Knowledge Graph Construction towards a Graph RAG-Enhanced Intelligent Maintenance Chatbot

Hansi Zhang[1,*,†], Wilma Johanna Schmidt[2,3,*,†], Xiaozhi Shen[4], Qiushi Cao[1], Sebastian Monka[2] and Adrian Paschke[3,5,†]

[1] *Robert Bosch GmbH, Corporate Research, Shanghai, China*
[2] *Robert Bosch GmbH, Corporate Research, Renningen, Germany*
[3] *Freie Universität Berlin, AG Semantic Web, Berlin, Germany*
[4] *Robert Bosch GmbH, Vehicle Motion, Suzhou, China*
[5] *Data Analytic Center (DANA), Fraunhofer Institute FOKUS, Berlin, Germany*

Abstract

In the context of Industry 4.0, effective maintenance is critical for minimizing manufacturing downtime and ensuring production reliability. While first Graph Retrieval-Augmented Generation (RAG) frameworks enhance contextual understanding and accuracy in maintenance chatbots, Knowledge Graph (KG) construction in manufacturing remains tedious and error-prone. To address this, we propose a semi-automated KG construction pipeline that integrates rule-based methods, Small Language Models (SLMs), and Large Language Models (LLMs), significantly reducing manual efforts in KG construction. We evaluate the constructed KG in a Graph RAG setting on real-world maintenance scenarios in a production line. Our results highlight the potential to significantly enhance the efficiency and intelligence of manufacturing maintenance workflows. Our work aims to spark discussions on efficient Graph RAG frameworks for maintenance scenarios in manufacturing.

Keywords

Knowledge Graph Construction, Graph RAG, LLM, Manufacturing

1. Introduction

In the context of Industry 4.0 — defined by the integration of advanced technologies such as the Internet of Things, Cyber-Physical Systems, Big Data, cloud computing, and AI into manufacturing and production systems [1] — industrial equipment has become increasingly intelligent and interconnected, generating vast amounts of data through shopfloor machinery, sensors and systems. Despite these advancements, equipment failures can disrupt whole production lines, resulting in downtime and reducing stability and reliability of the production. Maintenance has remained a critical component to mitigate these risks. Engineers and technicians must analyze generated data to identify potential causes and determine appropriate maintenance procedures to minimize downtime. While experienced engineers can leverage technical expertise and past insights to handle machine errors and failures effectively, less experienced personnel often struggle to navigate the complexities of modern maintenance systems. To address this, engineers typically follow standard processes to summarize and document their valuable experiences. These records not only serve as an educational resource for less experienced engineers, helping them build practical problem-solving skills and adapt more quickly to real-world challenges, but also provide experienced engineers with important insights for refining and optimizing maintenance strategies. With the rise of Large Language Models (LLMs), there is an opportunity to transform how maintenance knowledge is accessed, shared, and applied, particularly by improving its organization and utilization. LLMs perform exceptionally well in processing large volumes of unstructured data, providing contextual, accurate, and human-like responses. One motivating example as shown in Fig-

SGKi 2025: Scaling Knowledge Graphs for Industry Workshop, co-located with SEMANTiCS'25: International Conference on Semantic Systems, September 3–5, 2025, Vienna, Austria

[*] Corresponding author.

[†] These authors contributed equally.

hans.zhang@bosch.com (H. Zhang); wilma.schmidt@bosch.com (W. J. Schmidt); xiaozhi.shen@bosch.com (X. Shen); qiushi.cao@bosch.com (Q. Cao); sebastian.monka@bosch.com (S. Monka); adrian.paschke@fokus.fraunhofer.de (A. Paschke)
0009-0001-0872-4578 (W. J. Schmidt); 0000-0003-3156-9040 (A. Paschke)

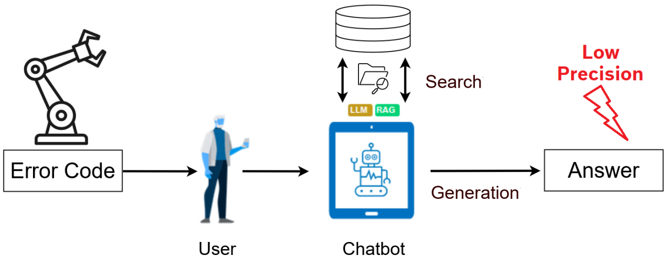

Figure 1: Motivating Scenario of Low Precision in Maintenance Question Answering with RAG Framework.

ure 1 involves our colleagues at Bosch, who leverage the power of Retrieval-Augmented-Generation (RAG) frameworks to enhance maintenance workflows using LLMs. They implement a standard RAG solution to better utilize existing maintenance knowledge. This approach involves splitting maintenance documents — such as slides, PDFs, and emails — into textual chunks, embedding them, and storing them in a vector database. Yet, this approach shows limitations such on both retrieval and generation accuracy, leading to low precision answers.

To overcome these limitations, a promising alternative is Graph RAG, which has recently gained attention [2, 3, 4, 5], also in the manufacturing domain [6]. Unlike traditional RAG, Graph RAG retrieves graph elements that contain relevant knowledge to a given query from an existing Knowledge Graph (KG). It addresses the aforementioned issues by leveraging the KG to extract factual knowledge and represent it through entities (classes and instances) and relationships (we focus on object properties), ensuring high-quality and structured information. Graph RAG considers relationships and interconnections within the data, enabling more accurate and comprehensive information retrieval especially for domain-specific scenarios. Moreover, KGs offer an abstraction and summarization of textual data, effectively reducing input length and alleviating concerns about verbosity.

However, KG construction and updates remain tedious tasks in the manufacturing domain, as both data and technical experts are needed to generate a KG. Hence, we present in our work a semi-automated construction pipeline of a maintenance KG to set the fundamental work for enabling Graph-RAG solutions to support maintenance tasks.

In summary, this paper presents the following contributions: 1) We develop a Manufacturing Maintenance Ontology (MMO) by extending a set of ontologies from Core Information Model for Manufacturing (CIMM) [7] with maintenance domain knowledge; 2) To reduce extensive manual efforts involved in KG construction, we develop a semi-automatic KG construction pipeline that integrates rule-based methods, SLMs, and LLMs; 3) We spark discussion and lay the ground work for Graph RAG frameworks supporting maintenance tasks in the manufacturing domain.

The remainder of the paper is organized as follows. Section 2 describes a semi-automated KG construction pipeline to build a maintenance KG. In Section 3 we analyze metrics and expert feedback on our approach. In Section 4 we discuss the lessons learned. Section 5 presents the related works and Section 6 concludes the paper with an outlook and future work.

2. Knowledge Graph Construction

In this section, we introduce data sources, ontology and our semi-automated KG construction pipeline to generate a maintenance KG. One major difference between Graph RAG and vector-based RAG lies in how data is presented. For use cases requiring high precision in answers, the quality of data representation is critical to success. Therefore, a KG is chosen for representing and storing raw data due to its structured and semantically rich format. However, building a high-quality KG is a resource-intensive process. To overcome this challenge, we propose a pipeline to speed up KG construction with reduced efforts from engineers, combining LLMs, SLMs, and traditional rule-based NLP techniques. Next, we describe the details of constructing a maintenance KG from unstructured data.

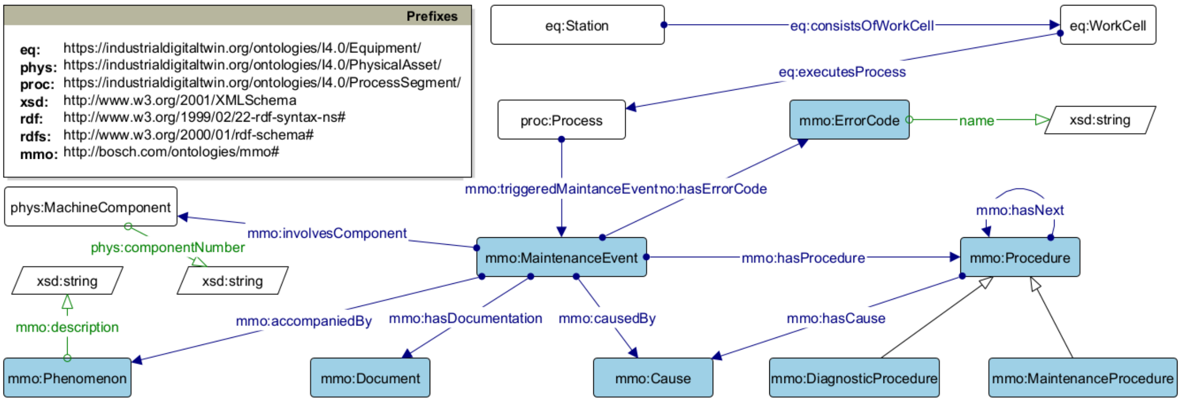
Prefixes
eq: https://industrialdigitaltwin.org/ontologies/I4.0/Equipment/
phys: https://industrialdigitaltwin.org/ontologies/I4.0/PhysicalAsset/
proc: https://industrialdigitaltwin.org/ontologies/I4.0/ProcessSegment/
xsd: http://www.w3.org/2001/XMLSchema
rdf: http://www.w3.org/1999/02/22-rdf-syntax-ns#
rdfs: http://www.w3.org/2000/01/rdf-schema#
mmo: http://bosch.com/ontologies/mmo#
eq:Station
eq:consistsOfWorkCell
eq:WorkCell
eq:executesProcess
proc:Process
mmo:ErrorCode
name
xsd:string
mmo:triggeredMaintanceEventno:hasErrorCode
phys:MachineComponent
mmo:involvesComponent
phys:componentNumber
mmo:hasNext
mmo:MaintenanceEvent
mmo:hasProcedure
mmo:Procedure
xsd:string
xsd:string
mmo:description
mmo:accompaniedBy
mmo:hasDocumentation
mmo:causedBy
mmo:hasCause
mmo:Phenomenon
mmo:Document
mmo:Cause
mmo:DiagnosticProcedure
mmo:MaintenanceProcedure

Figure 2: Main Concepts of the *Manufacturing Maintenance Ontology (MMO).*

2.1. Data Sources

The primary data sources comprise documented experiences and observations summarized after engineers have resolved machine errors. These records exist in various formats, including slides, excels, PDFs, emails, as well as chat histories. Typically, each record documents a specific maintenance event, detailing the complete maintenance workflow, such as observed phenomenon, error codes, diagnostic procedures with potential causes, maintenance procedures, i.e., final solutions with root causes, and summaries. The content is predominantly written in Chinese, with some terminologies in English. For this paper, we focus on *five* major stations within *one* production line, collecting *49* documents that describe various maintenance events. To evaluate our KG construction, engineers provide *two* question-answer (QA) pairs collected from their daily work, which we use to assess our solution.

2.2. Ontology

We develop a *Manufacturing Maintenance Ontology (MMO)* by extending a set of ontologies within CIMM [7] with domain-specific maintenance knowledge. This set of ontologies is published in the context of the Industrial Digital Twin association[1]. *MMO* is designed to represent complex relationships and entities involved in manufacturing maintenance, capturing standard maintenance workflow required for systematically observing, diagnosing, and resolving machine errors. This ontology integrates knowledge from *four* subdomains which we introduce next. Figure 2 illustrates the composition of *MMO* of the *four* interconnected sub-ontologies: 1) *Equipment Ontology* captures the hierarchical structure of manufacturing stations and their associated work cells; 2) *Physical Asset Ontology* represents machine components and their relationships to work cells, offering a detailed view of the physical elements within the manufacturing process; 3) *Process Segment Ontology* defines the processes performed within work cells; 4) *Manufacturing Maintenance Ontology* is designed with an event-driven structure to capture domain-specific maintenance concepts, e.g., error codes, maintenance procedures, observable phenomena, and root causes. It also provides detailed documentation of maintenance events, including production line downtime, timestamps, and other associated information. Further, we define a set of relationships, e.g., `mmo:hasNext`, `mmo:causedBy`, `mmo:hasDocumentation`, to provide a semantic description for linking maintenance events with their causes, procedures, and phenomenon. For instance, each maintenance event is documented in a source file, i.e., a document, which is linked via the `mmo:hasDocumentation` object property. Additionally, each event is addressed through a series of maintenance procedures, which are executed sequentially and connected using the `mmo:hasNext` object property.

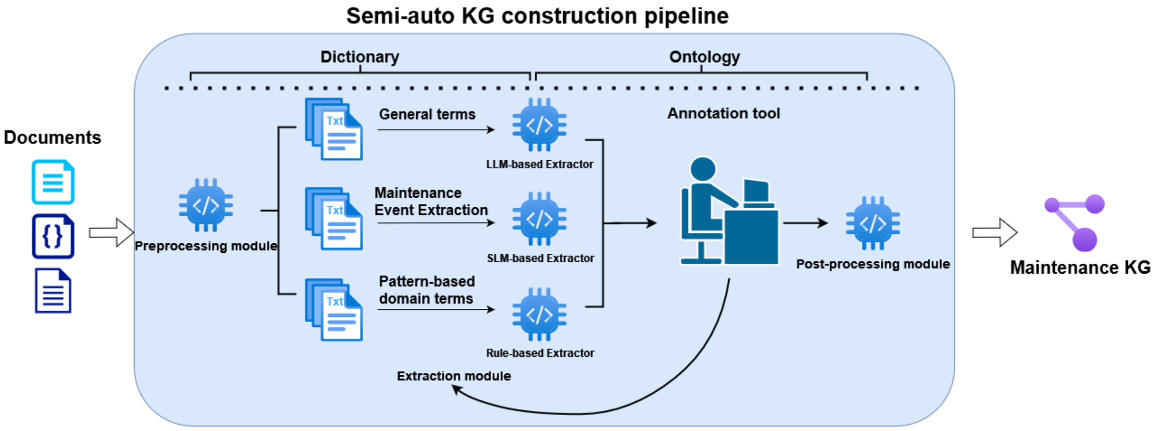

Figure 3: Semi-Automated Construction Pipeline for Maintenance KG Generation.

2.3. Semi-Automated KG Construction Pipeline

For KG construction, we develop a semi-automated pipeline, see Figure 3, that generates a KG from input documents via *four* modules: 1) Pre-processing, 2) Extraction, 3) Annotation, and 4) Post-processing.

Pre-processing Module: The first module processes raw unstructured maintenance documents in diverse formats. This includes textual content extraction, text cleaning, and normalization to ensure compatibility with the extraction module.

Extraction Module: This module combines a large language model (LLM), small language model (SLM), and a rule-based extractor to extract all entities from processed documents following the ontology design. The *LLM-based extractor* (we employ model *gpt-4*) extracts general terms and relationships, e.g., names, timestamps, and general maintenance procedures, offering foundation coverage for diverse, non-specific concepts across documents. The domain-specific *SLM-based extractor* is a lightweight, focused model, e.g., of 13B parameter, that specializes in extracting information from domain-specific, event-based documents. We utilize OneKE [8] as a baseline model, with the potential for future fine-tuning to optimize performance and automation level for processing maintenance event data in industry-specific context. The *Rule-Based Extractor* captures via rule-based methods pattern-based domain terms, such as equipment identifiers and structured codes. The methods rely on predefined dictionaries and regular expressions, ensuring high precision for extracting domain-specific information. This combination of techniques ensures a robust balance between the flexibility provided by LLMs, the domain knowledge of fine-tuned SLMs and the precision accomplished via rule-based extraction.

Annotation Module: The extracted information is refined and validated using an annotation tool, which plays a crucial role in ensuring data quality and ontology compliance. The tool allows users to verify whether extracted candidates, e.g., entities, relationships, or events align with the entities and schema defined in the ontology. Through a user interface, users can easily review, validate, and correct the extraction results.

Post-processing Module: After annotation, the post-processing module focuses on mapping instances to their corresponding classes, generating structured triples for the KG. This modules also eliminates redundancies, validates the accuracy of entities and relationships, as well as ensures overall data quality. The final triples are ingested into a graph database, making them available for potential reasoning, querying, and retrieval.

3. Results and Evaluation by Key Stakeholders

In this section, we show metrics on the generated KG and evaluate the semantic model in a qualitative manner. For the latter, we collect feedback via a questionnaire on the *2* QA pairs from *four* data scientists

[1] https://github.com/eclipse-esmf/esmf-manufacturing-information-model

Table 1
Metrics for Developed Ontology *MMO* and Generated Maintenance KG. *Key: OP: object properties; DP: datatype properties; US: unique subjects; UR: unique relationships; UO: unique objects*

Ontology				KG			
Classes	OP	DP	Namespaces	Triples	US	UR	UO
16	17	9	10	21,893	4,461	108	9,337

Table 2
Results of QA for Data Scientists. Result values in percent.

Question ID	Strongly Disagree	Disagree	Neither Agree nor Disagree	Agree	Strongly Agree
Q1	0	0	0	25	75
Q2	0	0	0	0	100

as they are responsible for continuously analyzing shopfloor data and their structure. The developed ontology *MMO* contains *16* classes and *26* properties, thereof *17* object and *9* datatype properties. The generated KG based on *49* documents and this ontology, results in *21,893* triples, *4,461* unique subjects, *108* relationships (object properties) and *9,337* objects. We show both metrics in Table 1.

Table 2 shows the responses from our evaluation on questions *Q1: Does the developed semantic model (ontology) effectively represents the domain knowledge related to machine maintenance?* and *Q2: Does the ontology bridge the gap between domain knowledge and the unstructured documents?*. We observe a strong agreement that the proposed ontology bridges the graph between domain knowledge and unstructured documents (*Q2*). Further, there is agreement to strong agreement that the ontology effectively represents the domain knowledge related to machine maintenance.

4. Discussion and Lessons Learned

In this section, we discuss results and lessons learned. From ontology and KG metrics, we can observe that the high number of generated triples suggests an impactful first step in knowledge extraction and structuring from unstructured maintenance documents. The final evaluation of helpfulness of the ontology and KG will be conducted in future work in an in-use setting within a Graph-RAG framework for manufacturing maintenance tasks. Also, both evaluated questions meet their expectations and hence, suggest a promising ground work for a future Graph-RAG framework. The data scientists highlighted that using this KG to represent unstructured documents effectively bridges the gap between their mental models (i.e., domain knowledge and the actual data). Furthermore, scalability and efficiency of KG construction is directly influenced by the automation level, which remains one of the biggest barriers to adopting approaches like Graph RAG. Constructing a high-quality KG is resource-intensive, as fully automated solutions often compromise precision and are unsuitable for applications demanding high answer accuracy. Our pipeline addresses this challenge by integrating LLMs, SLMs, and rule-based systems to enhance knowledge extraction, complemented by an annotation tool to collect expert feedback as labeled training data. This feedback loop enables continuous model refinement (SLM) through fine-tuning, effectively balancing automation efficiency with the precision required for reliable performance. Another dimension is the evolution from single- to multi-modality. Multi-modal integration is increasingly important in the maintenance context, particularly for handling screenshots and images that contain essential maintenance information, such as step-by-step device installation guidance or annotated components. Providing visual guidance alongside textual answers can significantly enhance the readability and usability of the system. Further, the involvement of domain experts is essential for accurately defining relationships, entities, and contexts within the KG. Their input ensures that the system aligns with real-world processes and delivers relevant results. Last but not least, the re-usability and scalability of the solution is crucial for adapting the system to new use cases and domains. A unified framework for unstructured knowledge extraction should integrate seamlessly with various internal data sources.

5. Related Work

There is no one-fits-all approach in manufacturing KG construction for the maintenance field, specifically considering the need for scalable solutions, handling of domain-specific knowledge and vast amount of unstructured data. Looking at manufacturing KGs in practice, we observe that our Bosch colleagues who are working with the RAG maintenance chat bot introduced in Figure 1 constructed the maintenance KG manually. Research has proposed different (semi-)automated approaches with strengths and weaknesses in recent years, from which we briefly discuss examples in the following. Liu et Lu [9] propose a maintenance KG construction approach from unstructured PDF files (textual, visual, and spatial) that comprises first a layout-based document understanding module and second the actual construction module. The first module models interdependencies among elements, performs multi-modal representation learning, and builds a heterogeneous graph, while the second module extracts maintenance task components and their relationships from this graph. Wang et al. [10] integrate a multi-source maintenance management method, called *Industrial Dataspace (IDS)* into their KG construction approach for equipment maintenance. The authors highlight that this approach reduces the costly involvement of experts, e.g., for ontology maintenance. With our industry-related approach, we utilize an ontology, as they are commonly available in practice, and evaluate it with expert feedback. Huang et al. [11] propose an ontology-guided KG construction for a multi-level KG for industrial purposes and integrate expert rules into their framework. Specifically, for the instance layer construction which includes the population of the semantic KG layer with actual production data, the authors do not appear to utilize LLMs or SLMs which we aim for in order to increase efficiency in our approach. We further extend the expert involvement by integrating expert feedback via an annotation tool for labeled training data. Also, we observe Guo et al. [12] stating the same challenges such as manual knowledge graph construction and unstructured relevant data which is not yet efficiently structured and exploited. Similarly to our work, the authors propose a top-bottom KG construction, i.e., first designing an ontology, and then extracting knowledge from data sources. Our work differs from their proposed knowledge extraction with *BERT-Improved TRANSFORMER-CRF* as we aim to leverage strengths of different approaches for different tasks, e.g., extraction of pattern-based domain terms via rules and extraction of general terms and relationships via an LLM. While several works have provided valuable insights into semi-automated KG construction from unstructured manufacturing data, few of these research efforts are anchored in real-world data and expert reviews. There is further a gap in existing literature addressing LLM-supported KG construction for manufacturing maintenance tasks. This gap underscores the significance of our contribution. We believe that high-quality KG construction will continue to require focused expert involvement in the future. Leveraging strengths of different KG construction methods, we combine three approaches, LLM-, SLM-, and rule-based, in our work for efficiency and involve experts for high quality.

6. Conclusion

In this paper, we introduce a semi-automated KG construction pipeline comprising LLM-, SLM-, and rule-based methods. This pipeline significantly reduces manual effort, enabling efficient KG construction while maintaining data quality and consistency. While conform to the introduced *Manufacturing Maintenance Ontology*, our approach balances automation with expert involvement on real-world data. The results are promising to motivate a Graph-RAG framework in order to improve the RAG-based maintenance solution from the shopfloor. Challenges such as data silos and handling multi-modal data like images are identified as areas for future work. Expanding the KG to incorporate additional stations, lines and plants, along with integrating visual data, is expected to further improve the system's usability and coverage.

Declaration on Generative AI

The author have not employed any Generative AI tools in creating the paper.

References

[1] S. Vaidya, P. Ambad, S. Bhosle, Industry 4.0–a glimpse, Procedia manufacturing 20 (2018) 233–238.

[2] D. Edge, H. Trinh, N. Cheng, J. Bradley, A. Chao, A. Mody, S. Truitt, D. Metropolitansky, R. O. Ness, J. Larson, From local to global: A graph rag approach to query-focused summarization, arXiv preprint arXiv:2404.16130 (2024).

[3] B. Peng, Y. Zhu, Y. Liu, X. Bo, H. Shi, C. Hong, Y. Zhang, S. Tang, Graph retrieval-augmented generation: A survey, arXiv preprint arXiv:2408.08921 (2024).

[4] B. Sarmah, D. Mehta, B. Hall, R. Rao, S. Patel, S. Pasquali, Hybridrag: Integrating knowledge graphs and vector retrieval augmented generation for efficient information extraction, in: Proceedings of the 5th ACM International Conference on AI in Finance, 2024, pp. 608–616.

[5] Z. Xu, M. J. Cruz, M. Guevara, T. Wang, M. Deshpande, X. Wang, Z. Li, Retrieval-augmented generation with knowledge graphs for customer service question answering, in: Proceedings of the 47th international ACM SIGIR conference on research and development in information retrieval, 2024, pp. 2905–2909.

[6] Y. Li, B. Starly, Building a knowledge graph to enrich chatgpt responses in manufacturing service discovery, Journal of Industrial Information Integration 40 (2024) 100612.

[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.

[8] H. Gui, L. Yuan, H. Ye, N. Zhang, M. Sun, L. Liang, H. Chen, Iepile: Unearthing large-scale schema-based information extraction corpus, arXiv preprint arXiv:2402.14710 (2024).

[9] Z. Liu, Y. Lu, A task-centric knowledge graph construction method based on multi-modal representation learning for industrial maintenance automation, Engineering Reports 6 (2024) e12952.

[10] Y. Wang, Y. Cheng, Q. Qi, F. Tao, Ids-kg: An industrial dataspace-based knowledge graph construction approach for smart maintenance, Journal of Industrial Information Integration 38 (2024) 100566.

[11] X. Huang, C. Yang, Y. Zhang, S. Lou, L. Kong, H. Zhou, Ontology guided multi-level knowledge graph construction and its applications in blast furnace ironmaking process, Advanced Engineering Informatics 62 (2024) 102927.

[12] L. Guo, X. Li, F. Yan, Y. Lu, W. Shen, A method for constructing a machining knowledge graph using an improved transformer, Expert Systems with Applications 237 (2024) 121448.