formal Ontologies Meet Industry: 7th International Workshop, FOMI 2015, Berlin, Germany, August 5, 2015, Proceedings 7, Springer, 2015, pp. 100–112.

18. O. Novo, M. D. Francesco, Semantic interoperability in the iot: Extending the web of things architecture, ACM Transactions on Internet of Things 1 (1) (2020) 1–25.
19. M. J. Kaur, V. P. Mishra, P. Maheshwari, The convergence of digital twin, iot, and machine learning: transforming data into action, Digital twin technologies and smart cities (2020) 3–17.
20. E. VanDerHorn, S. Mahadevan, Digital twin: Generalization, characterization and implementation, Decision support systems 145 (2021) 113524.
21. Y. Zheng, S. Yang, H. Cheng, An application framework of digital twin and its case study, Journal of Ambient Intelligence and Humanized Computing 10 (2019) 1141–1153.
22. H. Rijgersberg, M. Van Assem, J. Top, Ontology of units of measure and related concepts, Semantic Web 4 (1) (2013) 3–13.
23. R. Battle, D. Kolas, Enabling the geospatial semantic web with parliament and geosparql, Semantic Web 3 (4) (2012) 355–370.
24. M. H. Rasmussen, M. Lefrançois, G. F. Schneider, P. Pauwels, Bot: The building topology ontology of the w3c linked building data group, Semantic Web 12 (1) (2021) 143–161.
25. C. Blanco, J. Lasheras, E. Fernández-Medina, R. Valencia-García, A. Toval, Basis for an integrated security ontology according to a systematic review of existing proposals, Computer Standards & Interfaces 33 (4) (2011) 372–388.
26. G. Jocher, A. Chaurasia, J. Qiu, Yolo by ultralytics, https://github.com/ultralytics/ultralytics, accessed: 2024-06-16 (2023).
27. C. Yang, X. Tu, J. Autiosalo, R. Ala-Laurinaho, J. Mattila, P. Salminen, K. Tammi, Extended reality application framework for a digital-twin-based smart crane, Applied Sciences 12 (12) (2022) 6030.