Anomaly Detection and Diagnosis of Vehicle Steering Systems Using a Knowledge Graph-based Approach

Qiushi Cao[1,*], Patelis Alexandros[2] and Irlán Grangel-González[2]

[1]Corporate Research, Bosch (China) Investment Ltd., Shanghai, China
[2]Corporate Research, Robert Bosch GmbH, Renningen, Germany

Abstract

In the automotive industry, data-driven techniques have been widely used for the early detection of vehicle field issues. During data analysis processes, data heterogeneity is a crucial pain point that causes huge manual effort and time delay. To mitigate this pain point, in this paper we demonstrate a knowledge graph-based approach for the early detection of vehicle technical issues for automotive steering systems. The proposed approach enables semantic data integration, by which the efficiency of data Extract, Transform, Load (ETL) process is significantly improved. Based on the developed knowledge graph system, data visualization dashboards and Large Language Model (LLM) solutions can be easily developed to gain insights for failure-cause-effect analysis. The proposed knowledge graph-based approach has gained significant efficiency improvement. The time and manual efforts for data ETL and integration have been reduced up to 70%. The use of standardized domain ontologies enables the re-usability of the proposed approach for other use cases or products. Our work highlights the importance of industrial knowledge graphs for tackling data heterogeneity and data quality issues when developing data-driven applications.

Keywords

Knowledge Graph, Semantic Data Integration, Data Quality, Automotive Industry, Large Language Model

1. Introduction

Every day, millions of vehicles are driven on the road. Although most of them are able to assist drivers and passengers reach their destinations, some of them may face technical issues. When faced with a technical issue, vehicles are driven to service stations for repair services. When Original Equipment Manufacturers (OEM) try to fix unexpected technical field issues for the first time, all information regarding this problem and symptoms (from both vehicle and defect components) is collected and stored for further use. As more similar technical problems are encountered, OEMs start to investigate the source of the accumulating technical issues. As time goes by, when a sufficient number of similar issues are detected and reported, OEMs may decide to escalate the problem to automotive suppliers like Bosch. Bosch may also conduct investigation for the problem and trigger countermeasures such as root cause analysis. During this process, heterogeneous data is collected from various data sources, which requires a huge amount of human effort for data analysis. This leads to a significant time delay caused by data Extract, Transform, Load (ETL).

During this data ETL process, Semantic Interoperability Conflicts (SIC) [1] is a big challenge that causes data quality issues. This highlights the need for data standardization methods. Among the existing technologies for data standardization, ontologies appear to be a promising solution [2, 3]. Defined as “a specification of a conceptualization” [4], ontologies can promote the exchange and sharing of data and domain knowledge for the purpose of supporting both data-driven and knowledge-driven tasks [5, 6].

To tackle the aforementioned challenges regarding data ETL and data SIC, we propose a knowledge graph-based approach for managing and harmonizing heterogeneous data related to field issue detection. The proposed approach starts with collecting and pre-processing heterogeneous data sources regarding vehicle field, production, and auxiliary data, followed by the semantic data integration process enabled

ISWC 2025 Companion Volume, November 2–6, 2025, Nara, Japan
qiushi.cao@cn.bosch.com (Q. Cao); Alexander.Patelis@de.bosch.com (P. Alexandros);
irlan.grangelgonzalez@de.bosch.com (I. Grangel-González)

by data-ontology mappings. The data is integrated and ingested into a knowledge graph system that provides a unified view of field issues. Data traceability is also enabled by traversing the connected knowledge graphs that connect heterogeneous data sources. The constructed graph database also allows the development of data dashboards for answering complex failure-cause-effect questions. Exploratory charts, filters, and tables have been created by running specialized and customized queries that traverse the complex knowledge graph for retrieving important information.

2. Proposed Approach

We propose a three-layered knowledge graph system for semantic data integration and data analytics. Fig. 1 shows the overall architecture of the proposed system. We introduce the system architecture in details:

1. **Data Collection/Preparation Module** is responsible for data source management and data pre-processing. In this work, data is collected from various data sources hosted in different database systems, Vehicle Feedback DataBase Manuals, and CSV files. To handle these data for future exploitation, data cleaning and pre-processing capabilities are provided by this module for effective data ETL and integration. In this work, there are three main data sources:
2. **Semantic Data Integration Module**, which enables data access and integration via SPARQL queries. HTTP/REST endpoints are provided to access data sources from relational databases. The accessed data is ingested into the knowledge graph system by creating mappings that link data attributes to ontology classes. The mappings are key for semantic data integration, by which data is ingested to a single connected knowledge graph. This module consists of four sub-modules:
    - **Data Ingestion & Mapping:** data-ontology mappings are created for data ingestion. The ingested data together with domain ontologies form different knowledge graphs.
    - **Domain Ontology:** to support semantic data integration, domain ontologies are developed in this system module. These ontologies provide formal and high-level representations of the domain. In this work, we develop ontologies for describing core concepts and relationships regarding vehicles, diagnostic tasks, customer claims, vehicle production, and statistic calculations. These semantic models are developed based on the on the semantics defined in the Industry 4.0 Core knowledge graph developed at Bosch [7]. They are connected together to form the conceptual layer of the knowledge graph system.
    - **Knowledge Graph:** a unified knowledge graph is constructed by graph materialization. RDF triples are generated in a graph structure, under associate classes in semantic models. The generation of RDF triples is enabled by mappings written in SPARQL query language. In this way, the semantic models together with materialized RDF data form the whole knowledge graph, which provides access points to higher-level applications.
    - **Access Points:** the materialized knowledge graph provides access to actual data for end-users and software agents. Users and developers can query the underlying RDF data via SPARQL endpoints in the knowledge graph system. This allows the answering of key questions related to the early detection of vehicle field issues. The access points of the system also allow the deployment of Application Programming Interfaces, which can be used to simplify the process of software development (based on the developed knowledge graph system).
3. **Applications Module:** this module is built on top of the Semantic Data Integration Module. Data visualization features are provided in Applications Module to explore, make sense of, and show connections between data. Different data dashboards are developed for a field, production, and failure-related data. These dashboards display not only important statistics but also highlight data linkage across multiple data sources. The Applications Module also enables the deployment of Graph AI and LLM solutions. Leveraging the semantically-integrated data, AI algorithms will further improve the efficiency and automation level of the whole system.

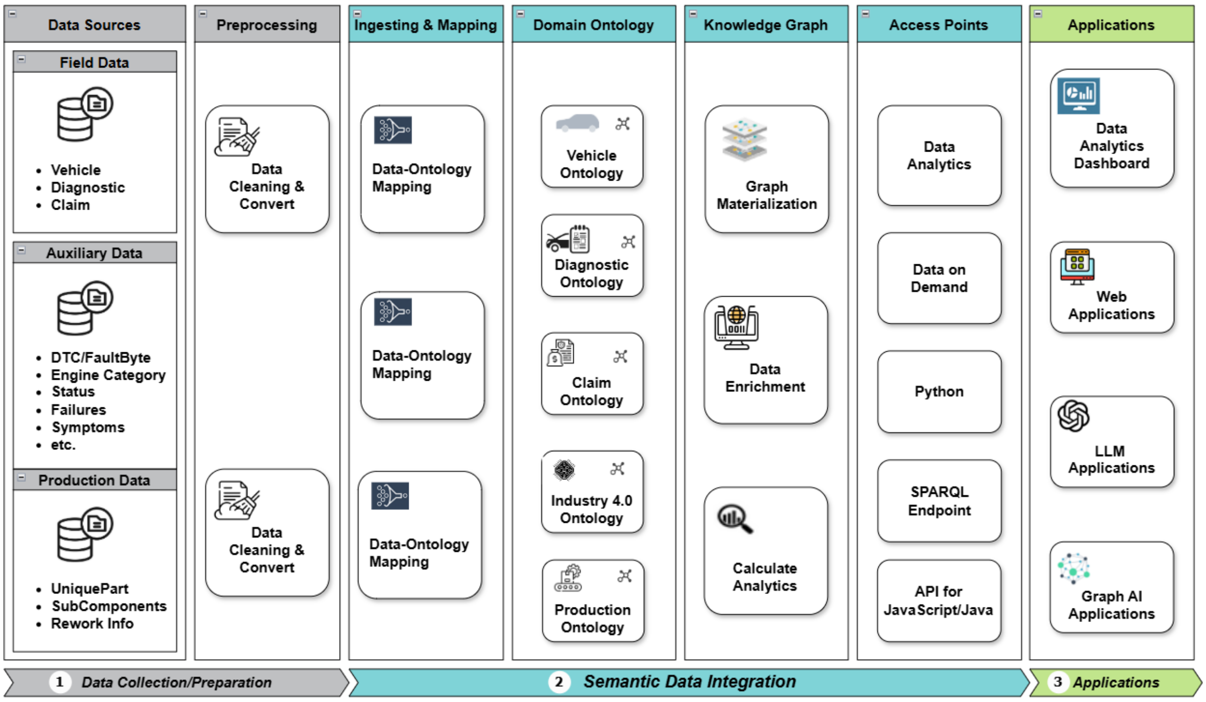

Figure 1: The overall architecture of the knowledge graph-based approach. The system consists of three main system components: i) Data Collection/Preparation Module, used for collecting data from heterogeneous sources. Vehicle field data, auxiliary data (information about failures), and production data are the three main data sources; ii) Semantic Data Integration Module, developed for integrating multiple data sources into a uniform knowledge graph. Raw data is mapped and ingested into semantic models (ontologies), which bring rich semantics and metadata into data sources; iii) Applications Module, which enables the development of data visualization solutions for root cause analysis purposes. The users can traverse the whole knowledge graph to search for answers to critical failure-cause-effect-related questions. Our solution also enables the use of LLM and Graph AI technologies for deriving new insights from graph data.

3. Results and Business Values

Currently, the developed knowledge graph system maintains data on about 1,750,000 vehicles. Data for more than 8,000,000 diagnostic sessions have been ingested and connected to vehicle data. For the ingestion of diagnostic data, more than 17,000,000 Diagnostic Trouble Code (DTC) occurrences have been recorded, and DTC-related information has been encoded in the knowledge graph system. Based on the knowledge graph system, three types of data visualization dashboards have been developed: i) Early Detection Dashboard for giving an overview of DTC Fault Byte occurrences on both monthly and weekly level; ii) DTC Overview Dashboard for displaying DTC codes, DTC code description, failure symptoms, failure criticality, and information about potential failure causes; iii) Early Warning Dashboard for showing the trend of DTC occurrence (according to time) and the link between DTC data with vehicle data.

The proposed knowledge graph-based approach has gained significant efficiency improvement. The time and manual efforts for data ETL and integration have been reduced up to 70%. The use of standardized domain ontologies enables the re-usability of the proposed approach for other use cases or products. The connected knowledge graph serves as an enabler for other solutions that demand integrated data silos for generating business insights, such as AI solutions including Graph Neural Networks and LLM (e.g. Graph Retrieval-Augmented Generation) applications. This research highlights the importance of industrial knowledge graphs for tackling data heterogeneity and data quality issues when developing data-driven applications.

Declaration on Generative AI

During the preparation of this work, the authors used Grammarly to help rephrase sentences or paragraphs to improve clarity and conciseness. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the publication's content.

References

[1] N. Melluso, I. Grangel-González, G. Fantoni, Enhancing industry 4.0 standards interoperability via knowledge graphs with natural language processing, Computers in Industry 140 (2022) 103676.

[2] Y. Chen, C. Yu, X. Liu, T. Xi, G. Xu, Y. Sun, F. Zhu, B. Shen, Pclion: An ontology for data standardization and sharing of prostate cancer associated lifestyles, International Journal of Medical Informatics 145 (2021) 104332.

[3] E. M. Sanfilippo, S. Borgo, What are features? an ontology-based review of the literature, Computer-Aided Design 80 (2016) 9–18.

[4] N. Guarino, P. Giaretta, Ontologies and knowledge bases, Towards very large knowledge bases (1995) 1–2.

[5] S. Geisler, M.-E. Vidal, C. Cappiello, B. F. Lóscio, A. Gal, M. Jarke, M. Lenzerini, P. Missier, B. Otto, E. Paja, et al., Knowledge-driven data ecosystems toward data transparency, ACM Journal of Data and Information Quality (JDIQ) 14 (2021) 1–12.

[6] Q. Cao, C. Zanni-Merk, A. Samet, C. Reich, F. D. B. De Beuvron, A. Beckmann, C. Giannetti, Kspmi: a knowledge-based system for predictive maintenance in industry 4.0, Robotics and Computer-Integrated Manufacturing 74 (2022) 102281.

[7] I. Grangel-González, F. Lösch, A. ul Mehdi, Knowledge graphs for efficient integration and access of manufacturing data, in: 2020 25th IEEE International Conference on Emerging Technologies and Factory Automation (ETFA), volume 1, IEEE, 2020, pp. 93–100.