Proceedings of the Thirty-Second International Joint Conference on Artificial Intelligence (IJCAI-23)
Demonstrations Track

SiWare: Contextual Understanding of Industrial Data for Situational Awareness

Anuradha Bhamidipaty, Elham Khabiri, Bhavna Agrawal, Yingjie Li
IBM Research
{anubham, ekhabiri, bhavna, yingjie}@us.ibm.com

Abstract

SiWare is an AI-powered knowledge discovery system, that helps unlock new insights and accelerates data-driven decisions with contextualized industrial data. SiWare links and fuses heterogeneous data sources with an industry semantic model leveraging multiple AI capabilities to provide system-wide visibility into operational characteristics. As part of this demo paper, we describe the requirements for such a system, along with its deployment aspects, and demonstrate the benefits in two industrial scenarios.

1 Introduction

Most industrial data, as much as 95%, is not connected and used [Bane, December 2020]. A 2022 study by the Manufacturing Institute indicates that 78% of companies are concerned about impending aging workforce and lack of knowledge transfer is estimated to cost large businesses $47 million per year due to time waste, missed opportunities, and delayed projects [Institute, July 2019]. This presents an opportunity for systematic approaches to organize, link, and contextualize the diverse and siloed industrial datasets into a knowledge representation that can support improved downstream analytics. Given a dataset of a particular modality, for example, textual log files, or manuals, there are techniques to create industry specific word embeddings and use such fine-tuned models for tasks such as classification [Khabiri *et al.*, 2019]. Similarly, given tabular datasets, there are AI approaches described to classify and link values [Sankhe *et al.*, 2021] based on the context embedding that surrounds a cell in a tabular setting. For timeseries datasets, there are multiple approaches to generate valuable insights such as data quality anomalies [Zerveas *et al.*, 2021; Shrivastava *et al.*, 2020; Patel *et al.*, 2022]. The primary focus so far has been either on analyzing datasets for one modality (text, tabular, time series, etc.) and generating insights, or on task-specific learning from multiple modalities [Erickson *et al.*, 2022]. Using this demonstration paper, we aim to address the gap of a systematic approach to leverage multiple such AI capabilities which may be data-modality-dependent and yet bring them together to create a unified representation of the multi-modal data sources thus enabling situational awareness.

1.1 Scenario 1: Semiconductor Manufacturing

Microelectronic manufacturing and development is a complex process involving data distributed across multiple data sources, structured and unstructured, including master databases, operational applications (manufacturing execution systems, scheduling applications), process trace data repository, maintenance reports, tool manuals, ad-hoc documentation of historically interesting events. Current practices rely on expert knowledge to extract, harmonize, aggregate and analyze data, and is done independently and repeatedly for each use case. In many cases, the subject matter expert (SME) has to work with data scientist to capture their domain expertise into the end-to-end analysis cycle. There is a need to organize data into semantically meaningful entities and relationships and enable retrieval of semantically related data without detailed knowledge of schemas or query languages. It is also imperative that an SME is able to interact with such a system to not only consume it but also to organize data leveraging their domain knowledge. Such semantic representation can enable high-value use cases like root cause diagnosis. For example, search for “metal liner deposition” returns data, documents referring to (1) “metal liner deposition” (2) liner deposition tool (3) graphs with electrical measurements influenced by liner deposition (4) others

It is important to note that semantic representation of the data is not reliant on the existence of an ontology or an industry standard but instead derived “bottoms-up” from the operational data that reflects the dynamic state of the environment.

1.2 Scenario 2: Continuous Flow Manufacturing

An operations center monitoring a plant floor can receive 1000s of alarms every day. Responding to each alarm is a critical but challenging task as it involves manual iterative steps of identifying relevant data including historical reports, adjacent related equipment in P&IDs, IoT sensor readings, analyzing trends and patterns, and performing root cause analysis by understanding the state of the related systems. There is a high cost of misreading an alert or inaction. There is a need to contextualize these diverse data sources in an (semi)automated fashion into a unified knowledge representation (such as a knowledge graph) that provides proactive recommendations and 24/7 guidance to operators.

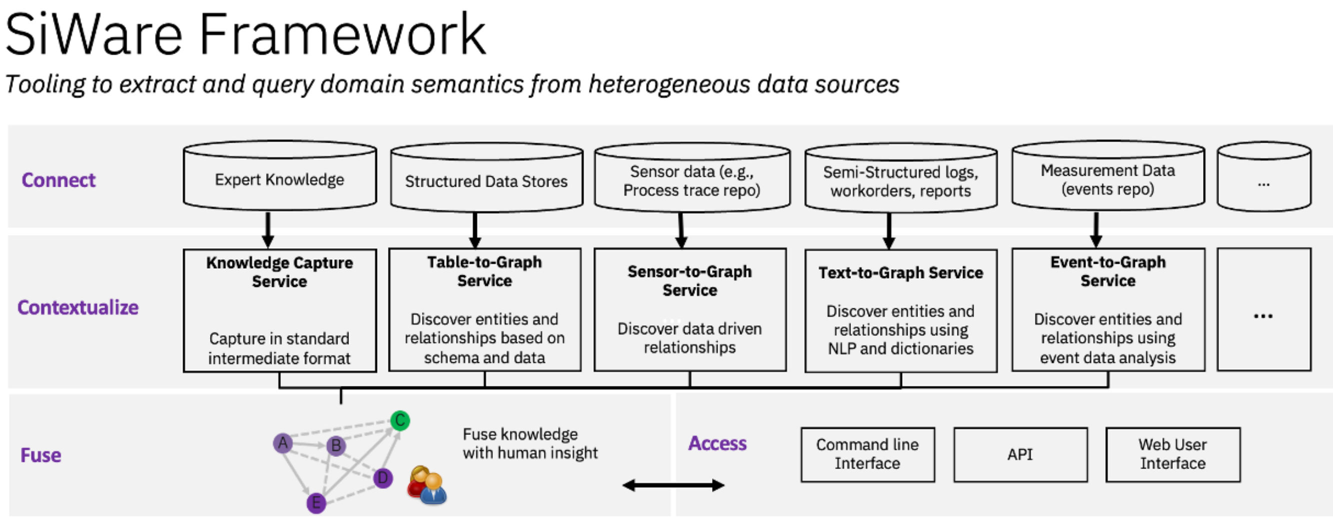

Figure 1: Basic components of SiWare including domain-independent but data-modality-dependent contextualization flows and a fusion service.

2 SiWare Overview

There are 4 key components of a SiWare flow (Figure 1):

- **Connect**: User can connect to the heterogeneous data sources in their environment by staging the data in Object Storage like IBM COS[1] or local file system and using the `\connect` API.
- **Contextualize**: User can initiate one or more contextualization flows by first invoking the `\contextualize\initialize` API to organize the data into a *Collection* which acts as a logical grouping of data that is needed for knowledge discovery. Each contextualization flow is aimed at extracting the knowledge - entities and relationships - from the collection and creates a graph representation. Examples of such flows include named entity recognition (NER) in a text data collection (`\contextualize\ner` API), or identifying relevant sensor correlations from a collection of sensor data (`\contextualize\sensor` API) informed by P&ID diagrams. The contextualization flows trigger AI models which run asynchronously returning a job id, the status of which can be queried using the `\status` API. The completed flow may result in purging of the staged data based on the configuration. Each contextualization flow has a standard interface with inputs being a collection and a configuration file and outputs a $.json$ artifact of entity-relationship mapping.
- **Fuse**: The user can fuse the output of one or more contextualization flows into an industrial knowledge graph using the `\fuse` API. The fusion service is responsible for reconciling the entities and relationships using either given business logic or inferred matching criteria [Al-Moslmi *et al.*, 2020]
- **Access**: The user can query the unified knowledge graph using the `\access` API. Predefined templates for industrial use cases allow the user to retrieve sub-graphs for root cause analysis given a problem, or creating data pipelines for developing a failure prediction model.

This plug-and-play architecture enables additional data modality-dependent AI-powered contextualization services to be added. For example, wafer monitoring in semiconductor manufacturing using scanning electron microscope images is a data source that could be added to **connect**, with wafer defect classification as a **contextualization** flow. Such classified defects can be **fused** in the knowledge graph. An engineer in the fab can now **access** historical defects which could have resulted from a similar anomaly in the measurements observed currently.

The ability to select and add new contextualization flows within the framework enables applicability in a wide variety of industrial use cases. Section 3 lists the deployment aspects and Section 4 describes the benefits in real-world usage.

3 Deployment

SiWare is packaged as containers that can be deployed on Hybrid Cloud via the Red Hat OpenShift[2] platform. The knowledge graph uses Janusgraph[3] as the graph database and is compatible with Apache Tinkerpop standards[4]. SiWare has been deployed in both private cloud and IBM public cloud.

3.1 User Personas

There are 2 personas interacting with SiWare, the ***domain expert*** who creates and updates the knowledge graph, and the ***operator*** who consumes the knowledge for efficient execution of operations.

[1]https://www.ibm.com/cloud/object-storage

[2]https://www.redhat.com/en/technologies/cloud-computing/openshift/container-platform

[3]https://janusgraph.org/

[4]https://tinkerpop.apache.org/

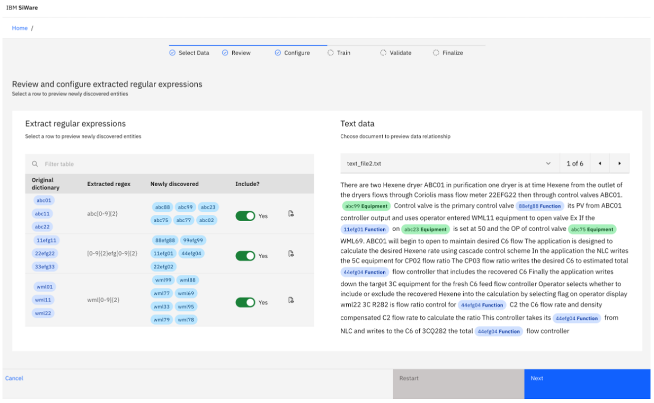

Figure 2: NER contextualization interface.

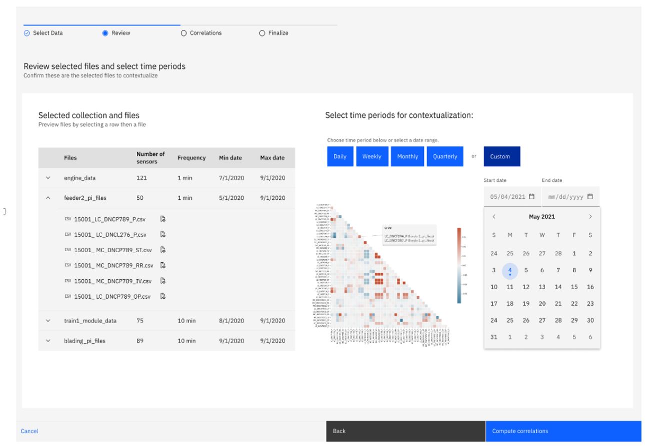

Figure 3: Sensor contextualization interface.

- The domain expert may or may not have data science knowledge and can interact with the system using a web user interface or a command line interface or via the APIs. Figure 2 presents the interaction screen for the NER contextualization flow where a domain expert can interactively train a deep learning model based on novel entity pattern extraction [Khabiri *et al.*, 2022]. Figure 3 presents the interaction for a sensor correlation contextualization flow. The domain expert can optionally integrate a domain semantic model of known entities and relationships, in a property graph format, using the Knowledge Capture Service. This can bootstrap the discovery process. An example of such a domain semantic model is illustrated in figure 4. The semantic model represents the connection between the common entities like tool, chamber, route, lot, wafer etc. The contextualization flows enable the linking of the operational data to the semantic model.

- The operator persona can consume the graph either through the web interface or via custom dashboards that are integrated with their workflow.

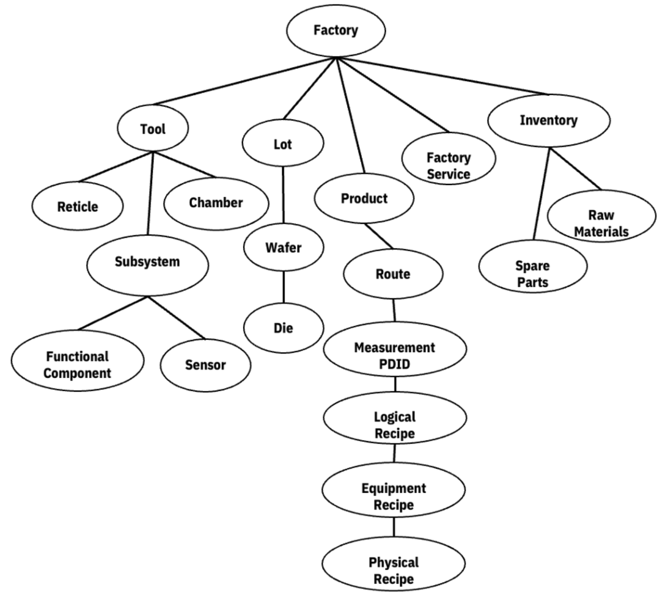

Figure 4: Semiconductor manufacturing domain semantic model containing tangible and intangible entities and relationships.

4 Evaluation

SiWare has been deployed in real-world scenarios to provide valuable operational insights using multi-modal data.

In **semiconductor manufacturing**, SiWare is being used to identify all relevant on-wafer measurements for a failure report that could suggest a root cause. Every time a failure report is searched for, the access mechanism in SiWare traverses the knowledge graph to return a measurement set of the order of 10s of values from $\sim 500K$ total number of measurements. This has dramatically reduced the mean time to diagnose a problem. In addition, our study indicates SME time savings of 50% in creating a semantic model and using it for semantic search and related problem diagnosis.

In **continuous flow manufacturing**, SiWare is being used to provide all contextual information in a single pane of glass for an operator to respond to an alarm. Given information about the distributed control systems, sensors, alarms, operation logs, and P&IDs, SiWare continuously organizes the knowledge into a representation that can be used to dynamically fetch the context needed for alarm response. Our experiments indicate a reduction of 80% of time spent on situation understanding and response.

Acknowledgments

We would like to thank Robert Baseman from IBM Research, Yorktown Heights, NY for sharing their semiconductor manufacturing expertise. The design input for the user interface was provided by Eileen Allen and realized with the help of John McGrath and Heena Purohit from the IBM Organic Growth Ventures team.

References

[Al-Moslmi *et al.*, 2020] Tareq Al-Moslmi, Marc Gallofré Ocaña, Andreas L Opdahl, and Csaba Veres. Named entity extraction for knowledge graphs: A literature overview. *IEEE Access*, 8:32862–32881, 2020.

[Bane, December 2020] Andy Bane. 95% of industrial asset data is unusable: It's time to change that – element analytics. https://www.elementanalytics.com/blog/95-of-industrial-asset-data-is-unusable-its-time-to-change-that, December 2020. Accessed: 2023-05-20.

[Erickson *et al.*, 2022] Nick Erickson, Xingjian Shi, James Sharpnack, and Alexander Smola. Multimodal automl for image, text and tabular data. In *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pages 4786–4787, 2022.

[Institute, July 2019] The Manufacturing Institute. The aging of the manufacturing workforce. https://www.themanufacturinginstitute.org/research/the-aging-of-the-manufacturing-workforce/, July 2019. Accessed: 2023-05-20.

[Khabiri *et al.*, 2019] Elham Khabiri, Wesley M. Gifford, Bhanukiran Vinzamuri, Dhaval Patel, and Pietro Mazzoleni. Industry specific word embedding and its application in log classification. In *CIKM*, pages 2713–2721. ACM, 2019.

[Khabiri *et al.*, 2022] Elham Khabiri, Bhavna Agrawal, Joseph Lindquist, Yingjie Li, and Anuradha Bhamidipaty. Big data techniques for industrial problems with little data. In *Big Data*, pages 2285–2289. IEEE, 2022.

[Patel *et al.*, 2022] Dhaval Patel, Giridhar Ganapavarapu, Srideepika Jayaraman, Shuxin Lin, Anuradha Bhamidipaty, and Jayant Kalagnanam. Anomalykits: Anomaly detection toolkit for time series. In *AAAI*, pages 13209–13211. AAAI Press, 2022.

[Sankhe *et al.*, 2021] Pranav Sankhe, Elham Khabiri, Bhavna Agrawal, and Yingjie Li. Tablenn: Deep learning framework for learning domain specific tabular data. In *IEEE BigData*, pages 4097–4102. IEEE, 2021.

[Shrivastava *et al.*, 2020] Shrey Shrivastava, Dhaval Patel, Nianjun Zhou, Arun Iyengar, and Anuradha Bhamidipaty. Dqlearn : A toolkit for structured data quality learning. In *IEEE BigData*, pages 1644–1653. IEEE, 2020.

[Zerveas *et al.*, 2021] George Zerveas, Srideepika Jayaraman, Dhaval Patel, Anuradha Bhamidipaty, and Carsten Eickhoff. A transformer-based framework for multivariate time series representation learning. In *KDD*, pages 2114–2124. ACM, 2021.