



Towards Semantic Reasoning in Knowledge Management Systems

Gulnar Mehdi, Sebastian Brandt, Mikhail Roshchin, Thomas Runkler

► To cite this version:

Gulnar Mehdi, Sebastian Brandt, Mikhail Roshchin, Thomas Runkler. Towards Semantic Reasoning in Knowledge Management Systems. 4th IFIP International Workshop on Artificial Intelligence for Knowledge Management (AI4KM), Jul 2016, New York, NY, United States. pp.132-146, 10.1007/978-3-319-92928-6_9 . hal-01950015

HAL Id: hal-01950015

<https://inria.hal.science/hal-01950015>

Submitted on 10 Dec 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

Towards Semantic Reasoning in Knowledge Management Systems^{*}

Gulnar Mehdi^{1,2}, Sebastian Brandt² Mikhail Roshchin², and
Thomas Runkler^{1,2}

¹ Siemens Corporate Technology, Munich, Germany
`firstname.lastname@siemens.com`,

² Technical University of Munich, Germany

Abstract. Modern applications of AI systems rely on their ability to acquire, represent and process expert knowledge for problem-solving and reasoning. Consequently, there has been significant interest in both industry and academia to establish advanced knowledge management (KM) systems, promoting the effective use of knowledge. In this paper, we examine the requirements and limitations of current commercial KM systems and propose a new approach to semantic reasoning supporting Big Data access, analytics, reporting and automation related tasks. We also provide comparative analysis of how state-of-the-art KM systems can benefit from semantics by illustrating examples from the life-sciences and industry. Lastly, we present results of our semantic-based analytics workflow implemented for Siemens power generation plants.

Keywords: knowledge management, semantic technology, data-access, analytics, automation

1 Introduction

It is well established today that knowledge is the core element of any AI based system, be it small robots like roomba [1] or large-scale applications such as IBM Watson [2]. Consequently, the transition of the global economy towards knowledge economy is an evident and prominent process in our information society. Even small-scale industries today value knowledge resources and use them for gaining a competitive edge. From a research and technology view point, much progress has been made in enabling information systems to leverage knowledge for decision-making and analysis. The scope of these systems is to construct, manage, share and process the applicable knowledge for their respective tasks. For example, KM systems are build to not only manage large repositories of biomedical data coming in from lab reports, patient records, research papers, and medical imaging, but also to store and analyze useful patterns and knowledge from them [3]. Nevertheless, implementing such a KM system is complex

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

and incorporates multi-faceted concepts from various disciplines and business practices. For instance, information scientists consider taxonomies, subject headings, and classification schemes to represent knowledge, whereas consulting firms actively promote practices and methodologies to capture corporate knowledge assets and organizational memory. In the biomedical industry, knowledge management practices often need to leverage existing clinical decision support, information retrieval, and digital library techniques to capture and deliver tacit and explicit biomedical knowledge [3]. Engineers, on the other hand rely on knowledge and data-driven strategies for design, manufacturing and maintenance of their artifacts. According to KPMG and the Conference Board [4], 80 percent of the world's biggest companies have knowledge management efforts under way, especially in the medical domain. Nevertheless, the full potential of KM systems is yet to be unlocked. Specifically, the challenges with respect to knowledge representation, search, integration, data-access and reasoning etc. are well understood in the research community but existing solutions have rarely resulted in widely adopted practical implementation. In this paper, we discuss requirements and limitations of the existing semantic and non-semantic solutions for KM and propose a new approach to semantic-based knowledge management that does not only enhance the feature set and usability but also supports analytics, reporting workflows, automation and big data infrastructures. In Section 2, we present the related approaches and further discuss ontology-based KM systems in Section 3. Section 4 and 5 presents success stories from the healthcare and engineering domain along with their extended requirement set. Section 6 describes the current challenges for semantic approaches and in Section 7 we propose our solution along with the results from Siemens Turbo-machinery use-case.

2 Related Approaches

Building and using KM system involves many tasks, see Figure 1. First and foremost is knowledge acquisition and representation, into which the scientific community has invested much time and effort. Knowledge engineering [5] methodologies for building expert systems have applied knowledge acquisition techniques (e.g. interviewing, protocol analysis, simulation, personal construct theory, card sorting, etc.) for eliciting the tacit knowledge from domain experts. Knowledge acquisition techniques are applied in order to develop knowledge repositories in knowledge management systems for formally documenting knowledge in a machine-processable way. To represent knowledge, a knowledge taxonomy and knowledge mapping are typically constructed for serving as a framework for building knowledge repositories [5]. Ontologies and ways for representing acquired knowledge (rules, cases, scripts, frames/objects, semantic networks, etc.) are typically created in the AI field for building expert and other intelligent systems [6]. Natural language and speech understanding front-ends as interfaces to knowledge management systems are important additions to enhance search over and dissemination of knowledge. Data mining and knowledge discovery techniques are employed to inductively look for trends, relationships, clusters

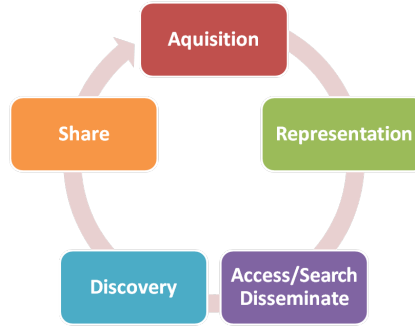


Fig. 1. Different facets of knowledge management systems

and possibly new insights and information from knowledge repositories [8]. On-line communities with a common interest in knowledge management are ways of sharing and distributing knowledge. Intelligent agents on the other hand are also applied to analyze the knowledge, email, web pages, and the like and to disseminate appropriate summaries or individual pieces of information and knowledge to those who should best make use of it [7].

Limitations Most previous works on KM systems has focused on its success factors [11], [9] and [10]. There are a few studies on challenges or limitations to KM systems [12], [13], [14]. However, they often offer intrinsic business value but KM systems do not always improve organizational performance because there exist some discrepancies between innovation and performance. We classify the requirements and limitations into two types: technological factors and social/cultural factors involving people [15]. Following are the significant weaknesses of current KM systems:

- i) **Searching often ignores context** in current KM system. Applications today provide key-word based search functionalities that often retrieve irrelevant information when terms have different meaning in different context or fail to relate different pieces of information into a meaningful context.
- ii) **Inefficient access and integration of information** is a major challenge in current systems. Human browsing and reading is required to extract and integrate information from different sources. Existing KM systems rely on labor intensive extract-transform-load jobs because the automatic agents do not integrate and possess common sense knowledge required to extract information from heterogeneous sources.
- iii) **Maintaining knowledge** is the main pain points of the state-of-the-art. It becomes difficult and time-consuming activity when the knowledge repositories become large or reach the level of Big Data infrastructures, for instance, hadoop clusters, teradata warehouse etc. Existing solutions lack transparency and usability with domain oriented interfaces as well as find it difficult to keep the knowledge consistent, correct and up-to-date. If the so called 'grain

size' of the knowledge representation is chosen properly (i.e. small enough to be comprehensible but large enough to be meaningful to the domain expert) then the KM system will allow great flexibility for adding, removing or changing as well as querying knowledge in the system.

- iv) **Generation of documents and reports** is too cumbersome and slow process to execute in any industrial setting. Whereas, automated and context-aware authoring can be of greater advantage in enabling content according to user profile or other aspects of relevance. The generation of such information presentation would require machine-accessible representations of the semantics of the information sources.
- v) **Lack of shared understanding**, is another aspect which involves how people, organization and KM systems communicate with one another. It is obvious that each stakeholder has different needs and background context.

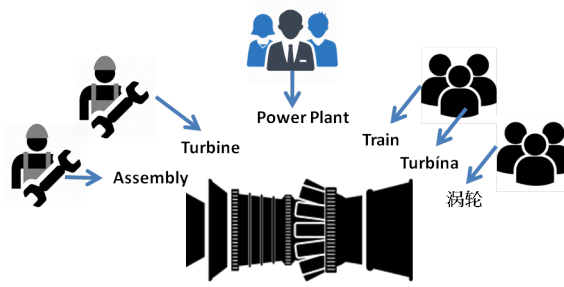


Fig. 2. Different naming but same semantics

Figure 2. represents a real-life example where a diverse set of users use different jargon in different languages to define the same concept or subject matter, that is Siemens Gas turbine in this case. This lack of shared understanding leads to disparate modelling methods, paradigm, languages and software tool which ultimately limits interoperability, reuse and sharing.

3 Semantic-based Knowledge Management System

The vision of semantic technologies is to provide human-readable artifacts annotated with meta-information. This meta-information defines what the artifact is about in a machine-processable way. Ontologies are at the core of semantic technologies. Ontology is a formal explicit description of concepts, relations and properties of a domain. Knowledge in ontologies can be formalised using five kinds of components: classes, relations, functions, axioms and instances [17]. The meta-information together with domain ontologies provides an arena of knowledge-driven systems and automated services such as information access,

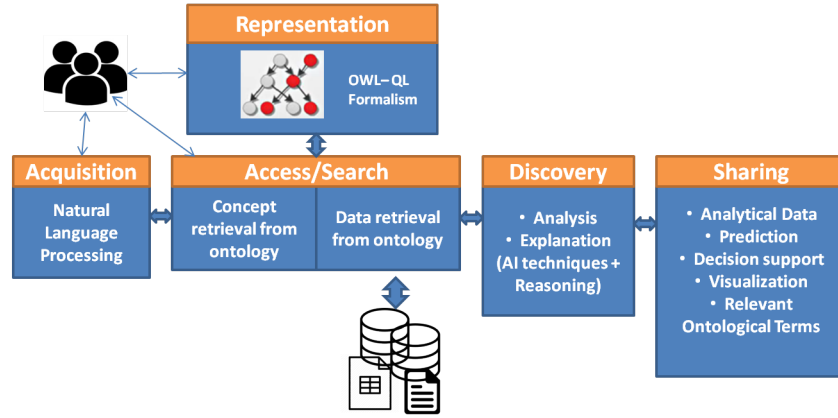


Fig. 3. Semantic Solution to knowledge management systems

reasoning services etc. It facilitates knowledge sharing and re-use and offers a wide feature set to support knowledge management capabilities. Figure 3. describes the unifying semantic architecture where ontologies are used throughout the knowledge management life cycle.

Ontologies are populated in an automated or semi-automated fashion from heterogeneous data sources. This serves the purpose of knowledge acquisition. This utility helps users to relate or map their domain concepts with the data model underneath and extract relevant information when required without performing cumbersome ETL jobs. As knowledge is acquired, it is then represented in an ontology language. The ontology language (for example: ontology web language - OWL, RDF etc.) is able to represent domain knowledge with a clear semantics as well as to provide redundancy or inconsistency checks. Knowledge maintenance is simplified by following standard design and modular approach to build and manage semantic models. Finally, the knowledge is made available to the end-user by means of semantic-based search, sharing, summarizing, visualization and organization.

3.1 Benefits of using ontologies

A key advantage of ontologies over many other knowledge representation formalisms is their formally well-defined semantics. They specifically support subsumption relations, multiple view points and hierarchies including partonomies, and inferred relationships. Ontology inference engines are used to derive implicit knowledge from explicit statements, detect redundancies and inconsistencies, and discover relationships that may not have been clear to the author of the ontology in the first place [17]. Currently, the most commonly used ontology formalism is OWL and its sub-languages. Some additional characteristics of ontology [22] addressing key challenges in the KM domain are:

- Ontologies clarifies the structure of knowledge and domain for an effective KM system.

- They separate factual knowledge about the domain from problem-solving knowledge.
- They facilitate sharing and re-using knowledge as well as interoperability of information resources between humans and software agents.
- They make searching, querying and browsing information more effective. For instance, a web site or a corporate intranet can organise its content according to some ontology which then can be utilised to improve the quality of searches. For instance, generalisation or specialisation of information can help in assisting users. It is a short distance from general search applications to knowledge management applications. One of the big challenges in knowledge management is to find knowledge and information that is relevant, and here ontologies have a lot of potential.
- They promote ease of maintenance of knowledge models and artefacts in KM system. The unified semantic solution can keep the information up-to-date with minimum extra effort, and the link between different types of knowledge can be examined by means of automated well-defined procedures.
- They provide a layer of abstraction over KM system services and are able to integrate heterogeneous knowledge resources. Various applications of ontology-based data access and integration are success stories in many industrial use-cases.

4 Healthcare Use-Case

Expert systems in the healthcare domain dates back to the early 70s, when the MYCIN program was developed to support consultation and decision-making. This expert system relied on expert knowledge in form of IF-THEN rules. Creating and encoding these rules was a time-consuming and labor-intensive process [3]. Later, medical terminologies were represented as ontologies along with many AI techniques such as data-mining, text mining, natural language processing etc. A prominent development is the construction of SNOMED CT and the gene ontology. SNOMED CT is a systematic collection of medical terms including codes, synonyms and definitions used in clinical documentation and reporting [18]. It also includes: clinical findings, symptoms, diagnoses, procedures, body structures, organisms and other ethologies, substances, pharmaceuticals, devices and specimens. Such a knowledge model facilitates sharing and aggregation of patient records and findings, the use of standards, access to heterogeneous data, ensures quality screening, facilitates treatment and electronic recording. Figure 4. gives an example of lung disease and its causative agent in SNOMED CT. The example combines is-a and attribute relationships. Such a representation of terms in SNOMED ontology helps users in performing subsumption that is testing pairs of expressions to see whether one is a subtype of the other and vice versa as well as classification that is to structure a set of expressions according to their subsumption relationships.

Existing solutions in the healthcare focus on reference terminology applications where reasoning is often hidden behind the terminological efforts. Though

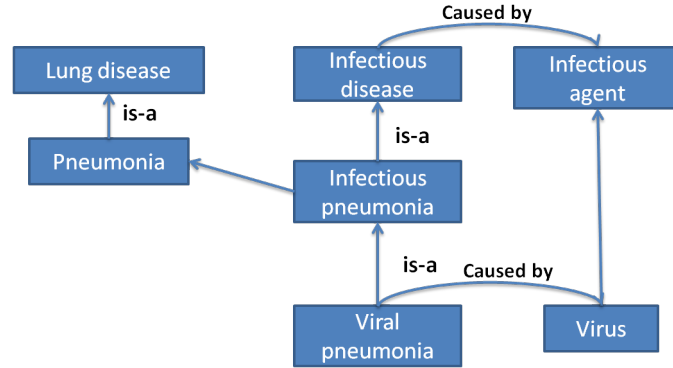


Fig. 4. SNOMED ontology example

systems without semantic models would find difficult to detect and repair inconsistencies in their knowledge repositories. Organizations such as IHTSDO, WHO etc. provide standard terminologies to be used by the industry and thus the healthcare domain is not much affected by knowledge authoring problems.

5 Engineering Use-Case

The engineering domain is well-defined and deals with known concepts and relations. Nevertheless, KM systems in the engineering domain often provide limited capabilities of search, data-access, integration and analytics [19] [20] [21]. The prime focus of many manufacturing and service industries today is to adopt data-driven strategies where they aim to move analytics and decision-support to the data itself. These strategies involve a variety of tasks from data-access, to integration, to storage, analytics, reporting and automation. Thus, a task such as data-access does not live in isolation anymore. Information systems are required to provide workflows with clear semantics to support these strategies and enable autonomous systems. For example, automatic shut-down of a power plant in case safety checks are violated.

It is important to realize the utility of semantic reasoning and the benefits it brings to the engineering applications. Current state-of-the-art applications of data integration, search and interoperability either use manual intervention of experts performing complex Extract-Transform-Load jobs or involve managing large set of configuration files. All these solutions require greater expertise and consume much time and effort. The existing implementations also lack automated reasoning capabilities because most of them are not based on logic-based formalisms. Semantic reasoning can provide better knowledge management services. It promotes reuse of models including wide range of ontologies such as standard sensor network (SSN) ontology ³ to address the domain requirements. Figure 5. shows a snapshot of ontology developed for industrial gas and steam

³ <https://www.w3.org/2005/Incubator/ssn/ssnx/ssn>

turbines. It captures concepts related to compositional structure of the plant, its processes and configurations. This semantic model helps engineers to infer relationships about plant configurations, processes and supports multiple hierarchies to represent part-of and is-a relationships of the related physical entities.

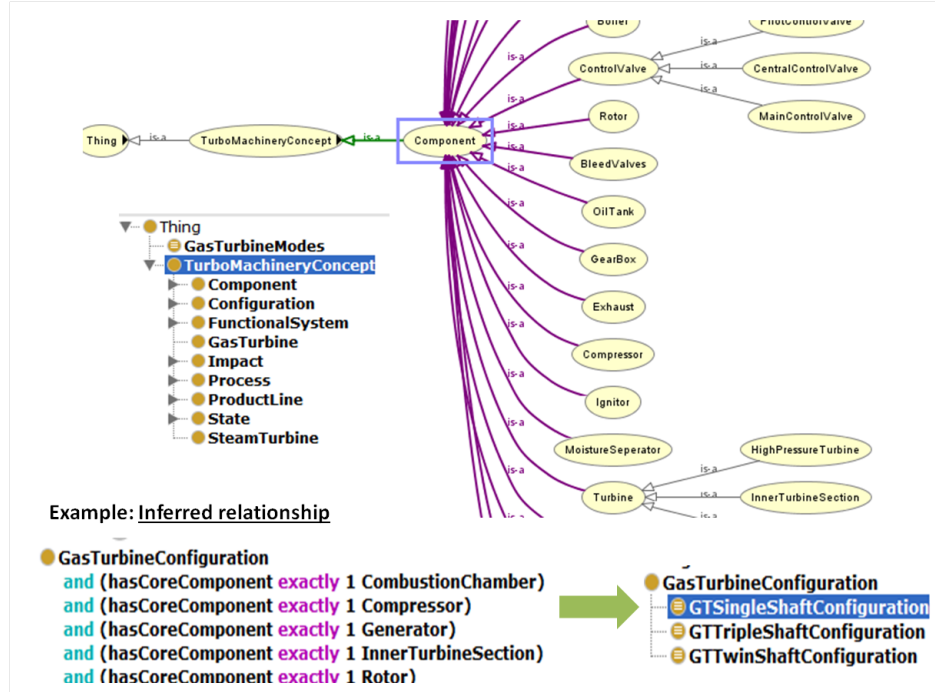


Fig. 5. Power generation - turbine ontology example

6 Challenges of the Semantic Approach

Although early semantic-based KM approaches have shown the benefits of using ontologies to support the KM life cycle, there still exist a large number of challenges from the automation and digitalization of industrial resources. These have to be addressed in order to make semantic technologies fully functional in an industrial setting. In this paper, we aim to answer the following questions:

- How to support analytics and make analytical workflows closer to the data? This new situation demands that semantic paradigm should be able to adjust according to the available data sources and make analytics easier to implement and deploy.

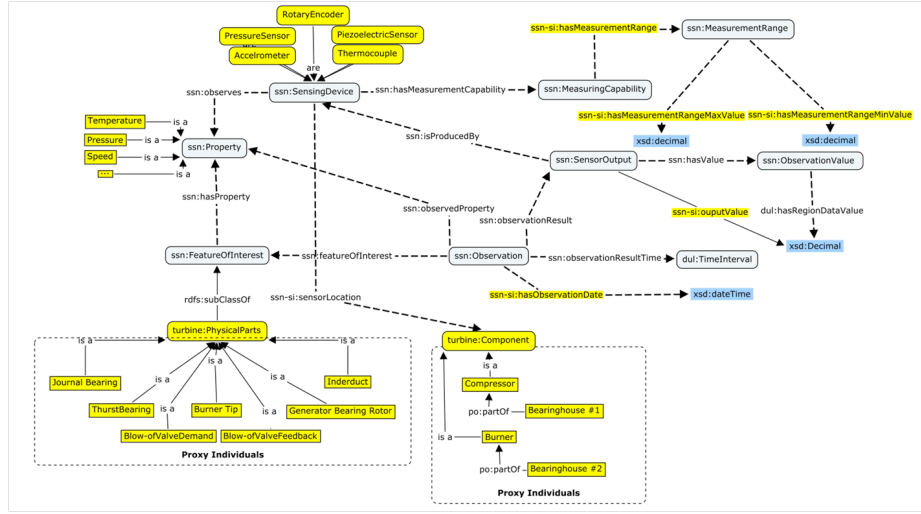


Fig. 6. Extended turbine ontology

- How to remain 'abstract'? This means that semantic-based interfaces should not only support data-access and integration but also help end-users in developing semantic-based analytics and reporting workflows. Thus, an appropriate level of abstraction is required over existing data sources, analytical tools and reporting technologies.
- How to describe analytics outcome? A level of abstraction is required to define the outcomes from analytics and to use the results for decision-support tasks and reporting mechanisms.
- How to cater Big-Data? Semantic-based solution must adhere to the requirements of big-data architectures, where semantic layer can be adopted to manage, reuse and share large-scale data.
- How to manage authoring problems? Industries that have shallow standards and few specialized tools encounter problems with domain modelling and efficient authoring. Thus, semantic solution should be able to support use of reporting mechanism, workflows and design templates to understand, find and display relevant information.

7 Our Proposed Solution

Figure 7. shows the architecture of our proposed solution. It comprises of several components. The Ontology-base data-access (OBDA) middleware provides an abstraction over the existing data sources including interfaces to very large data sets such as hadoop clusters or Spark. The OBDA middleware allow users to formulate their information requirement (i.e. queries) without any knowledge about the data model or source and retrieve the relevant data automatically. A

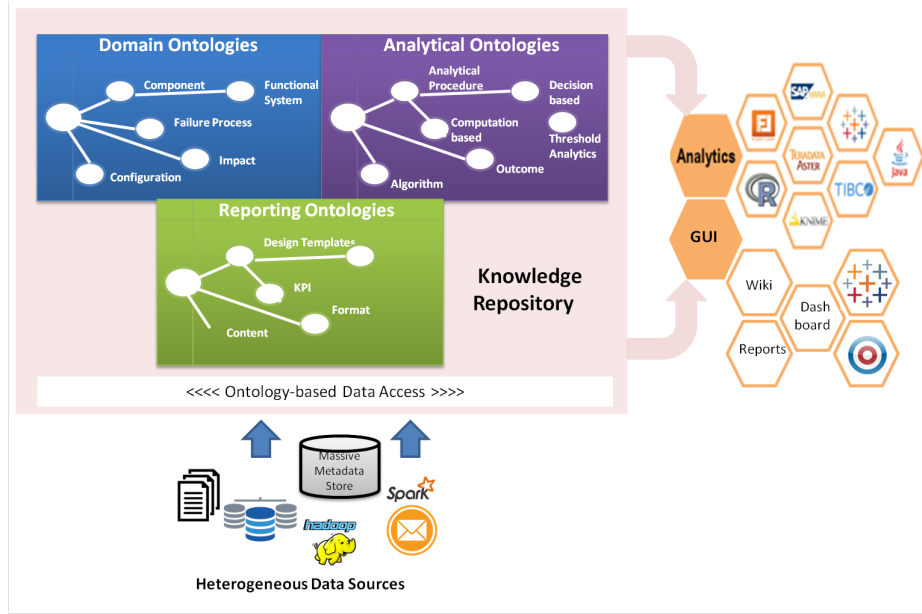


Fig. 7. New approach for semantic-based knowledge management

set of mapping is maintained which describes the relationship between the terms in the ontology and the corresponding terminology in the data source specifications, e.g. table and column names in the relational database schemas.

The domain ontology is extended to include concepts related to system configurations, failure processes, causality and more. Furthermore, the extension to analytics and reporting ontology is an important component of our solution. Knowledge about analytical workflows such as feature set or algorithms to be used and outcomes from analytics can be represented in terms of ontology to support automation and integration of knowledge resources for decision-support. Whereas, reporting ontologies represent different types of reports, content, design templates to be used and more. For example, *"Give me an analysis report X for all machines of type Y that had a 'shut-down' in last three months"*.

Our solution also supports existing analytical and reporting workflows by exposing a Java-based OBDA node with SPARQL endpoint. This sort of integration can be made into any analytics environment (such as in KNIME or R analytics) and leverage the existing analytical models with semantic interfaces.

7.1 Preliminary Results from Siemens Turbo-machinery Use-Case

We setup our solution for Siemens Turbo-machinery use-case where users analyse key performances indicators (KPI) of different gas turbines and compare the reliability, availability and outages as per different parameters (such as product,

service region etc.). The difficulty today is that various machines have different sensors that contribute towards performance, their location within the partonomy may also vary and sensor tags are unknown. Today, users enumerate all sensor tags by hand and formulate customized KPI rules for each machine individually because they have different sensor type and tags along with different threshold values. With our solution, we overcome these problems by using ontology, OBDA mappings and exposing all this to analytical tool.

Figure 6. shows SSN ontology together with our extended turbine ontology to exemplify the idea of capturing part of relations between unknown individuals, sensors, measurable, and sensor meta-data including measurement capabilities. Use of OBDA mappings (shown in Figure 8.) is to connect the sample data set against ontology.

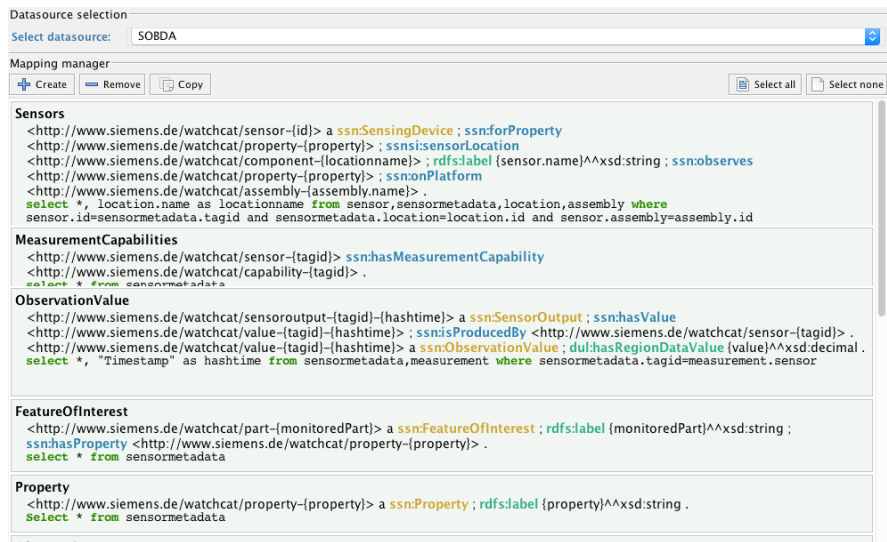


Fig. 8. Semantic mappings

As users were well-equipped with KNIME analytics tool, we provided an integration of our semantic node into this analytics platform. The added-value of our solution is the automation of KPI calculations across machines by using single KPI analytical workflow to determine service hours, period hours and outages.

Figure 9. shows a KNIME⁴ based workflow where the our implemented source node for turbine semantics is exposed to computational procedure of performance analysis.

⁴ <https://www.knime.org/>

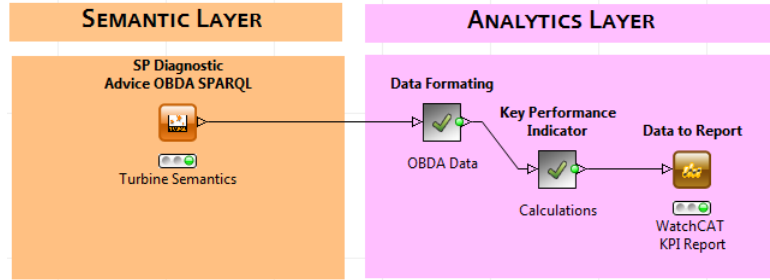


Fig. 9. Semantic node for KNIME-based workflow

Figure 10. shows a detailed view of the SPARQL query that uses terminologies from the domain ontology to extract data and resulting reports are made available to the user. Figure 11 shows snapshots of such visualizations.

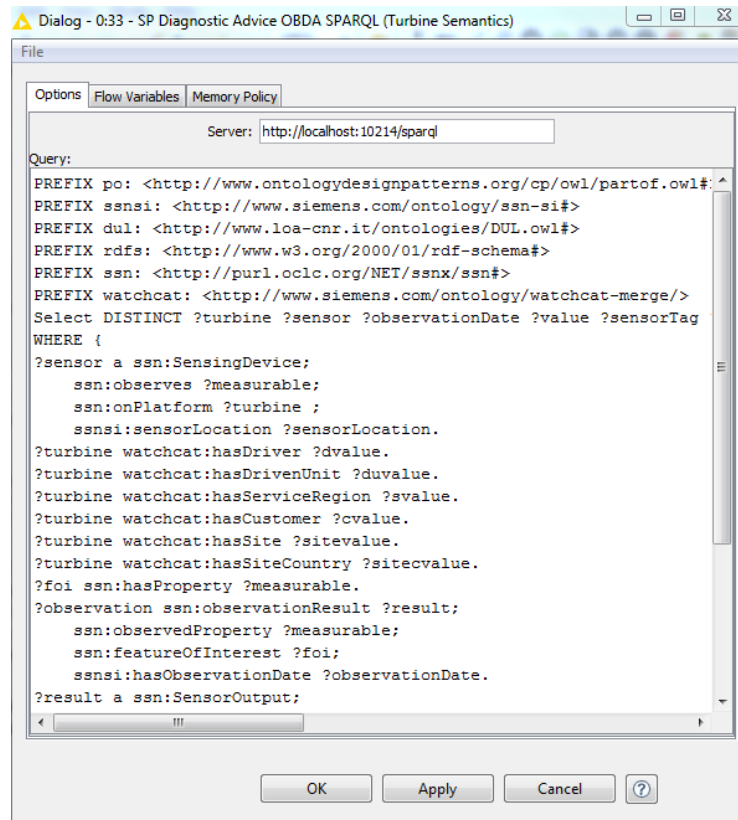


Fig. 10. SPARQL query to access turbine data

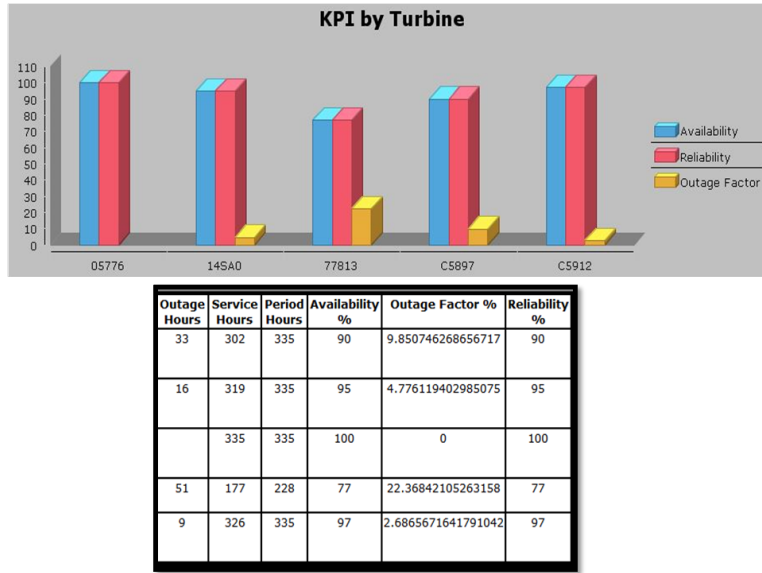


Fig. 11. Results from KPI report

8 Conclusion

In order to make effective use of knowledge, it needs to be classified, defined and related in conventional terminologies. In this paper, we have discussed the basic requirements and limitations of knowledge management systems along with existing semantic and non-semantic approaches to knowledge acquisition, representation, modelling, discovery and distribution. We have presented results from the life sciences and engineering where ontology is used for various application tasks, together with an analysis of the feature set that semantic reasoning brings to the domain. Furthermore, we have described the current challenges for semantic approach and proposed a new solution to solve for big data-access, analytics, reporting and automation. Preliminary results have been presented for Siemens Turbo-machinery use-case where we used the OWL-QL language to define components of turbine ontology, and identified axioms involved in entities - and their interactions to analyse turbine performance indicators. The semantic node was integrated into analytics platform KNIME to automate the workflow and make knowledge discovery tasks easy to follow.

References

1. Jones, J. L. (2006). Robots at the tipping point: the road to iRobot Roomba. IEEE Robotics & Automation Magazine, 13(1), 76-78.
2. Carroll, J. M., & Rosson, M. B. (1992). Getting around the task-artifact cycle: how to make claims and design by scenario. ACM Transactions on Information Systems (TOIS), 10(2), 181-212.

3. Chen, H., Fuller, S. S., Friedman, C., & Hersh, W. (Eds.). (2006). Medical informatics: knowledge management and data mining in biomedicine (Vol. 8). Springer Science & Business Media.
4. Sajeve, S. (2015). Critical analysis of knowledge management maturity models and their components. *ECONOMICS AND MANAGEMENT*, (14), 611-623.
5. Edwards, J. S. (2015). Knowledge Management Concepts and Models. In *Advances in Knowledge Management* (pp. 25-44). Springer International Publishing.
6. Smith, B., & Welty, C. (2001, October). Ontology: Towards a new synthesis. In *Formal Ontology in Information Systems* (Vol. 10, No. 3, pp. 3-9). ACM Press, USA, pp. iii-x.
7. Liebowitz, J. (2001). Knowledge management and its link to artificial intelligence. *Expert systems with applications*, 20(1), 1-6.
8. Russell, S., Norvig, P., & Intelligence, A. (1995). A modern approach. *Artificial Intelligence*. Prentice-Hall, Englewood Cliffs, 25, 27. Chicago.
9. Delone, W. H., & McLean, E. R. (2003). The DeLone and McLean model of information systems success: a ten-year update. *Journal of management information systems*, 19(4), 9-30.
10. Alavi, M., & Leidner, D. E. (2001). Review: Knowledge management and knowledge management systems: Conceptual foundations and research issues. *MIS quarterly*, 107-136.
11. Davenport, T. H., De Long, D. W., & Beers, M. C. (1998). Successful knowledge management projects. *Sloan management review*, 39(2), 43.
12. Chircu, A. M., & Kauffman, R. J. (2000). Limits to value in electronic commerce-related IT investments. *Journal of Management Information Systems*, 17(2), 59-80.
13. Damodaran, L., & Olphert, W. (2000). Barriers and facilitators to the use of knowledge management systems. *Behaviour & Information Technology*, 19(6), 405-413.
14. Evgeniou, T., & Cartwright, P. (2005). Barriers to information management. *European Management Journal*, 23(3), 293-299. Chicago.
15. Benbya, H., Passiante, G., & Belbaly, N. A. (2004). Corporate portal: a tool for knowledge management synchronization. *International Journal of Information Management*, 24(3), 201-220.
16. Joo, J., & Lee, S. M. (2009). Adoption of the Semantic Web for overcoming technical limitations of knowledge management systems. *Expert Systems with Applications*, 36(3), 7318-7327.
17. Davies, J., Fensel, D., & Van Harmelen, F. (Eds.). (2003). *Towards the semantic web: ontology-driven knowledge management*. John Wiley & Sons.
18. Donnelly, K. (2006). SNOMED-CT: The advanced terminology and coding system for eHealth. *Studies in health technology and informatics*, 121, 279.
19. Kharlamov, E., Jimnez-Ruiz, E., Zheleznyakov, D., Bilidas, D., Giese, M., Haase, P., ... & Rodriguez-Muro, M. (2013, May). Optique: Towards OBDA systems for industry. In *Extended Semantic Web Conference* (pp. 125-140). Springer Berlin Heidelberg.
20. Mehdi, G., Brandt, S., Roshchin, M., & Runkler, T. Semantic Framework for Industrial Analytics and Diagnostics.
21. Kharlamov, E., Grau, B. C., Jimnez-Ruiz, E., Lamparter, S., Mehdi, G., Ringsquandl, M., ... & Horrocks, I. (2016, October). Capturing Industrial Information Models with Ontologies and Constraints. In *International Semantic Web Conference* (pp. 325-343). Springer International Publishing.
22. Uschold, M., & Gruninger, M. (1996). Ontologies: Principles, methods and applications. *The knowledge engineering review*, 11(02), 93-136.