Towards Simplification of Analytical Workflows With Semantics at Siemens (Extended Abstract)

Evgeny Kharlamov
University of Oxford, UK
University of Oslo, Norway

Gulnar Mehdi
Siemens CT, Germany

Ognjen Savkovic
Free University of Bozen-Bolzano, Italy

Guohui Xiao
Free University of Bozen-Bolzano, Italy

Steffen Lamparter
Siemens CT, Germany

Ian Horrocks
University of Oxford, UK

Arild Waaler
University of Oslo, Norway

***Abstract*—Analytical workflows are heavily used in large and data intensive companies. An important application of such workflows in Siemens is equipment analytics when equipment KPIs and reports are computed by aggregating equipment's operational, master, and analytical data. In Siemens this data satisfies big data dimensions and this dependence poses significant challenges in authoring, reuse, and maintenance of analytical workflows by engineers and data scientists. In this work we propose to address these problems by relying on semantic technologies: we use ontologies to give a high level representation of equipment's operational and master data and offer a high level language to express KPIs over ontologies. We implemented our approach, integrated it with KNIME, and evaluated at Siemens. This is a preliminary work and we are excited about its further extensions.**

Motivation. An analytical workflow typically consists of the following steps: *(1) data access* when users obtain permissions to enterprise data on different levels, *(2) data analysis and discovery*, when users extract and analyse data by interacting with the existing templates for dashboards and extract relevant knowledge from data, *(3) collaboration and sharing* when users find extra insights from the data and knowledge when shared with colleagues. Modern Business Intelligence systems and analytical platforms allow to combine these steps in analytical workflows and to iterate over them.

Step 2 in such workflows is where *self-service* is crucial. Indeed, an analytical platform should be easy to use so that business users from all skill levels can easily re-use a dashboard or modify and add components. In data intensive companies such as Siemens such self-service is often hampered by the fact that re-use and modification of dashboards and their components require deep knowledge of schemata and formats of underlying data. Due to the big data dimensions, such knowledge is only affordable to IT specialists, as illustrated by the following example.

Siemens Example. Siemens diagnostic engineers work at service centres and monitor complex industrial equipment such as power generating turbines. To this end they rely on analytical workflows that compute key performance indicators (KPIs) and reports by aggregating equipment's operational, master, and analytical data that together satisfy several important big data dimensions: variety, velocity, and volume. Indeed, operational data comes from sensors installed in turbines, a typical Siemens gas turbine has about 2,000 sensors that record and report temperature, pressure, vibration, etc in raw and aggregated form, and even in one turbine sensors of the same kind report data in at least a dozen of different formats. Master data is stored in several hundreds relational and XML databases, it describes equipment specification and other knowledge about the equipment manufacturing, configurations, deployment and use, it also contains history of weather forecasts, and even information about databases that store sensor data. Finally, analytical data are results of monitoring tasks conducted by Siemdens service centres for the last five years. All in all the data available for a diagnostic engineer is stored in several thousand databases and files with hundreds of different schemata and it is in the order of hundreds of terabytes: only operational and master data of each turbine is at least 15 GB, and for a fleet of turbines the data grows in average 30 GB per day. Thus, a dashboard where the engineer checks for a relatively simple task whether the purging[1] of a turbine is over has to rely on about 300 queries over at least seventy turbine operational and master databases, where most of these queries differ only in the format of the input data, while they compute essentially the same functions / outputs. In other words the queries differ in *how* they compute the same type of answers. Thus, modifying the purging dashboard would require the engineer to update these 300 'how-queries' which is time consuming, error prone, and require a relatively strong IT background: one should be able to understand the variety of database schemata behind the 300 queries. Finally, keeping such dashboard up-to-date requires to add new 'how-queries' and update them in order to be up-to-date with the changes in

[1]Purging is the process of flushing out liquid fuel nozzles or other parts which may contain undesirable residues.

978-1-5386-5035-6/18/$31.00 ©2018 IEEE

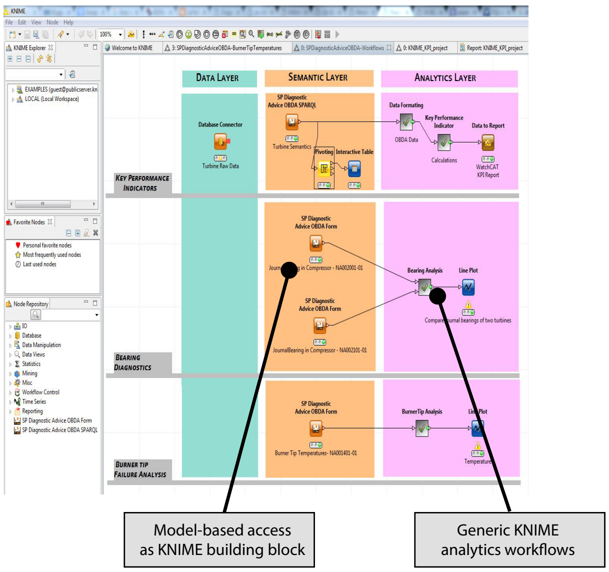

Figure 1. Integration with KNIME

the schemata of the underlying DBs.

Semantic Approach: General Idea. In order to address this aforementioned challenge at Siemens and to enable the self-service for a large community of Siemens engineers, we propose to add on top of Siemens databases an abstraction layer that consists of ontologies [7], [16].[2] Then, Siemens engineers will be shielded from the variety of data formats and instead of modifying the 300 'how-queries' over Siemens databases, they will have to modify 1 'what-query' over the abstraction layer, that is, the query that essentially states *what* has to be extracted from turbine data. The ontology in this case will provide 'a single point of semantic data access' for the engineers, and allow to pose queries over the integrated data sources in terms of a user-oriented conceptual model that abstracts away both the variety of database schemata and the complexity of implementation-level details typically encountered in database schemata. In order to relate the abstraction and the data layer, the ontology should be connected to the data via a set of ETL-style *mappings*: declarative specifications that relate ontological terms with queries over the underlying data. Then, a system that implements query processing over the integrated abstraction and data layer will automatically translate ontological queries, e.g., expressed in the SPARQL query language [3], into database queries, i.e., expressed in SQL, and delegate execution of SQL queries to the database systems hosting the data. We refer the reader for further details of such semantic approach to the literature on Ontology-Based Data Access (OBDA) [26].

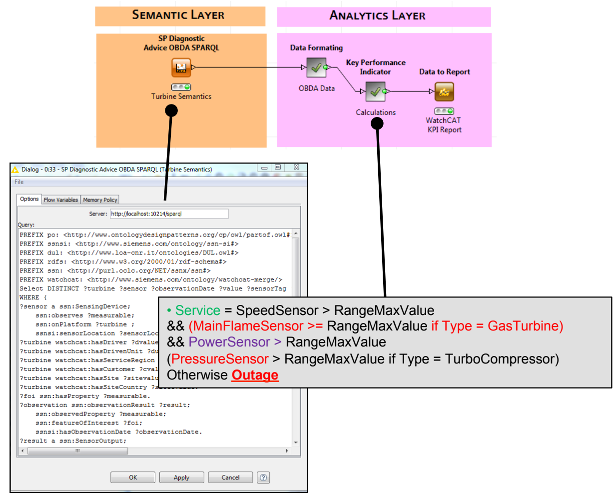

Figure 2. A semantic query and an analytical function over an ontology

Implementation of Semantic Approach for Siemens. We implemented the semantic approach and deployed our system at Siemens for evaluation and testing. For this purpose we extended KNIME [4] data analytics system to streamline query answers into analytics, and a configurable plug-in for R analytics [5]. In Figure 1 we give an example of three KNIME analytical workflows extended with our semantic layer. The first workflow computes a turbine KPI, the second compares joint bearings of two compressors, while the third performs failure analyses for burner tips of a turbine. Observe that the engineer has to define nodes in the analytical workflows only in the semantic and analytical layers (the yellow and pink vertical stripe in the screenshot), and the connection to the data layer is then done automatically (the blue stripe).

Example Semantic Analytical Task. To further illustrate our approach, observe in Figure 2 an example SPARQL query over the Siemens turbine ontology. The data is produced by this query are then used to compute KPIs and to check for outage in the turbines. Observe that the node of

[2] An ontology us a semantically rich conceptual model of the problem domain that captures the domain in terms of classes and properties that relate entities that populate classes and assign data values to entities. A number of standardised machine processable ontology languages have been proposed and standardised, e.g., RDF, OWL 2. Ontologies were successfully used in many applications [7], including smart (Web) search [2], [12], [1] and query formulation [30], [29], [28], [32], [4], [27], medicine [21], e-commerce [3], media, data integration [25], [9], knowledge modelling [20], [13], [14], [31], and industry [17], [8], [15], [11], [10].

[3] https://www.w3.org/TR/rdf-sparql-query/

[4] KNIME, the Konstanz Information Miner, is an open source data analytics, reporting and integration platform. KNIME integrates various components for machine learning and data mining through its modular data pipelining concept. A graphical user interface allows assembly of nodes for data preprocessing (ETL: Extraction, Transformation, Loading), for modeling and data analysis and visualization.

[5] https://www.r-project.org/

the workflow diagram that corresponds to KPI computation is defined using KNIME rules that rely on the our extension of the standard KNIME syntax and they are formulated against the ontological concepts. We developed the syntax and semantics of these rules in order to support Siemens turbine diagnostic and KPI computation tasks [23], [22], [24], [19], [18]. These rules say that a turbine is deemed to be in service at any time if either it is a gas turbine or a turbo compressor that satisfies extra conditions. In the former case these conditions say that the sensor signal of the rotor speed sensor should have readings above its characteristic operational speed value, the main flame sensor reading should show that the flame is on, and the turbine should generate power, i.e., Power Sensor should be above the value characteristic for that turbine while generating power. In the latter case the conditions are that the generated pressure should be at or above the nominal pressure specified for the machine.

Observe that due to ontologies each occurrence of “RangeMaxValue” in the KNIME rules is a different type of value read from the static configuration data of the machine and this can be encoded using property hierarchy. Indeed, for the latter case one can achieve it by stating that “RunningSpeedConfigValue” is a sub-property or “RangeMaxValue” and in the latter case that “MainFlameOnSignal” is a sub-property of “RangeMaxValue”. Moreover, an advantage of such semantically-backed KNIME diagrams is that one can talk about specific values in turbines and even compressors of different types at an abstract level, without giving details of such appliances. Finally, observe that KNIME has a sophisticated reporting functionality which we exploit in our system: the last node in the work-flow diagram, called Data to Report, summarises the KPI for all turbines across a specific fleet and builds a ready-to-use report for them.

Conclusions and Future Work. Analytical workflows are heavily used in large and data intensive companies. In Siemens this kind of analytics is heavily data dependent and this dependence poses significant challenges in authoring, reuse, and maintenance of analytical workflows by engineers and data scientists due to the big data dimensions. In this work we propose to address these problems by relying on semantic technologies: we use ontologies to give a high level representation of equipment’s operational and master data and offer a high level language to express KPIs over ontologies. We implemented our approach and integrated it with KNIME. We are currently evaluating our work at Siemens and plan to develop rule learning techniques [5], [6]. This is a preliminary work and we are excited about further steps and would like to share it with the IEEE Big Data community.

Acknowledgements: This work is partially funded by the EU projects Optique (FP7-ICT-318338) and TheyBuyForYou (H2020-780247), by the EPSRC projects MaSI^3, DBOnto, ED^3, and by the SIRIUS Centre, Norwegian Research Council project number 237898.

REFERENCES

[1] Marcelo Arenas, Bernardo Cuenca Grau, Evgeny Kharlamov, Sarunas Marciuska, and Dmitriy Zheleznyakov. Faceted search over ontology-enhanced RDF data. In *CIKM*, pages 939–948, 2014.

[2] Marcelo Arenas, Bernardo Cuenca Grau, Evgeny Kharlamov, Sarunas Marciuska, and Dmitriy Zheleznyakov. Faceted search over RDF-based knowledge graphs. *J. Web Sem.*, 37-38:55–74, 2016.

[3] Bruno Charron, Yu Hirate, David Purcell, and Martin Rezk. Extracting semantic information for e-commerce. In *ISWC*, pages 273–290, 2016.

[4] Gong Cheng and Evgeny Kharlamov. Towards a semantic keyword search over industrial knowledge graphs (extended abstract). In *2017 IEEE International Conference on Big Data, BigData 2017, Boston, MA, USA, December 11-14, 2017*, pages 1698–1700, 2017.

[5] Vinh Thinh Ho, Daria Stepanova, Mohamed H. Gad-Elrab, Evgeny Kharlamov, and Gerhard Weikum. Learning rules from incomplete kgs using embeddings. In *ISWC Posters & Demonstrations*, 2018.

[6] Vinh Thinh Ho, Daria Stepanova, Mohamed H. Gad-Elrab, Evgeny Kharlamov, and Gerhard Weikum. Rule learning from knowledge graphs guided by embedding models. In *ISWC*, pages 72–90, 2018.

[7] Ian Horrocks. Ontologies and the semantic web. *Commun. ACM*, 51(12):58–67, 2008.

[8] Ian Horrocks, Martin Giese, Evgeny Kharlamov, and Arild Waaler. Using Semantic Technology to Tame the Data Variety Challenge. *IEEE Internet Computing*, 20(6):62–66, 2016.

[9] Ernesto Jiménez-Ruiz, Evgeny Kharlamov, Dmitriy Zheleznyakov, Ian Horrocks, Christoph Pinkel, Martin G. Skjæveland, Evgenij Thorstensen, and José Mora. BootOX: Practical Mapping of RDBs to OWL 2. In *ISWC*, pages 113–132, 2015.

[10] Evgeny Kharlamov, Sebastian Brandt, Martin Giese, Ernesto Jiménez-Ruiz, Yannis Kotidis, Steffen Lamparter, Theofilos Mailis, Christian Neuenstadt, Özgür L. Özçep, Christoph Pinkel, Ahmet Soylu, Christoforos Svingos, Dmitriy Zheleznyakov, Ian Horrocks, Yannis E. Ioannidis, Ralf Möller, and Arild Waaler. Enabling semantic access to static and streaming distributed data with optique: demo. In *DEBS*, pages 350–353, 2016.

[11] Evgeny Kharlamov, Sebastian Brandt, Ernesto Jiménez-Ruiz, Yannis Kotidis, Steffen Lamparter, Theofilos Mailis, Christian Neuenstadt, Özgür L. Özçep, Christoph Pinkel, Christoforos Svingos, Dmitriy Zheleznyakov, Ian Horrocks, Yannis E. Ioannidis, and Ralf Möller. Ontology-based integration of streaming and static relational data with optique. In *SIGMOD*, pages 2109–2112, 2016.

[12] Evgeny Kharlamov, Luca Giacomelli, Evgeny Sherkhonov, Bernardo Cuenca Grau, Egor V. Kostylev, and Ian Horrocks. Semfacet: Making hard faceted search easier. In *CIKM*, pages 2475–2478, 2017.

[13] Evgeny Kharlamov, Bernardo Cuenca Grau, Ernesto Jiménez-Ruiz, Steffen Lamparter, Gulnar Mehdi, Martin Ringsquandl, Yavor Nenov, Stephan Grimm, Mikhail Roshchin, and Ian Horrocks. Capturing industrial information models with ontologies and constraints. In *The Semantic Web - ISWC 2016 - 15th International Semantic Web Conference, Kobe, Japan, October 17-21, 2016, Proceedings, Part II*, pages 325–343, 2016.

[14] Evgeny Kharlamov, Bernardo Cuenca Grau, Ernesto Jiménez-Ruiz, Steffen Lamparter, Gulnar Mehdi, Martin Ringsquandl, Yavor Nenov, Stephan Grimm, Mikhail Roshchin, and Ian Horrocks. SOMM: industry oriented ontology management tool. In *ISWC Posters & Demonstrations*, 2016.

[15] Evgeny Kharlamov, Dag Hovland, Martin G. Skjæveland, Dimitris Bilidas, Ernesto Jiménez-Ruiz, Guohui Xiao, Ahmet Soylu, Davide Lanti, Martin Rezk, Dmitriy Zheleznyakov, Martin Giese, Hallstein Lie, Yannis E. Ioannidis, Yannis Kotidis, Manolis Koubarakis, and Arild Waaler. Ontology Based Data Access in Statoil. *J. Web Sem.*, 44:3–36, 2017.

[16] Evgeny Kharlamov, Yannis Kotidis, Theofilos Mailis, Christian Neuenstadt, Charalampos Nikolaou, Özgür L. Özçep, Christoforos Svingos, Dmitriy Zheleznyakov, Sebastian Brandt, Ian Horrocks, Yannis E. Ioannidis, Steffen Lamparter, and Ralf Möller. Towards analytics aware ontology based access to static and streaming data. In *ISWC*, pages 344–362, 2016.

[17] Evgeny Kharlamov, Theofilos Mailis, Gulnar Mehdi, Christian Neuenstadt, Özgür L. Özçep, Mikhail Roshchin, Nina Solomakhina, Ahmet Soylu, Christoforos Svingos, Sebastian Brandt, Martin Giese, Yannis E. Ioannidis, Steffen Lamparter, Ralf Möller, Yannis Kotidis, and Arild Waaler. Semantic access to streaming and static data at Siemens. *J. Web Sem.*, 44:54–74, 2017.

[18] Evgeny Kharlamov, Ognjen Savkovic, Martin Ringsquandl, Guohui Xiao, Gulnar Mehdi, Elem Güzel Kalayci, Werner Nutt, Mikhail Roshchin, Ian Horrocks, and Thomas A. Runkler. Diagnostics of trains with semantic diagnostics rules. In *ILP*, pages 54–71, 2018.

[19] Evgeny Kharlamov, Ognjen Savkovic, Guohui Xiao, Rafael Peñaloza, Gulnar Mehdi, Mikhail Roshchin, and Ian Horrocks. Semantic rules for machine diagnostics: Execution and management. In *CIKM*, pages 2131–2134, 2017.

[20] Evgeny Kharlamov, Dmitriy Zheleznyakov, and Diego Calvanese. Capturing model-based ontology evolution at the instance level: The case of DL-Lite. *J. Comput. Syst. Sci.*, 79(6):835–872, 2013.

[21] Dennis Lee, Ronald Cornet, Francis Y. Lau, and Nicolette de Keizer. A survey of SNOMED CT implementations. *Journal of Biomedical Informatics*, 46(1):87–96, 2013.

[22] Gulnar Mehdi, Evgeny Kharlamov, Ognjen Savkovic, Guohui Xiao, Elem Guzel Kalayci, Sebastian Brandt, Ian Horrocks, Mikhail Roshchin, and Thomas A. Runkler. Semantic rule-based equipment diagnostic. In *ISWC Posters & Demonstrations*, 2017.

[23] Gulnar Mehdi, Evgeny Kharlamov, Ognjen Savkovic, Guohui Xiao, Elem Güzel Kalayci, Sebastian Brandt, Ian Horrocks, Mikhail Roshchin, and Thomas A. Runkler. Semantic rule-based equipment diagnostics. In *ISWC*, pages 314–333, 2017.

[24] Gulnar Mehdi, Evgeny Kharlamov, Ognjen Savkovic, Guohui Xiao, Elem Güzel Kalayci, Sebastian Brandt, Ian Horrocks, Mikhail Roshchin, and Thomas A. Runkler. Semdia: Semantic rule-based equipment diagnostics tool. In *CIKM*, pages 2507–2510, 2017.

[25] Christoph Pinkel, Carsten Binnig, Ernesto Jiménez-Ruiz, Evgeny Kharlamov, Wolfgang May, Andriy Nikolov, Ana Sasa Bastinos, Martin G. Skjæveland, Alessandro Solimando, Mohsen Taheriyan, Christian Heupel, and Ian Horrocks. RODI: benchmarking relational-to-ontology mapping generation quality. *Semantic Web*, 9(1):25–52, 2018.

[26] Antonella Poggi, Domenico Lembo, Diego Calvanese, Giuseppe De Giacomo, Maurizio Lenzerini, and Riccardo Rosati. Linking data to ontologies. *J. Data Semantics*, 10:133–173, 2008.

[27] Ahmet Soylu, Martin Giese, Ernesto Jiménez-Ruiz, Evgeny Kharlamov, Dmitriy Zheleznyakov, and Ian Horrocks. Why not simply Google? In *NordiCHI*, pages 1039–1042, 2014.

[28] Ahmet Soylu, Martin Giese, Ernesto Jiménez-Ruiz, Evgeny Kharlamov, Dmitriy Zheleznyakov, and Ian Horrocks. Ontology-based end-user visual query formulation: Why, what, who, how, and which? *Universal Access in the Information Society*, 16(2):435–467, 2017.

[29] Ahmet Soylu, Martin Giese, Rudolf Schlatte, Ernesto Jiménez-Ruiz, Evgeny Kharlamov, Özgür L. Özçep, Christian Neuenstadt, and Sebastian Brandt. Querying industrial stream-temporal data: An ontology-based visual approach. *JAISE*, 9(1):77–95, 2017.

[30] Ahmet Soylu, Evgeny Kharlamov, Dmitriy Zheleznyakov, Ernesto Jiménez-Ruiz, Martin Giese, Martin G. Skjæveland, Dag Hovland, Rudolf Schlatte, Sebastian Brandt, Hallstein Lie, and Ian Horrocks. OptiqueVQS: A visual query system over ontologies for industry. *Semantic Web*, 9(5):627–660, 2018.

[31] Dmitriy Zheleznyakov, Evgeny Kharlamov, and Ian Horrocks. Trust-sensitive evolution of dl-lite knowledge bases. In *AAAI*, pages 1266–1273, 2017.

[32] Dmitriy Zheleznyakov, Evgeny Kharlamov, Vidar Klungre, Martin G. Skjæveland, Dag Hovland, Martin Giese, Ian Horrocks, and Arild Waaler. Keywdb: A system for keyword-driven ontology-to-rdb mapping construction. In *Proceedings of the ISWC 2016 Posters & Demonstrations Track co-located with 15th International Semantic Web Conference (ISWC 2016), Kobe, Japan, October 19, 2016.*, 2016.