Query-based Industrial Analytics over Knowledge Graphs with Ontology Reshaping

Zhuoxun Zheng[1,2], Baifan Zhou[3], Dongzhuoran Zhou[1,3], Gong Cheng[4], Ernesto Jiménez-Ruiz[5,3], Ahmet Soylu[2], and Evgeny Kharlamov[1,3]

[1] Bosch Center for Artificial Intelligence, Renningen, Germany
[2] Department of Computer Science, Oslo Metropolitan University, Oslo, Norway
[3] SIRIUS Centre, University of Oslo, Oslo, Norway
[4] State Key Laboratory for Novel Software Technology, Nanjing University, China
[5] Department of Computer Science, City, University of London, London, UK

Abstract. Industrial analytics that includes among others equipment diagnosis and anomaly detection heavily relies on integration of heterogeneous production data. Knowledge Graphs (KGs) as the data format and ontologies as the unified data schemata are a prominent solution that offers high quality data integration and a convenient and standardised way to exchange data and to layer analytical applications over it. However, poor design of ontologies of high degree of mismatch between them and industrial data naturally lead to KGs of low quality that impede the adoption and scalability of industrial analytics. Indeed, such KGs substantially increase the training time of writing queries for users, consume high volume of storage for redundant information, and are hard to maintain and update. To address this problem we propose an ontology reshaping approach to transform ontologies into KG schemata that better reflect the underlying data and thus help to construct better KGs. In this poster we present a preliminary discussion of our on-going research, evaluate our approach with a rich set of SPARQL queries on real-world industry data at Bosch and discuss our findings.

1 Introduction

Industrial analytics includes among others equipment diagnosis and anomaly detection [1]. It helps to reduce the downtime of manufacturing equipment, resource consumption, error rates, etc. and aims at enhancing the overall production value-chain which is one of the key goals of Industry 4.0 [2, 3]. Industrial analytics heavily relies on integration of heterogeneous production data. Knowledge Graphs (KGs) as the data format for integration and ontologies as the unified data schemata are a prominent solution that offers not only a high quality data integration [4] but also a convenient and standardised way to exchange data and to layer analytical applications over it [5, 6].

However, poor design of ontologies or high degree of mismatch between them and industrial data, e.g., when the ontology is designed to reflect the general domain of knowledge or exported, rather than data particularities [7] naturally lead to KGs of low quality that impede the adoption and scalability of industrial analytics. Indeed, such KGs often have deep structure or have many blank nodes thus they are sparse and consume a high volume of storage for redundant information, and are hard to maintain and
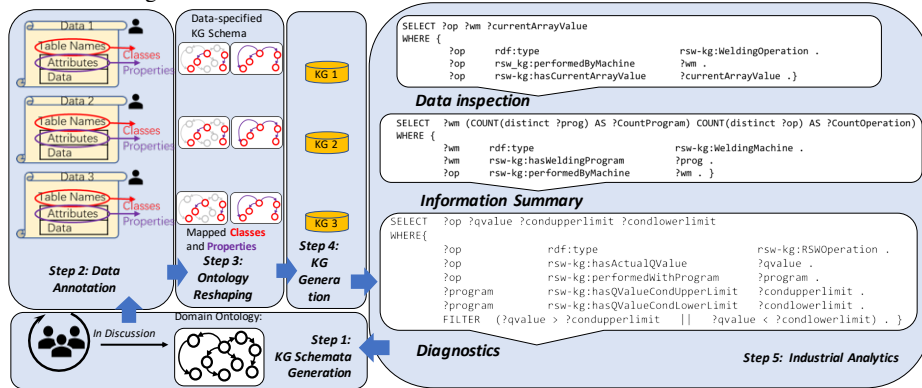
Fig. 1. Our workflow of enhancing industrial analytics with ontology reshaping ($\mathcal{OR}$)

update. Moreover, accessing such data requires long and cumbersome SPARQL queries that are counterintuitive and this substantially increases the training time of users such as engineers [8] in writing such queries.

In order to address this problem, we propose an ontology reshaping approach to transform ontologies into KG schemata that better reflect the given underlying industrial data and thus help to construct better KGs. In this poster, we present a preliminary discussion of our on-going research, evaluate our approach with a rich set of SPARQL queries on real-world industry data at Bosch and discuss the our findings.

2   Our Approach

We now describe our approach of enhancing KG generation with $\mathcal{OR}$ that consists of five steps and summarised in Fig. 1.

KG Schemata Generation (Step 1, left-bottom part of Fig. 1). In this step, we adopt the approach of generating KG schemata by extending domain ontologies from upper level ontologies [9, 10]. The users, i.e. domain experts, participate heavily in ontology extension. They have intensive discussions and create good quality ontologies that reflect the general domain knowledge.

Data Annotation (Step 2, Fig. 1) In this step experts annotate heterogeneous industrial raw data with diverse formats and structures collected from production with unified terms in the domain ontology created in Step 1. Here we consider the raw data in the format of relational tables. The table names are mapped to classes in the ontology and attribute names to properties. In some complex cases, attribute names with endings like "ID" or "NAME" are elevated to classes according to users' inputs.

Ontology Reshaping (Step 3, Fig. 1) In this step our ontology reshaping algorithm $\mathcal{OR}$ takes a knowledge-oriented domain ontology, a raw dataset, a mapping between them, and some optional user heuristics as the input and outputs a data-orientated ontology, which serves as the KG schemata. The resulting ontology is a (compact) version of the original one and essentially consists of (1) all corresponding information (table names and attributes) from the raw data and (2) other essential connecting elements, which are partially from the original ontology and partially from users, to attain some optimality defined by user heuristics, efficiency, simplicity, etc. We adopt our $\mathcal{OR}$ algorithm [11]. In the nutshell $\mathcal{OR}$ firstly selects a subset of nodes and edges in the domain

ontology [12], creating its sub-graph, which is a sparse sub-graph that consists of many disconnected small graph fragments; secondly, it extends the sub-graph to a KG schema with the help of two sources of information: (a) retaining other nodes and edges of the domain ontology to preserve part of its knowledge, (b) some optional information given by users (welding experts) that help to connect the fragments in the sub-graph.

KG Generation (Step 4, Fig. 1) In this step the KG schemata are populated in the ETL fashion with the actual data based on the annotated table names and attributes in the relational table to generate the KGs. The generated KGs are data-oriented and (often) more compact.

Industrial Analytics (Step 5, Fig. 1) In this step, we layer industrial analytics upon the generated KG. In this paper, we consider three types of SPARQL queries written by engineers or generated by users' keywords inputs. The queries account for retrieving or summarising specific information of interest from huge datasets that come from running production lines, and to perform basic diagnostics over them. The details follow in the next section. During the application, the user feedback is constantly collected and the workflow can go back to Step 1 and the whole process restarts. For example, after the application stage, users may realise they can scale the system to more tasks and thus go back to extend the domain ontology for more tasks.

Related Work. The most related work to ours are the one on ontology modularisation [13], summarisation [14, 15], and summarisation forgetting. They focus on selecting subsets of ontologies, but they do not address the data particularity issue. We thus propose to rely on ontology reshaping ($\mathcal{OR}$), which transforms domain ontologies to its often compacter versions that reflect more data specificities. The resulting KGs will contain no blank-nodes and become less deep, and thus the queries will become less deep and more user-friendly.

3 Evaluation with Industrial Dataset

We now present evaluation of our approach with a real industrial dataset.

Data Description. The dataset D is collected from a German factory, in which reside production lines that consist of 27 welding machines of an impactful automated welding process widely applied in automotive industry: the resistance spot welding [16, 17]. D contains a high number of welding operation records and a series of welding sensor measurements. These data account for 1000 welding operations, estimated to be related to 100 cars. In this work, we select a section of D for discussion, which contains 4.315 million records and 176 attributes. The knowledge-oriented domain ontology O is generated by welding experts and contains 206 classes, 203 object properties, and 191 datatype properties. The mapping maps all 176 attributes in D to classes in O.

Query Description. We consider SPARQL queries of three types as follows: *Type I: Data inspection*, where the experts need to inspect these data for generating a first handful of insights. The desired answer to such query is a listing of some attributes. An example see Step 5 in Fig. 1, which will return all arrays of currents with the corresponding operation-names and machine-IDs. *Type II: Information summary*, where the welding experts need to gain overview information of arbitrarily selected datasets, e.g.

Table 1. Our approach enhanced by ontology reshaping (Onto-Reshape) outperforms the baseline significantly in terms of query simplicity, avg.: average, max: maximum

	Subset	Set 1	Set 2	Set 3	Set 4	Set 5	Set 6
raw data	#attributes	20	40	60	80	100	120
Baseline	avg. query depth	4.2	4.4	4.3	4.3	4.2	4.3
	max. query depth	5.0	5.0	5.0	5.0	5.0	5.0
Onto Reshape	avg. query depth	2.3	2.5	2.6	2.7	3.0	3.1
	max. query depth	3.0	4.0	3.0	4.0	3.0	4.0

how many different programs does every welding machine perform (Step 5 in Fig. 1)? *Type III: Diagnostics*, where the welding experts need to perform various diagnostic tasks, such as detecting abnormal machines, operations, etc. Moreover, the users would also like to find the surroundings of the abnormalities, to figure out what happened near the abnormalities, so that they can better understand for root-causes. One example for this kind of queries would be: Where are the abnormal welding operations whose quality indicators exceed the conditional tolerance limit? (Step 5 in Fig. 1)

Experiment Design. To test our approach, we randomly sub-sample D to 6 sub-datasets (Set 1-6 in Table 1). Each set contains a subset of the attributes of D, reflecting different data complexity. The numbers of attributes in the subsets increase by twenty each time, from 20 to 120. We repeat the sub-sampling for each subset 10 times to decrease the randomness. We compare our approach with a *baseline* of KG generation without ontology reshaping, which is a naive approach to use the domain ontology directly as the KG schema. This work considers 729 queries and 324, 189, 216 queries for query Type I, II, III, respectively. The evaluation metrics are set as the average and maximal query depth, which characterises the number of edges to connect two nodes in the query graph via the shortest path (the query is essentially also a graph with variable nodes).

Results and Discussion. The results (Table 1) show that for retrieving the same answers, the queries are significantly simplified with our approach: the query depths are reduced by about 2 for both its average and maximum. This indicates the generated KG becomes more practical, because shorter queries are needed to get the same information. In addition, we also observe that the KGs after Onto-Reshape become more efficient and simpler: their generation becomes 7 to 8 times faster, the number of entities are reduced to 1/2 to 1/6 of the baseline, storage space to 2/3, and the number of blank nodes to zero.

4   Conclusion and Outlook

In this paper, we present our ongoing research of knowledge graph-based industrial query analytics at Bosch. We have preliminary results that show the benefit of our approach for industrial analytics. Our work falls into the big picture of KG-based industrial applications [18]. As future work we will investigate transforming queries across different KG schemata, and study more query properties.

References

1. M. H. ur Rehman, et al., The role of big data analytics in industrial Internet of Things, Future Generation Computer Systems 99 (2019) 247–259.
2. H. Kagermann, Change through digitization – value creation in the age of Industry 4.0, in: Management of Permanent Change.
3. B. Zhou, T. Pychynski, M. Reischl, R. Mikut, Comparison of machine learning approaches for time-series-based quality monitoring of resistance spot welding (RSW), Archives of Data Science, Series A (Online First) 5 (1) (2018) 13.
4. I. Horrocks, M. Giese, E. Kharlamov, A. Waaler, Using semantic technology to tame the data variety challenge, IEEE Internet Comput. 20 (6) (2016) 62–66.
5. D. Zhou, B. Zhou, Z. Zheng, E. V. Kostylev, G. Cheng, E. Jimenez-Ruiz, A. Soylu, E. Kharlamov, Enhancing knowledge graph generation with ontology reshaping – Bosch case, in: ESWC (Demos/Industry), Springer, 2022.
6. X. Zou, A survey on application of knowledge graph, in: Journal of Physics: Conference Series, Vol. 1487, IOP Publishing, 2020, p. 012016.
7. M. Yahya, B. Zhou, Z. Zheng, D. Zhou, J. G. Breslin, M. I. Ali, E. Kharlamov, Towards generalized welding ontology in line with ISO and knowledge graph construction, in: ESWC (Posters & Demos), Springer, 2022.
8. A. Soylu, E. Kharlamov, D. Zheleznyakov, E. Jiménez-Ruiz, M. Giese, M. G. Skjæveland, D. Hovland, R. Schlatte, S. Brandt, H. Lie, I. Horrocks, OptiqueVQS: A visual query system over ontologies for industry, Semantic Web 9 (5) (2018) 627–660.
9. B. Zhou, Y. Svetashova, A. Gusmao, A. Soylu, G. Cheng, R. Mikut, A. Waaler, E. Kharlamov, SemML: Facilitating development of ML models for condition monitoring with semantics, Journal of Web Semantics 71 (2021) 100664.
10. Y. Svetashova, B. Zhou, T. Pychynski, S. Schmidt, Y. Sure-Vetter, R. Mikut, E. Kharlamov, Ontology-enhanced machine learning: A Bosch use case of welding quality monitoring, in: ISWC, Vol. 12507, Springer, 2020, pp. 531–550.
11. D. Zhou, B. Zhou, J. Chen, G. Cheng, E. V. Kostylev, E. Kharlamov, Towards ontology reshaping for KG generation with user-in-the-loop: Applied to Bosch welding, in: IJCKG, 2021.
12. Y. Svetashova, B. Zhou, S. Schmid, T. Pychynski, E. Kharlamov, SemML: Reusable ML for condition monitoring in discrete manufacturing, in: ISWC (Demos/Industry), Vol. 2721, 2020, pp. 213–218.
13. P. Doran, Ontology reuse via ontology modularisation, in: KnowledgeWeb PhD Symposium, Vol. 2006, Citeseer, 2006.
14. X. Zhang, G. Cheng, Y. Qu, Ontology summarization based on RDF sentence graph, in: WWW 2007, 2007, pp. 707–716.
15. S. Pouriyeh, M. Allahyari, Q. Liu, G. Cheng, H. R. Arabnia, M. Atzori, F. G. Mohammadi, K. J. Kochut, Ontology summarization: Graph-based methods and beyond, Int. J. Semantic Comput. 13 (2) (2019) 259–283.
16. B. Zhou, T. Pychynski, M. Reischl, E. Kharlamov, R. Mikut, Machine learning with domain knowledge for predictive quality monitoring in resistance spot welding, J. Intell. Manuf. 33 (4) (2022) 1139–1163.
17. B. Zhou, D. Zhou, J. Chen, Y. Svetashova, G. Cheng, E. Kharlamov, Scaling usability of ML analytics with knowledge graphs: Exemplified with A Bosch welding case, in: IJCKG, ACM, 2021, pp. 54–63.
18. B. Zhou, Z. Zheng, D. Zhou, E. Jimenez-Ruiz, G. Cheng, T. Tran, D. Stepanova, M. H. Gad-Elrab, N. Nikolov, A. Soylu, E. Kharlamov, The data value quest: A holistic semantic approach at Bosch, in: ESWC (Demos/Industry), Springer, 2022.