Smart Data Access:
Semantic Web Technologies for Energy Diagnostics

Dr. Ulli Waltinger
Siemens AG - Corporate Technology - Research & Technology Center
Otto-Hahn-Ring 6 - 81739 Munich, Germany
ulli.waltinger@siemens.com

In today`s (big) data-intensive world, scalable technologies enabling the efficient management, storage and analysis of large data set are needed. However, the underlying logic of the emerging data-driven business is very different to the established understanding of the traditional often technology-driven industries. As large and complex data are generate almost everywhere in exponentially growth, it is becoming challenging to process and analyze them efficiently by utilizing traditional data analytic and mining techniques.

Semantic web technologies and data mining techniques for unified information access and predictive analytics bring together a multidisciplinary skill set that allows and supports the combination of actual and expected values to plan, predict, and monitor business scenarios and their impact throughout an organization. These techniques play nowadays a key role for challenges such as the optimization of complex system behavior, real-time decision support in operational processes, condition monitoring for predictive maintenance such as failures and fatigue detection, and to increase the efficiency of remote monitoring operations.

Especially the processing of data in diagnostics and search related purposes as for instance in alarm management systems become more and more complicated, which can be attributed to the following constraints: [Volume] The diagnosis process, the search for root causes or the calculation of key performance indicators relies on handling large amounts of data. Nowadays, collected data sums up to hundreds of TB for individual use cases (Waltinger et al. 2014). [Velocity] In addition to the large amounts of data, more and more data is generated every day. Archived and/or continuous incoming live/streaming data have to be included into the diagnose process to achieve proper results (Giese et al. 2013). [Variety] Different vendors of machines or single components, coupled with historical or compatibility reasons, lead to multiple different logical and physical data representations and forms. Providing a unified and efficient access to all the different logical models is complex and cumbersome. [Veracity] Finally, the aspect of data quality - faulty or missing information leads to high expenses for companies for several reasons. Bad decisions based on wrong information may lead to accidents, resulting in machine damage or even human harm. Additional costs are generated when internal employees are unable to find their required knowledge in time or at all. Consequently, expensive external experts are required (Feldman and Sherman 2001). Varying data representations and the difficulties with processing unstructured data, require additional support for engineers with predefined search queries or diagnostic tools. The search queries have to be updated and adapted to the different logical representations or new unstructured events regularly. Engineers in the oil and gas industry spend about 30% to 70% of their time searching for data and assessing the quality of the data (Alcook 2009).

Due to the steady development of new key technologies within the area of semantic web and standards like SPARQL Protocol and RDF Query Language (SPARQL) or Web Ontology Language (OWL), new approaches and promising ideas emerge to solve diagnosis and search problems also in the area of energy diagnostics. As for instance, automating and offering a general applicable natural language interface (Waltinger et al. 2013) and/or Ontology-based

interpretation (Tran et al. 2007) reduces the error-proneness and simplifies the query optimization, therefore speeding up the response time. Hence, reducing this amount of time will lead to great benefits for the engineers and companies itself.

In this talk, we present two different business-driven use cases derived from the domain of Energy diagnostics that builds heavily upon semantic web technologies. We describe the motivation and current needs for semantic web technologies to industry data, where eligible technologies and data storage possibilities are analyzed. Within the first use case, we describe the benefit of automatic SPARQL query construction (Lehmann et al 2011) for effective natural language queries by unifying the information derived from the Linked Data Cloud with Corporate Repositories. In the second use case, we describe the benefit of separating Ontology-based data modeling and associated large-scale diagnostic sensor data within a real-time processing setup. We analyze the performance (Schmidt et al., 2010) of using RDBMS, RDF and Triple Stores for the knowledge representation. The proposed approaches will be evaluated on the basis of a query catalog by means of query efficiency, accuracy, and data structure performance. The results show, that natural language access to industry data using ontology's, is a simple but effective approach to improve diagnosis and data search for a broad range of users. Furthermore, virtual RDF graphs do support the DB-driven knowledge graph representation process (Kumar et al, 2011), but do not perform efficient under industry conditions in terms of performance and scalability.

References:

- Alcook, P. 2009. R. Crompton (2008), class and stratification, 3rd edition. cambridge. Journal of Social Policy 38.
- Damljanovic, D.; Agatonovic, M.; and Cunningham, H. 2012. Freya: An interactive way of querying linked data using natural language. In The Semantic Web: ESWC 2011 Workshops, 125–138. Springer.
- Feldman, S., and Sherman, C. 2001. The high cost of not finding information. IDC Whitepaper.
- Giese, M.; Calvanese, D.; Haase, P.; Horrocks, I.; Ioannidis, Y.; Kllapi, H.; Koubarakis, M.; Lenzerini, M.; Mller, R.; Rodriguez-Muro, M.; zcep, .; Rosati, R.; Schlatte, R.; Schmidt, M.; Soylu, A.; and Waaler, A. 2013. Scalable end-user access to big data. In Akerkar, R., ed., Big Data Computing. CRC Press.
- Kumar, A. P.; Kumar, A.; and Kumar, V. N. 2011. A comprehensive comparative study of SPARQL and SQL. International Journal of Computer Science and Information Technologies 2(4):1706–1710.
- Lehmann, J., and Bühmann, L. 2011. Autosparql: Let users query your knowledge base. In The Semantic Web: Research and Applications. Springer. 63–79.
- Schmidt, M.; Meier, M.; and Lausen, G. 2010. Foundations of SPARQL query optimization. In Proceedings of the 13th International Conference on Database Theory, 4–33. ACM.
- Tran, T.; Cimiano, P.; Rudolph, S.; and Studer, R. 2007. Ontology-based interpretation of keywords for semantic search. In The Semantic Web. Springer. 523–536.
- Waltinger, U.; Tecuci, D.; Olteanu, M.; Mocanu, V.; and Sullivan, S. 2013. USI Answers: Natural language question answering over (semi-) structured industry data. In Munoz-Avila, H., and Stracuzzi, D. J., eds., Proceedings of the Twenty-Fifth Innovative Applications of Artificial Intelligence Conference, IAAI 2013, July 14-18, 2013, Bellevue, Washington, USA.
- Waltinger, U.; Tecuci, D.; Picioroaga, F.; Grigoras, C.; and Sullivan, S. 2013. Market Intelligence: Linked Data-driven Entity Resolution for Customer and Competitor Analysis. Aaloborg, North Denmark. Proceedings of the 13th International Conference on Web Engineering (ICWE 2013)
- Waltinger, U.; Tecuci, D.; Olteanu, M.; Mocanu, V.; and Sullivan, S. 2014. Natural Language Access to Enterprise Data, in: AI Magazine, Vol 35, No 1, pp 38-52.