From Raw Data to Smart Manufacturing:

AI and Semantic Web of Things for Industry 4.0

Pankesh Patel
Fraunhofer USA

Muhammad Intizar Ali
National University of Ireland

Amit Sheth
Kno.e.sis Center

AI techniques combined with recent advancements in the Internet of Things, Web of Things, and Semantic Web—jointly referred to as the Semantic Web—promise to play an important role in Industry 4.0. As part of this vision, the authors present a Semantic Web of Things for Industry 4.0 (SWeTI) platform. Through realistic use case scenarios, they showcase how SweTI technologies can address Industry 4.0's challenges, facilitate cross-sector and cross-domain integration of systems, and develop intelligent and smart services for smart manufacturing.

Industry 4.0 refers to the Fourth Industrial Revolution—the recent trend of automation and data exchange in manufacturing technologies. To fully realize the Industry 4.0 vision, manufacturers need to unlock capabilities such as vertical integration through connected and smart manufacturing assets of a factory, horizontal integration through connected discrete operational systems of a factory, and end-to-end integration throughout the entire supply chain. Several architectures and conceptual platforms (for example, RAMI 4.0 in Europe, IIRA in the US) have been proposed to develop Industry 4.0 applications.[1] However, these reference architectures are largely missing the granularity of Semantic Web and AI technologies. We believe that combined AI and Semantic Web technologies are a good fit for the plethora of complex problems related to interoperability, automated, flexible, and self-configurable systems such as Industry 4.0 systems.[2] In this article, we present a Semantic Web of Things for Industry 4.0 (SWeTI) platform for building Industry 4.0 applications.

We first present a representative set of Industry 4.0 use cases, followed by a layered SWeTI platform for building these use cases. Each layer contains a variety of tools and techniques to build smart applications that can process raw sensory information and support smart manufacturing using Semantic Web and AI techniques. Finally, we present our existing tools and middleware for realizing the SWeTI platform.

Published by the IEEE Computer Society
1541-1672/18/$33.00 ©2018 IEEE

USE CASES

Below is a representative set of Industry 4.0 use cases that can potentially leverage Semantic Web technologies.

Use Case 1: Deep Integration

In this use case, an integrated and holistic view of a factory is established to improve decision-making across different departments and to reduce overall complexity. This includes the inter-linking of diverse data sources such as sensor measurements (for example, temperature, vibration, pressure, power), the manufacturing execution system (for example, work orders, material needed for the production, incoming material number), business processes, work force, and so on. Although much of this data is already captured by IT systems, it remains largely inaccessible in an integrated way without investing significant manual effort. Thus, the objective of this use case is to make all data available in a unified model to support users (factory planners, machinists, controllers, field technician, and so forth) in decision-making. For example, consider the following scenarios:

- A factory planner needs input from diverse sources regarding order plans, machine maintenance schedules, workforce availability, and so on.[3]
- A field technician must quickly troubleshoot an onsite industrial asset, and is seeking a solution that combines a summary of the problem, including difficulty and time estimates; links to relevant manuals and necessary parts; additional physical tools to resolve the problems and the current location of these tools; and, if the problem is difficult to resolve, additional support from people with the necessary expertise.
- During production, a machinist needs to know which tools are required to perform the task at hand, the location of these tools and materials, and quality control standards to be adhered to.[3]

Use Case 2: Horizontal Integration

This use case extends the vertical integration of all factory operation into the horizontal dimension, knitting together the relevant players in a manufacturing supply chain—the raw materials and parts suppliers, logistics, inventory of supplied goods, production process, warehouses and distributors of finished products, sales and marketing, customers—through an interconnected networks of Internet of Things (IoT) devices and external information sources such as social media and web services (for example, financial services or weather forecasting), overseen via an overarching semantic-enabled engine. Some examples include the following:

- A smart factory manager wants to optimize supply chain and warehouse facilities to ensure that the right amount of raw material is always available in the warehouse to support production processes. Because the factory produces customized products, allowing customers to choose their food products' ingredients, it is difficult to estimate the amount and type of raw material required to fulfill customer orders. However, efficient data-mining and machine-learning techniques that harness social media data can provide insights into ongoing trends and customer preferences, which will help the warehouse manager optimize the supply chain and ensure that the right amount of raw ingredients is always available in the warehouse to replenish the processing machine at the food production factory.
- A production manager in a manufacturing unit needs an integrated view of the supply chain, including raw materials and the distribution network, to optimize internal manufacturing processes. A horizontal integration of smart factory production processes, supply, and the distribution network including fleet management data and external datasets such as traffic congestion, weather, and social media is required to build an optimal strategy for the production of perishable food products. Using this integrated information, the production manager can adapt internal business processes on the fly.

Use Case 3: Autonomous System

This class of use cases deals with enabling factory devices to cooperate to achieve the factory's overall objectives. The following are just a few examples:

- **Self-organization.** When a production order comes down to the factory, machines can communicate and exchange information with on another to organize resources to complete the orders on time. Resource allocation is determined at runtime, rather than pre-allocated, depending on the machines' current conditions, including current factory workload at factory, machine availability and maintenance schedule time, backlog of customer orders, and machine capacity. Moreover, resource allocation could consider external electricity rates data to achieve the goal of reducing the factor's energy consumption and carbon footprint.
- **Flexible manufacturing.** Decentralized control is very useful when market demands lead to the introduction of new machines in factories. The new machines can participate by simply announcing their services and features during the resource allocation process. This illustrates the flexibility and adaptability of a factory, where new machines can be integrated in a plug-and-produce fashion according to market demands with minimal downtime.
- **Fault tolerance.** The highly dynamic and self-organization features result in a fault-tolerant system—faulty sensors can be replaced by discovering new sensors with similar functionality to prevent downtime during the production process.

A SEMANTIC WEB OF THINGS FOR INDUSTRY 4.0 PLATFORM

In this section, we present our SWeTI platform's layered architecture for Industry 4.0 application design. Figure 1 depicts the overall architecture beginning with the data processing pipeline at the device-, sensor-, or machine-level and moving toward intelligent autonomous applications or dashboards at the application layer. In what follows, we briefly describe each layer and its components and functionality.

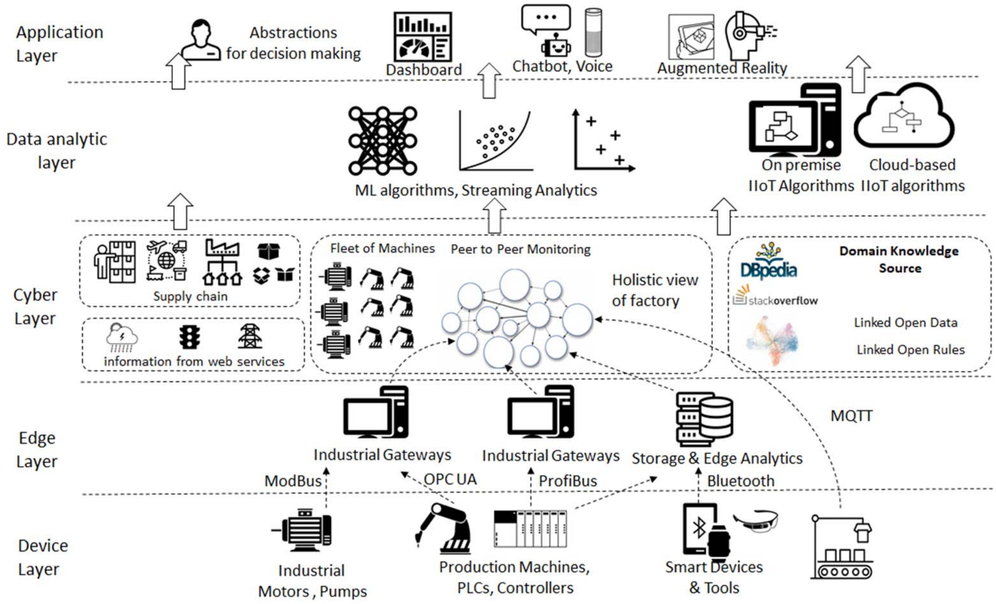

Figure 1. A layered view of the Semantic Web of Things for Industry 4.0 (SWeTI) platform.

Device Layer

At the factory floor, devices ranges from production machines (such as PLCs, industrial motors, pumps, and robots) to smart devices and tools (such as smartwatches, glasses, sensors, and smartphones) that provide extended human–machine interface and functionality. From a connectivity viewpoint, these could be legacy devices (which communicate through legacy protocols such as profibus, modbus, opc) as well as new equipment with embedded technology that allows these devices to communicate through recent IoT standards such as OPC-UA, MQTT, Bluetooth, and so on.

Edge Layer

The edge layer transforms data generated by factory floor device into information. Typically, industrial gateways are deployed to this layer, which are relatively powerful devices compared to the lowest layer. The broad functionality at this layer is presented as follows:

- **Interoperability.** For instance, OPC-UA specifies real-time communication of plant data between control devices from different vendors (such as ABB or Siemens). Production Performance Management Protocol (PPMP; https://bit.ly/2P0WGUc) describes the payload definition of three types of messages generated from industrial assets: machine messages, measurement messages, process messages. This protocol is independent of transport protocols such as HTTP/REST, MQTT, and AMQP.
- **Analytics-based actions.** Actions include reactions to various production events such as executing rule-based alerts and/or sending commands back (for example, reconfiguration of industrial machines) to the production equipment and tools.
- **Edge analytics.** These provide data aggregation techniques, data filtering, and cleansing techniques to refine data, implementing a device-independent model as a base for decisions, and analytical logic for specific device domains and types.

Greengrass (https://aws.amazon.com/greengrass) is an edge-analytic software solution from Amazon. Similarly, Microsoft offers Azure IoT Edge. Greengrass has a small footprint that can run on gateway devices such as BeagleBone and Raspberry PI. Using Lambda functions, AWS Greengrass provides data filtering and computation capabilities. Developers can push small analytical capabilities by deploying lambda functions from AWS cloud (for further details, see "On Using the Intelligent Edge for IoT Analytics"[4]).

Cyber Layer

The cyber layer acts as a distributed information hub. Having massive information gathered from diverse distributed information sources, this layer prepares ground for the data-analytic layer for specific analytics. Key forms these information sources take are enterprise wide knowledge: Industry 4.0 envisions the access of data across different players (logistics, customers, distributors, and supplier) in the supply chain. Information from various machines on factory floors (through the edge layer or directly from the device layer) is pushed to form the linked network of information (that is, Linked Data; https://bit.ly/29YZz5b) generated from production machines. Linked data is a natural fit because it provides an abstraction layer on top of a distributed set of data, stored across the supply chain.

An alternate technology solution could be blockchain (https://bit.ly/2nna7Ac). It is decentralized and distributed across peer-to-peer networks. Each participant in the network can read and write data in blockchain. The blockchain network remains in sync. For instance, blockchain can be used for timely industrial asset maintenance, sharing necessary information across different organizations.[5] Repair partners could monitor the blockchain for maintenance to minimize downtime in factory production and record their work on the blockchain for further use. The regulators of an industrial plant equipment would have access to asset records, allowing them to provide timely certification to ensure that the asset is safe for factory workers.

Domain knowledge sources include the Linked Open Data (LOD), Linked Open Reasoning (LOR), and Linked Open Services (LOS). These approaches reuse data, rules, and services, respectively, from the web.[6] For instance, Sensor-based Linked Open Rules (S-LOR)[6] is a data set of interoperable rules used to interpret data produced by sensors. LOS is intended to share and reuse services and applications that can be used to compose crossdomain Industry 4.0 applications.

External sources are also available. Real-time data streams from external sources such as social media and tracking devices from logistics, weather, traffic, and news feeds can be brought into the cyber layer. This data can be linked with other information (for example, events affecting supply shipments, possibly with the support of information extraction techniques such as for event identification) at the data-analytic layer for strategic optimization, such as route network improvements. This would provide a fully transparent view of the supply chain and let managers make decisions to keep the supply chain flow moving.

An alternative to the distributed information hub is to use the cloud to build Industry 4.0 applications. Cloud-based manufacturing is a centralized single-shop place that allows manufacturers to apply industrial analytics on top of stored data. For example, Microsoft Azure (https://bit.ly/2KLKbs9) use the term “data lake” to store big data. A cloud-based centralized repository allows users to store structured and unstructured data at any scale. Users can run different analytics—from simple analytics such as data visualization to complex analytics such as real-time analytics, machine learning, and big data processing.

Additional analytics on top of the data sources described above enables several capabilities, described in the next section.

Data-Analytic Layer

The massive amount of data available on the cyber layer creates a distributed data lake, which creates an opportunity to add industrial analytics on top of this data lake by leveraging AI algorithms. The purpose of industrial analytics is to identify invisible relationships among data at the application layer (discussed in the next section), thus enabling decision-makers to make optimized decisions.

The industrial analytics algorithms can be on-premise (written by an enterprise) or cloud-based. Various cloud vendors allow manufacturers to store and integrate data, apply industrial analytics available in a cloud marketplace, and develop customized business solutions leveraging cloud-based tools. As an example, GE developed Predix (www.predix.io/catalog), an industrial Internet platform that offers a marketplace to deploy various apps and services, including predictive maintenance, anomaly detection, and algorithms for the intelligent edge. Similarly, Azure AI gallery (https://gallery.azure.ai) offers a catalog service that finds different machine learning algorithms for the industrial Internet. Microsoft has integrated AzureML (https://studio.azureml.net), a visual programming environment, to develop machine-learning algorithms. Therefore, AzureML developers can make their algorithms available to the community by publishing an Azure AI gallery. Recently, Siemens launched MindSphere (https://siemens.mindsphere.io; hosted on AWS), a cloud-based Industry 4.0 OS that lets manufacturers connect their industrial machines to the cloud and offers a marketplace (like an AppStore) to use deployment-ready Industrial applications.

Application Layer

The application layer presents the acquired knowledge from the cyber layer to users (such as domain experts and decision-makers) so that correct decisions can be taken. This layer is about building meaningful and customized applications on top of services and data exposed by the data-analytic layer and presenting the acquired knowledge to the user in an appropriate manner. A broad variety of machine learning approaches[7] have been developed to analyze and extract higher-level information from diverse IoT data.

In recent years, industrial vendors have demonstrated a wide variety of Industry 4.0 applications, combining a range of technologies such as digital twin, chatbot/natural language processing, and augmented reality. For instance, by combining the functionality exposed by the data-analytic layer with an industrial asset, developers can create a *digital twin*—a virtual representation of an industrial asset. An extension of this work is to integrate different interfaces with the digital twin. GE Digital has demonstrated an extended demo of a digital twin (https://youtu.be/2dCz3oL2rTw); customers can ask the digital twin questions about its performance and potential issues in natural language and receive answers back in natural language. Moreover, customers can interact with digital twin through augmented reality (AR) devices such as the Microsoft HoloLens and obtain a 3D view of an asset (for example, a steam turbine) to analyze its internal parts.

The Industry 4.0 application-building approaches are divided into two broad categories:

- **Rapid application development (RAD).** To address the longer development time problem, RAD tools provide abstractions that hide low-level programming details. For example, Node-RED (https://nodered.org) offers node to read data from an OPC-UA server. Similarly, Kura Wires (www.eclipse.org/kura) offers visual programming constructs to acquire data using industrial protocols such as Modbus and OPC-UA.
- **Cloud-based tools.** These tools offer services that implement common application development functions, such as connecting industrial assets to the cloud, data storage, and data visualization. Developers can configure these tools and develop applications very rapidly. Various vendors have started offering cloud-based tools for building Industry 4.0 applications. For instance, Amazon Lex (https://aws.amazon.com/lex) offers a service that develops conversational interfaces into any application using text and voice. Microsoft Azure Accelerator (https://bit.ly/2P1qEan) offers preconfigured and customizable solutions. Developers can deploy these solutions in minutes, connect factory floor assets, and deploy Industry 4.0 solutions such as remote monitoring, connected factory, and predictive maintenance.

SWETI PLATFORM COMPONENTS

We continue to leverage our existing tools and middleware to build Industry 4.0 applications. We plan to use some open source tools for Industry 4.0 (https://bit.ly/2gIbVow) from Eclipse Foundation. Here, we present our existing open source tools for building Industry 4.0 applications.

IoTSuite

IoTSuite (https://bit.ly/2M9FldX) is a framework for prototyping IoT applications by making application development easier by hiding development-related complexity from developers. It takes platform-independent high-level specifications as inputs, parses them, and generates platform-specific code that results in a distributed software system collaboratively hosted by heterogeneous IoT devices. The high-level specification includes specification about sensors, actuators, storage devices, and computational component specification as well as deployment specification that describe device properties. Thus, developers need not concern themselves with the platform- and runtime-specific aspects of development. The key characteristics of IoTSuite that make it suitable for our Industry 4.0 projects are its flexibility:

- It can generate code for different target programming languages (for example, C, C++, Python, Java) as the code generator for target programming language is exposed as a plug-in. To generate a new programming language, IoTSuite developers simply write a plug-in to generate code in a target programming language.
- It can plug different runtime systems (for example, MQTT, CoAP, OPC-UA) as it exposes well-defined interfaces to plug different runtime systems. IoTSuite developers simply have to implement runtime specific interfaces to plug a target runtime system.

This framework has been used for industrial-grade devices such as ABB's RIO 600 as well as popular devices such as Raspberry PI, Arduino, and Android devices. The framework's current version integrates standards such as OPC-UA, MQTT, CoAP, and WebSocket, and generates code in programming languages such as Node.js, Android, and Java.

Machine-to-Machine Measurement

The Machine-to-Machine Measurement (M3) framework (https://bit.ly/2OWMWdk) is intended to help build crossdomain IoT applications. It uses semantic technologies to achieve interoperability among heterogeneous IoT systems. Reasoning over semantically annotated data produced by IoT devices can help create user suggestions. For example, a body temperature value of 38 degrees could be associated with a naturopathy application that would suggest home remedies when a constant high fever was sensed. This framework uses LOD, LOR, Linked Open Vocabularies (LOV), and LOS to enhance interoperability and get meaningful knowledge from data.[6]

ACEIS (https://bit.ly/2OYQtYA) middleware contains a set of tools designed for IoT data analytics and uses Semantic Web technologies to build various components including automated streaming data discovery, integration on the fly, event detection, and contextually aware decision support systems.[8] This middleware can be used to build smart applications for Industry 4.0.

CONCLUSION

Industry 4.0 is an emerging area of research. In this article, we described the example of the SWeTI platform, which augments Semantic Web, AI, and data analytics to support the building of smart IoT applications for Industry 4.0. We presented a set of realistic use case scenarios to advocate for the SWeTI platform.

ACKNOWLEDGMENTS

We acknowledge the fruitful contributions of the SWeTI workshop (https://swetiworkshop.wordpress.com) participants and contributors. A productive discussion concluded by the workshop inspired some of the ideas presented in this article. This work is partially funded by SFI under grant no. SFI/16/RC/3918 and SFI/12/RC/2289.

REFERENCES

1. I. Grangel-González et al., "The Industry 4.0 Standards Landscape from a Semantic Integration Perspective," *Proc. 22nd IEEE Int'l Conf. Emerging Technologies and Factory Automation* (ETFA), 2017, pp. 1–8.
2. A. Sheth, "Internet of Things to Smart IoT Through Semantic, Cognitive, and Perceptual Computing," *IEEE Intelligent Systems*, vol. 31, no. 2, 2016, pp. 108–112.
3. N. Petersen et al., "Monitoring and Automating Factories Using Semantic Models," *Proc. 6th Joint Int'l Conf. Semantic Technology* (JIST), 2016, pp. 100–115.
4. P. Patel, M.I. Ali, and A. Sheth, "On Using the Intelligent Edge for IoT Analytics," *IEEE Intelligent Systems*, vol. 32, no. 5, 2017, pp. 64–69.
5. D. Miller, "Blockchain and the Internet of Things in the Industrial Sector," *IT Professional*, vol. 20, no. 3, 2018, pp. 15–18.
6. A. Gyrard et al., "Building the Web of Knowledge with Smart IoT Applications," *IEEE Intelligent Systems*, vol. 31, no. 5, 2016, pp. 83–88.
7. M.S. Mahdavinejad et al., "Machine Learning for Internet of Things Data Analysis: A Survey," *Digital Communications and Networks*, vol. 4, no. 3, 2017, pp. 161–175.
8. F. Gao et al., "Automated Discovery and Integration of Semantic Urban Data Streams: The ACEIS Middleware," *Future Generation Computer Systems*, vol. 76, 2017.

ABOUT THE AUTHORS

Pankesh Patel is a senior research scientist at Fraunhofer USA's Center for Experimental Software Engineering (CESE). His current focus is implementation of Industry 4.0 techniques and methodologies in commercial environments. Contact him at dr.pankesh.patel@gmail.com.

Muhammad Intizar Ali is an adjunct lecturer, research fellow, and research unit leader of the Reasoning, Querying, and IoT Data Analytics Unit at the Insight Centre for Data Analytics, National University of Ireland. His research interests include Semantic Web, Internet of Things, and data analytics. Contact him at ali.intizar@insight-centre.org.

Amit Sheth is the LexisNexis Ohio Eminent Scholar and executive director of the Ohio Center of Excellence in Knowledge-Enabled Computing (Kno.e.sis), and a Fellow of IEEE and AAAI. Contact him at amit@knoesis.org.