Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence (IJCAI-16)

# Semantic Framework for Industrial Analytics and Diagnostics*

**Gulnar Mehdi[1,2], Sebastian Brandt[1], Mikhail Roshchin[1] and Thomas Runkler[1,2]**
[1]Siemens Corporate Technology, Munich, Germany
{firstname.lastname}@siemens.com
[2]Technical University of Munich, Germany

## Abstract

Massive data streams from sensors and devices are prominent form of industrial data generated during condition-monitoring and diagnosis of complex systems. Data analytics and reasoning has emerged as a vital tool to harness massive data sets, providing insights into historical and real-time system conditions; enhanced decision support, reliability and cost reduction. However, application of data analytics is mainly challenged by the complexity of data-access, integration, domain-specific query support and contextual reasoning capabilities. The current state-of-the-art only uses dedicated scenarios and sensors, but this limits reuse, scalability and are not sufficient for an integrated solution. Our thesis investigates if semantic technology can be a potential solution to interact and leverage data analytics for operational use. First, we have studied related work and utilized ontology-based data access (OBDA) techniques for semantic interpretation of diagnosis for Siemens Turbine use-case. Secondly, we have extended our solution to support any analytical workflow by means of an ontology.

## 1 Introduction

Today, industries everywhere accumulate massive amount of data. This data may come from various heterogeneous systems, devices, sensors and applications in a variety of formats (structured and unstructured), velocities (stream and static) and volumes. Clearly, processing and analyzing diverse, complex and large volumes of data has many challenges. Nevertheless, leveraging this data for operations and predictive analysis can create a competitive edge and accelerate growth. Therefore, many industrial businesses have adopted data-driven strategies [Runkler, 2012] to enhance their diagnostic analytical capabilities and support decision-making. These data-driven strategies are often combined with knowledge models to provide better context for acquisition, validation, optimization and analytical reasoning.

*This research is supported by the Optique project with the grant agreement FP7-318338.

### 1.1 Challenges and Contribution

Roughly 74% of the organizations as per [Russom, 2011] have adopted some form of analytics today. Yet they still claim a number of potential barriers to implementation and exploitaton. Most of the reported technical challenges are due to the complexity problem. Here we define complexity in terms of:

i) **data access and integration**, where existing information systems, sensors and devices fail to form a unified new whole due to several kinds of heterogeneity (e.g. data models, schemas, software, integrity constraints etc.).

ii) **domain-specific query support**, where currently domain experts depend largely on IT experts to translate their requirements into dedicated set of queries and reports. This translation needs to be automated in order to reduce cost and allow domain-specific interaction to improve deeper analysis especially for non-IT experts [Kharlamov and others, 2013].

iii) **elaborative contextual reasoning capabilities**, where experts can retrieve and explore contextual information that is relevant (semantically similar) to their task at hand. For example, traversing data that relate to the design properties of a part exhibiting certain failures (such as blade crack).

iv) **system-level requirements**, where the complex structure of systems, processes, behaviors, dependencies, policies, heterogeneous actors with diverse interests and interactions provide a holistic view of the system for complete and integrated analysis.

Basic in-depth but semantics-less state-of-the-art analytics limits scalable data integration, interoperability, reusability and use dedicated sensors and scenarios mainly exposing numerical data without further qualification. This is why such applications today typically require collaboration of engineers, database analysts, statisticians/data miners and software developers. This large resource commitment makes it difficult to incorporate analytics as part of an overall business process and ultimately create a direct financial investment link. In our view, Semantic technology coupled with specialized programs for analytics can be an effective solution to reduce resource requirements and reduce complexity problems. This means that we provide an abstraction (semantic) layer above existing IT technologies to connect data, context and analytical processes. We aim to demonstrate our prototype for semantic analytics for different diagnostic tasks and describe our experiences in the context of a class of semantic applications for Siemens use-case.

## 2 Related Work

A large body of work has been published for model-free and model-based methods of diagnosis [Gertler, 1998]. Both approaches are computationally expensive, have large development efforts and fail to provide a view of the whole system. Current trend focuses on data-driven methods and ontologies as a knowledge representation approach for a domain. To the best of our knowledge, large number of applications exists for ontology-based data access (OBDA) [Kharlamov and others, 2013], semantic integration and search [Sheth, 2005]; whereas, very little progress has been made in the area of semantic-based analytics.

## 3 Approach

Our approach is to design generic, semantically annotated analytical models for diagnostic applications, to allow easy adaptation of analytical procedures over multiple abstractions, domains and needs. First, we plan to extend an existing OBDA system [Kharlamov and others, 2013] to model technical systems, analytical procedures and query reports. Our models will be described in an ontology. Secondly, we plan to model the context of the diagnostic state, in order to provide complete adaptation of the reasoning task.

## 4 Preliminary Results

Our first milestone is an upper-level ontology of a technical system. The ontology is expressed in OWL 2 QL, a tractable profile of the OWL 2 ontology language compatible with our current OBDA system [Kharlamov and others, 2013]. Currently, we support partonomy, configuration, functional system, product lines, process and system state concepts for any type of technical system. We have incorporated the standard sensor network ontology[1] to exemplify the idea of capturing part of relations between unknown individuals, sensors, measurables, and sensor metadata including measurement capabilities. R2RML mappings are used to connect data to the ontology and a SPARQL endpoint to query the instance data. Our initial use-case comes from the domain of remote monitoring of Siemens gas and steam turbines where users today have difficulty to analyze their performance because different machines have different sensors to contribute, different configurations, sensor tags, thresholds values and compositional structure. With our solution as depicted in Figure. 1, we overcome these problems by i) quering in a domain-specific language against an abstract domain model rather than the actual heterogeneous data sources. For example, as a domain expert would formulate, we are able to return
"***sensor observation values*** *including the* ***time of observation*** *between '2015-08-14' and '2015-08-20' being* ***produced by sensors*** *that observes* ***performance*** *'Speed & Power' and that are of* ***interest for parts*** *of type 'Gas Turbine' and where the observed values are within the sensors* ***measurement range***". and ii) making all this available even to analytics tool such as KNIME[2] in our case. This means that by using a semantic framework, we are able to support easy data-access, context reasoning and reduce customizations of any analytical workflow.

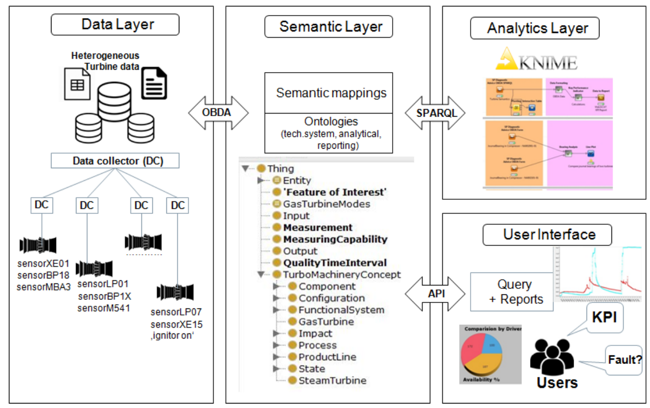


Figure 1: Semantic Framework for Turbine Analytics

## 5 Conclusion and Future Work

In this thesis, we go beyond existing OBDA system to provide a solution for data analytics addressing a number of important industry requirements for condition-monitoring and diagnosis. The focus is to incorporate domain specific knowledge and provide a simpler means of formulating and adapting analytical procedures relevant to real-world industrial applications. We also aim to make the semantic framework accessible to a range of professionals (design, service engineers, manufacturers etc.) who need assistance in integrating expert knowledge with data-storage and analytics practices used in Big Data infrastructure.

## References

[Gertler, 1998] Janos Gertler. *Fault detection and diagnosis in engineering systems*. CRC press, 1998.

[Kharlamov and others, 2013] Evgeny Kharlamov et al. Optique: Towards obda systems for industry. In *The Semantic Web: ESWC 2013 Satellite Events*, pages 125–140. Springer, 2013.

[Runkler, 2012] Thomas A. Runkler. *Data Analytics: Models and Algorithms for Intelligent Data Analysis*. Springer Science and Business Media, 2012.

[Russom, 2011] Philip Russom. Big data analytics. *TDWI Best Practices Report Fourth Quarter*, 2011.

[Sheth, 2005] Amit Sheth. From semantic search and integration to analytics. In *Dagstuhl Seminar Proceedings*. Schloss Dagstuhl-Leibniz-Zentrum fr Informatik, 2005.

[1] https://www.w3.org/2005/Incubator/ssn/ssnx/ssn

[2] https://www.knime.org/