Complex Systems Informatics and Modeling Quarterly (CSIMQ)
eISSN: 2255-9922
Published online by RTU Press, https://csimq-journals.rtu.lv
Article 197, Issue 36, September/October 2023, Pages 1–15
https://doi.org/10.7250/csimq.2023-36.01

Design Objectives for Evolvable Knowledge Graphs

Anna Teern[1*], Markus Kelanti[1], Tero Päivärinta[1,2], and Mika Karaila[3]

[1] M3S, University of Oulu, Oulu, 90014, Finland
[2] Information Systems, Luleå University of Technology, Luleå, 97187, Sweden
[3] Valmet Automation Systems, Valmet Oyj, Tampere, 33900, Finland

anna.teern@oulu.fi, markus.kelanti@oulu.fi, tero.paivarinta@oulu.fi, mika.karaila@valmet.com

Abstract. Knowledge graphs (KGs) structure knowledge to enable the development of intelligent systems across several application domains. In industrial maintenance, comprehensive knowledge of the factory, machinery, and components is indispensable. This study defines the objectives for evolvable KGs, building upon our prior research, where we initially identified the problem in industrial maintenance. Our contributions include two main aspects: firstly, the categorization of learning within the KG construction process and the identification of design objectives for the KG process focusing on supporting industrial maintenance. The categorization highlights the specific requirements for KG design, emphasizing the importance of planning for maintenance and reuse.
Keywords: Knowledge Graph, Semantic Web, Knowledge Engineering, Learning System, Design Science Research.

1 Introduction

Knowledge bases (KBs) are used to represent human knowledge and create artificial intelligence (AI) solutions for people [1]. Since Google's Knowledge Graph (KG) in 2012, many domains have implemented KGs to organize knowledge [2]. Conceptually, a KB consists of stored knowledge in a specific field. On the other hand, a KG is an approach that models and implements a KB by representing a schema of information entities and their relations as a graph, allowing for flexible interconnections among the entities [3].

Previously, academia was interested in the semantic web that models information as a graph, but where open information sharing is one of the fundamental principles. KGs, which do not hold such principles, have been more keenly adopted by companies [4]. KGs are applied to question-answering, recommendations, information retrieval and many other tasks [5]. Given their ability

[*] Corresponding author

Reference: A. Teern, M. Kelanti, T. Päivärinta, and M. Karaila, "Design Objectives for Evolvable Knowledge Graphs," Complex Systems Informatics and Modeling Quarterly, CSIMQ, no. 36, pp. 1–15, 2023. Available: https://doi.org/10.7250/csimq.2023-36.01

Additional information. Author ORCID iD: A. Teern – https://orcid.org/0000-0002-8214-3181, M. Kelanti – https://orcid.org/0000-0003-1886-8521, T. Päivärinta – https://orcid.org/0000-0002-7477-0783, and M. Karaila – https://orcid.org/0000-0002-1458-7960. PII S225599222300197X. Received: 15 February 2023. Accepted: 2 October 2023. Available online: 31 October 2023.

to organize human knowledge, KGs offer a valuable tool for explaining AI-driven decisions [6]. Semantic web and KGs are also applied within the Industry 4.0 context, where KGs are especially helpful in integrating heterogeneous information from various devices [7].

However, the literature still needs to address the details of KG construction. Although insights from ontology construction can be leveraged [8], there is currently no standardized construction process. Nonetheless, some examples exist that can serve as a starting point for organizations embarking on their KG journey [9].

Our practical motivation for this research stems from a collaborative project between industry and academia known as Oxilate[†]. Together with a company, we developed a prototype of an intelligent assistant to support maintenance personnel in managing a complex cyber-physical system (CPS). We selected the creation and maintenance of a KG as the overall approach to organizing knowledge for the assistant. Maintenance requires time and resources in the manufacturing industry, and AI is increasingly used to optimize activities. For instance, AI systems offer suggestions to maintenance personnel based on past cases [10] or context-optimized assistance [11], and original equipment manufacturers (OEMs) rely on failure diagnostics to enhance their products and services [12]. KGs are often at the core of such systems.

Our research reviewed literature and developed an integrated process model of KG creation and maintenance, published in [13]. The literature review focused on process elements of KG creation. The results of the review are summarized in Section 3. As the project progressed, we defined design objectives and examined the learning aspects of the KG construction process (KGCP). This article covers the definition and initial verification of design objectives as a part of a design science research (DSR) process [14], utilizing both literature and a case study interactively. Hence, this article takes the research from the problem definition, published in [13], to the objectives' definition in DSR.

The article is organized as follows. The next section outlines the research methodology. Section 3 summarizes the relevant literature, culminating in an integrated model for the KGCP. Section 4 describes the case, illustrating an operationalization of the KGCP and the emerging design challenges. Section 5 elaborates on the design objectives, and section 6 discusses the contributions of this research. Lastly, Section 7 outlines our future research.

2 Methodology

Our research follows the principles of the design science research process [14]. The progress of the study is depicted in Figure 1. Firstly, the end-users, service specialists, and maintenance engineers require support in their fieldwork. Currently, they can contact service specialists in the company who assist the other workers. An intelligent assistant can offer support in the future, while experts can focus on particularly challenging cases to create novel solutions. The software engineer's responsibility is to design and implement the digital assistant for the service specialists. At the same time, researchers observe how the KG is created and develop the process.

The methodology to develop the process in this study encompasses a literature review and a case study, which together form a part of the design science research process [14]. The literature review published in detail in [13] focused on the process elements involved in creating KGs. Both knowledge graph and knowledge base were used as search terms to ensure comprehensive coverage. The authors drew literature from Google Scholar, Web of Science, and Scopus in 10/2021. The main inclusion criterion was: "The paper describes a process for creating a KB/KG", regardless of the application domain. In the end, the authors reviewed 37 studies from 1997 to 2021.

[†] Oxilate = Operational eXcellence by Integrating Learned information into AcTionable Expertise https://itea4.org/project/oxilate.html

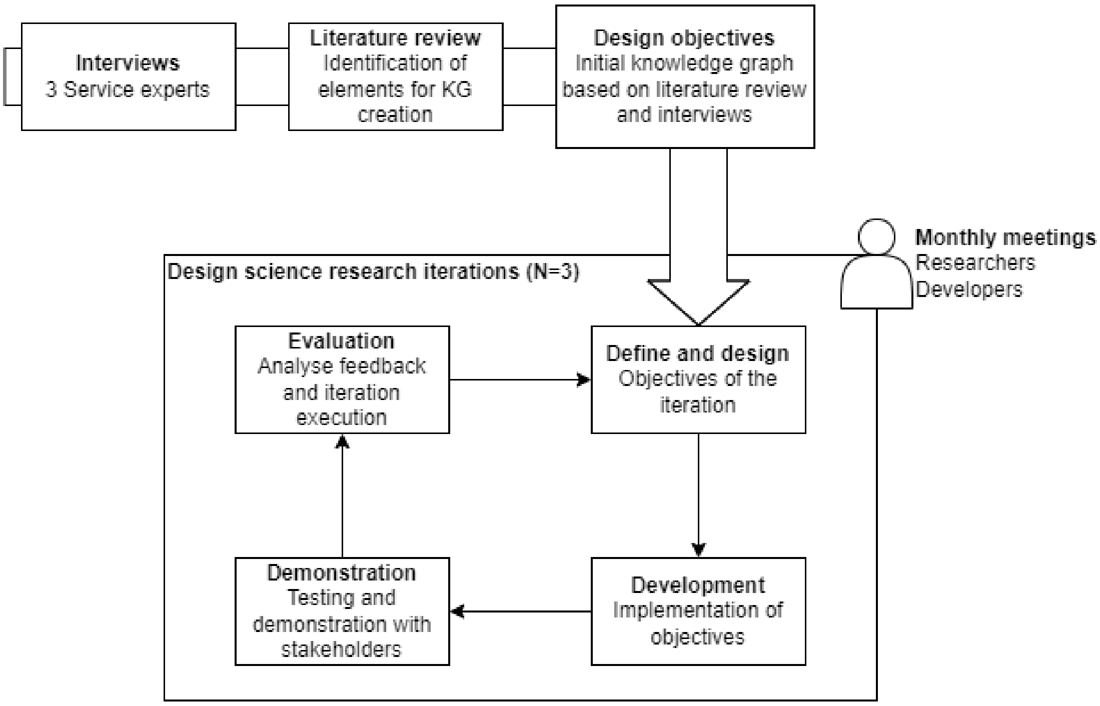

Figure 1. Design science research process of the study

Researchers obtained the initial problem definition and design challenges through three expert interviews with experienced service specialists in the case study. One co-author then developed the actual KG for the collaborating company, providing design research input and verifying the initially defined design objectives. This design-oriented exploration offers professional and practical interaction, leading to the crystallization of the design objectives. The academia-industry exchange took place through workshops from May 2021 to December 2022. The collaborative project aimed at knowledge mobilization for experts in industrial maintenance using a digital assistant.

The development process for the KG-based digital assistant involved approximately three iterations. In the first iteration, a mock-up assistant was built upon technologies such as Pandora and Media Wiki to serve as the foundation for a chatbot. However, challenges remained in creating content to enable a question-answering interface with appropriate content maintenance responsibilities.

The second development cycle focused on leveraging existing documents to support maintenance engineers. The developer and researcher chose a prefabricated ontology [15] and found that it captured all relevant information for the maintenance task scenario. Eventually, the developer used the company's maintenance document structures to build a proof-of-concept for the company. This demonstrates KG's usability as a knowledge reuse solution for supporting service specialists and maintenance engineers in the field. Furthermore, a comprehensive understanding of the KG creation and maintenance process and its subtasks was essential. Based on the literature, a task model for such a process was created, with a more crystallized set of intermediate design challenges [13].

In the third cycle, the emphasis shifted towards developing the interface of the digital assistant. The developer expanded the proof-of-concept system to incorporate a stream of structured machine health data combined with an improved view of documents, thus providing enhanced support for the maintenance task. This iteration highlighted the need for evolvable KGs and identified the design objectives focusing on how the KG can dynamically evolve through learning.

The authors worked iteratively designing and evaluating the prototype in a company context. We followed a build-evaluate approach and incorporated the evaluation possibilities presented by Sonnenberg and vom Brocke [16]. Throughout the process, we gained theoretical insights into creating a KG through literature and practical output by developing a prototype of an intelligent assistant based on a KG. As a result, the study produced a problem statement validated within the context and established design objectives for an evolvable KG that supports an intelligent assistant for industrial maintenance.

The evaluation process was conducted in two phases: phase 1 involved ensuring the novelty and importance of the problem; Phase 2 involved assessing the clarity and applicability of the design objectives and the completeness and understandability of the artefact [16].

The following section describes the initial process diagram, but more details about the literature review process are given in [13]. The case study section (4) describes the interaction of researchers and the use case context on industrial maintenance.

3 Integrated Process Diagram

This section summarizes the literature on KG construction processes in and beyond industrial maintenance.

Knowledge engineering (KE) is an iterative process with the stages of knowledge acquisition, representation, knowledge base, validation, inferencing, and explanation and justification [1]. Knowledge acquisition encompasses gathering data from various sources, e.g., documents and experts. The stage and subsequent tasks can be adapted to a modern data-intensive environment. Knowledge representation describes tasks for modelling knowledge for computer processing, then stored in a knowledge base. Knowledge validation ensures data integrity and resolves factual disputes. Expert systems often explain suggested actions, requiring the stage of explanation.

For many years, the descriptions of the KB followed similar lines of reasoning with the KE process. In 2017, Pan et al. [17] proposed an iterative process for knowledge management and KG creation for businesses, introducing a different naming convention for the stages compared to previous approaches. They outlined a high-level process with three main stages: construction, storage, and consumption, and elaborated on the KG life cycle. They included KG consumption as a stage, recognizing that its planned utilization impacts other stages. [3] introduced knowledge fusion as a step in the process, drawing inspiration from data fusion to identify actual triples instead of accurate data.

We identified five overarching stages for the process: knowledge acquisition, knowledge fusion, knowledge processing, knowledge storage and knowledge utilization. These stages are commonly found in many sources, including traditional KE principles [1] with updates from modern data science [18]. The stages include 14 tasks, which may consist of sub-tasks, as reported, e.g. in [19].

We define knowledge acquisition as tasks required to discover or structure data to collect knowledge. It is essential to differentiate between data acquisition and knowledge extraction because raw or pre-processed data still needs to be transformed to offer understanding to users. Knowledge extraction involves discovering patterns in the data, often through entity, relation, and attribute extraction [19].

Studies do not define the terms knowledge fusion and knowledge processing uniformly and use them interchangeably. This paper defines knowledge fusion as validation and integration with sub-tasks such as entity co-reference or disambiguation. Ontology matching is a well-known field for KB integration [20], and new techniques are developed for more complex cases and automating the process [21], [22]. Knowledge processing, on the other hand, encompasses model creation, quality evaluation (e.g., fact-checking), and knowledge inferencing [3].

Knowledge storage focuses on tasks directly related to stored knowledge. It includes knowledge representation: modelling knowledge for computer processing and storage technologies, data retrieval and visualization [1]. Knowledge utilization involves tasks that exploit the KG for specific use cases.

We distinguish between ontology selection and creation/update because these tasks have distinct characteristics. Most articles adopt a bottom-up approach, creating an ontology from data. However, some discuss the possibility of selecting a predefined ontology, although it often requires adaptation for the specific use case [23], [24]. Consequently, they may need to update ontologies during the knowledge-processing stage.

Although previous studies from 1997 to 2021 follow a similar process, with knowledge acquisition, fusion, further processing, and storage in a KB, knowledge utilization is often overlooked despite many papers describing a process for an application-specific KG. There are already helpful reviews available for many sub-tasks [19], [25], so we focus on providing a higher-level view of the process. The results of this review support creating a unified process representation that incorporates the previous KGCPs, as shown in Figure 2.

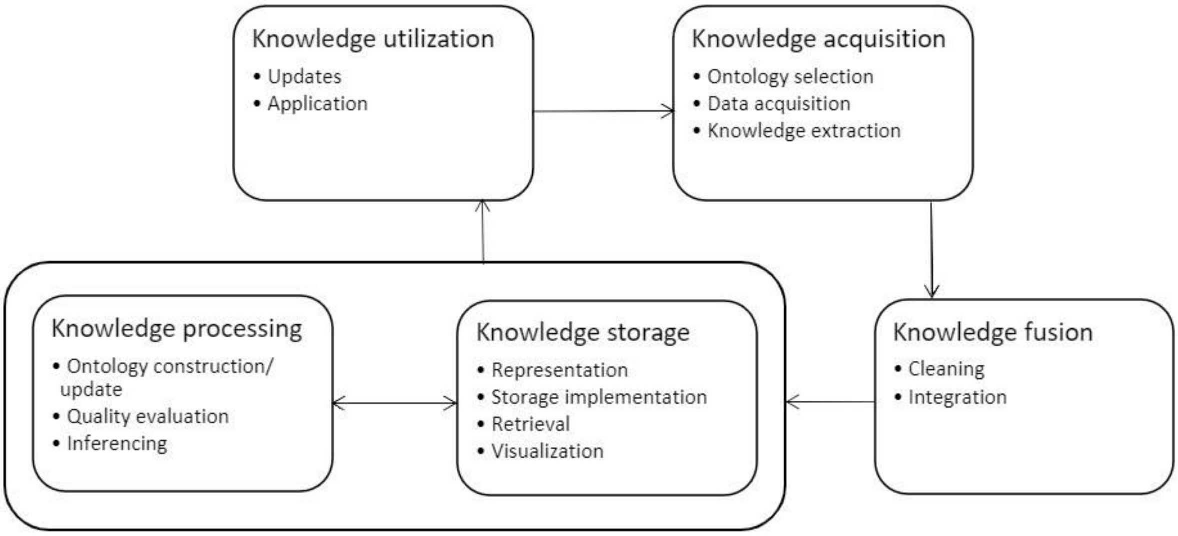

Figure 2. The proposed KG construction process diagram

The top-down KG construction approach starts with the knowledge acquisition stage, where data for ontology building is acquired. Then, knowledge fusion tasks are performed as needed. Knowledge representation occurs in knowledge storage, followed by ontology creation in the knowledge processing stage. When ontology has been built, knowledge utilization may not be relevant unless users or representatives evaluate it. Instead, data is obtained with ontology-based knowledge extraction in the knowledge acquisition stage. Knowledge fusion tasks are then performed again where needed, followed by knowledge storage tasks, ultimately leading to knowledge utilization.

The bottom-up approach starts with knowledge acquisition but goes straight into data acquisition and knowledge extraction utilizing ML methods. Then, knowledge fusion tasks are performed as required. Again, functions in knowledge storage are carried out, and an ontology may be created in the knowledge processing stage. In both approaches, knowledge processing tasks may be performed on accumulated knowledge, even if knowledge is already being utilized.

It is important to note that knowledge and related structures evolve due to changes in industrial equipment, applications, processes, and people. KG construction and maintenance must be treated as a continuous process to provide timely knowledge for work processes in such dynamic environments. [26] suggests that the current environment's ever-improving nature requires a spiral evolution of KGs, illustrated in our iterative process.

4 The Case of Industrial Maintenance

This case involves a company that offers maintenance services to the systems they manufacture. The systems of machines are sold globally, and the maintained machines are often spread in various locations, while the service specialists (possessing the most comprehensive knowledge) work near the headquarters in Finland. Industrial maintenance (IM) relies heavily on human effort aided by technology. When a local maintenance engineer meets trouble, they contact a service specialist. The experts are often overloaded with requests, which increases maintenance time. To solve this, the company developed a digital assistant for the onsite maintenance engineers that should ease the workload of the service specialists and allow more time for challenging cases. The assistant will also reduce the need for travel by the experts, thus reducing cost and environmental impact.

The digital assistant allows maintenance engineers to identify causes for the problem and offers relevant documentation, depicted in Figure 3. Computers or robots cannot fully replace engineers, but the issue can be addressed with an intelligence augmentation application that supports problem-solving. We have identified the need to understand better the expert knowledge to support the engineers in their knowledge-based tasks as one of the design objectives in a broader context [27]. It is vital to notice that new solution material is added to documents through humans, whereas the assistant uses existing material to offer suggestions.

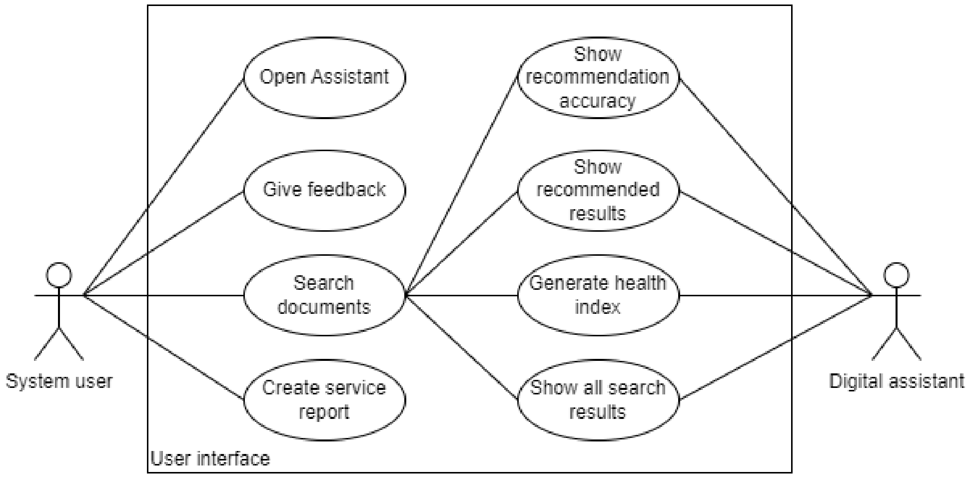

Figure 3. Use case diagram of the digital assistant in industrial maintenance

Firstly, we needed to decide whether we needed an ontology and then if we build it or select one. Reference ontology is a domain-specific vocabulary of common entities, and it can be used to create application ontologies and KGs [15]. We evaluated two IM ontologies for the use case. Yahya et al. developed an ontology for Industry 4.0 based on an extensive literature review [7]. However, this ontology presented only high-level concepts for IM and was not applicable. Karray et al. developed a reference ontology designed directly for IM [15], first used in the ideation phase. Also, RAMI 4.0 was helpful in the development phase [28]. In the end, the OPC UA-based process automation device information model (PA-DIM) is used [29].

While the ontology gave the structure and data needs, data was gathered from various sources, and entities, attributes, and relations were matched from documents to the ontology structure. An equivalent format for the same document type was essential to ease the automatic detection of nodes and edges. However, many documents are completed by humans, leading to various contents and styles in individual copies. The iterative nature of KGCP becomes visible when new data is added to the system and knowledge is inferred based on existing data. Several data types are variably changeable, e.g., metadata is more static, whereas runtime data has a higher velocity. The suggestions can be drawn from various kinds of documents with different attributes on update times and other contents. Furthermore, the content format and structure may also vary between data sources.

In addition to affecting knowledge acquisition, the diverse sources affect the integration task, i.e., duplicate information is identified and removed. For instance, when data is integrated, and suggestions are given, the first mock-ups link documents beyond the user's access rights. Because the company trades internationally, workers often write the records in local languages, complicating finding similar cases and connecting the same entities. Computer translation tackled a part of this problem and worked well enough between the main languages (Finnish and English); hence, computer translations can be tested for other languages.

As content structures vary and relevant information for solving a problem involves multiple information sources, problem solution usually requires human thinking. For instance, understanding and solving a single issue can require analysis of tickets, updates, warnings, manuals, and other sources. For a digital assistant to work, text analysis with natural language processing (NLP) techniques could be used to understand similar or most common cases. The sameness of an issue has multiple levels, i.e., it is possible to compare only the recurring problem of the same equipment instance, the same type of equipment in the same factory, or the same type of equipment across all factories. This poses challenges for inferencing, as the algorithms must be carefully designed to address the needs.

Machine learning methods for knowledge processing must be studied further. Numerous machine-learning techniques have been introduced for KG refinement in the literature. One learning method might involve workers' communications with the digital assistant. The reliability of the information must also be evaluated, e.g., the data can contain measuring mistakes or incorrect information.

4.1 Architecture

While developing the KG graph databases, GraphQL and Neo4j were found to be valuable tools. A web server application, shown in Figure 4, was employed as a proof-of-concept to test and evaluate the system. The initial application requirements will be reviewed and revised with a group of test users. In the current state of the art excellent data import tools such as LangChain [30] are available for seamless data integration.

A mobile application capable of indoor location detection Is ideal for users. This would allow users to access information about nearby assets effortlessly. For more detailed asset-specific information, the asset´s tag or serial number can be used to identify the object in question.

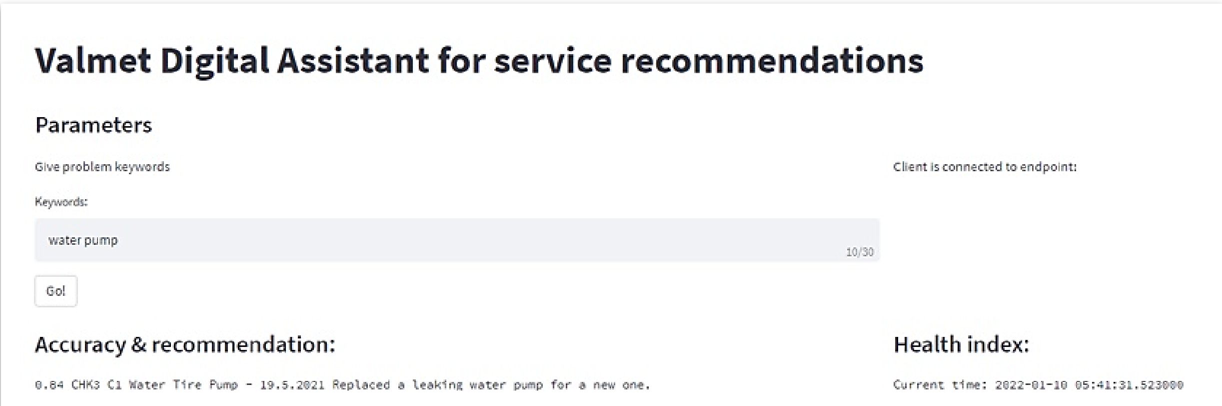

Figure 4. Web server application as a proof-of-concept

We encountered certain scalability and performance challenges along the way. When training the document to vector-model, utilizing an extensive number of service reports can lead to substantial memory consumption. A practical approach would be to limit the number of documents to around 2000–3000 and remove similar and older records from Neo4j before training the model, thus reducing memory usage.

The KG serves as the underlying foundation for the assistant's operations. A diagram depicting the activities of the assistant is available in Figure 5. While the KG is used, users can provide feedback on the usability of the suggestions. This feedback influences the stored knowledge and how it is represented later. Additionally, the assistant can gather usage data and assess the usefulness of knowledge. For instance, the speed and effectiveness with which an issue is resolved and whether it can be solved without human expert involvement serve as indicators of the assistant's performance. These aspects enable the assistant to learn and improve over time.

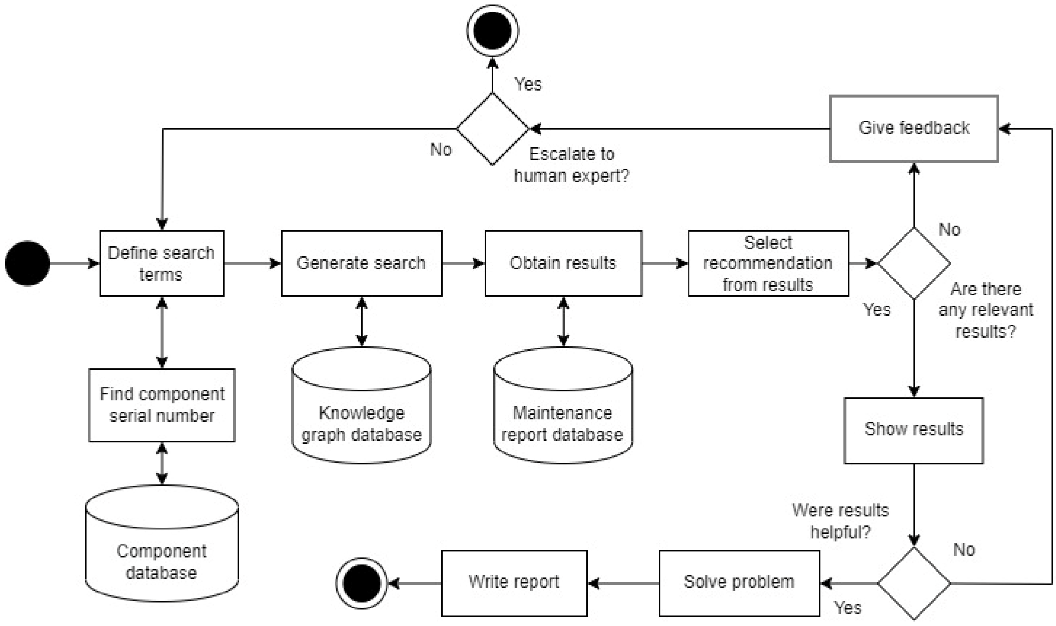

Figure 5. Activity diagram of the intelligent assistant

5 Design Objectives for an Evolvable KG

In the last paper [13], we discussed the challenges posed by the practical implementation of KG according to the process description. These challenges have been brought forward with the continued development of the KG and intelligent assistant to define design objectives for such a system. Before discussing these objectives, we first characterize learning in an intelligent system utilizing a KG.

5.1 Learning in the Intelligent System

Based on several reviews on the meaning of intelligence, the driving force for it is learning [31]–[33]. Therefore, understanding the learning methods in the KGCP is essential. As we conducted our literature review and collaborated with industry, we identified two distinct methods of incorporating learning in the intelligent system through KGCP.

The first method concerns machine learning algorithms that can find new entities or relations through inferencing in the knowledge processing stage [3]. These methods are prominent in the literature and are also known as KG completion or knowledge reasoning. These methods are not just machine learning methods used to construct the KG that "learn" from the data and build the core elements of the KG. Instead, the inferencing methods add new knowledge to the KB based on existing knowledge.

Logical inferencing has already been used for early expert systems, and inference structures were introduced by Parpola in connection with their KB construction process [34]. Soon after the rising interest towards KG, Ktob and Li suggested that these learning methods should be developed generically for any KG [3]. Such general reasoners are available on some KB services, one of which is applied in the process described by Niknam and Karshenas [35]. A few years ago, Chen et al. reviewed knowledge reasoning methods and categorized them based on the approach into logic-based, embedding-based and neural network-based reasoning [36].

The second method of incorporating learning concerns knowledge utilization, discussed in the case. Contemporary systems can collect information about the usage of the system and feedback from the user, which can be used to update the KG with added information or human preferences.

This seems especially important for offering the workers relevant information through the intelligent assistant. There are cases in the literature where this type of learning is suggested, for instance, in personalized education [37], [38], where learning happens in the models that enable continuous improvement in teaching.

One more way of incorporating learning concerns meta-learning [39], [40], which happens in the machine learning methods used in the KG construction. This type of learning occurs on a different level, and we will not investigate the effects of meta-learning at this point.

5.2 Design Objectives

Now that we have characterized the two ways of incorporating learning in the KGCP, we will discuss their implications for designing the KG. We identify two design objectives supported by the third objective and further endorsed by design mechanisms. Design objectives are structured according to design principles [41].

Existing documents in industrial maintenance offer many opportunities to reuse knowledge, was the Oxilate project's original goal. However, thousands of new records are added monthly, making it impossible for humans to benefit from the added knowledge. It is also highly challenging to see similar cases or connections between different cases in such an immense amount of data. Therefore, machine learning methods are needed to make those connections and bring summarized knowledge to the user.

The first design objective, described in Table 1, defines the need for inferencing. Inferencing algorithms have been researched extensively, and it is possible to proceed initially with general reasoning algorithms. However, there are few reported successes of their use in the industrial maintenance or manufacturing field [42], so these must be tested and experiences gathered to understand how these methods affect KG.

Table 1. Describing the first objective of learning

Title: Objective of Learning Through Knowledge Processing		
Components	**Structure**	**KGCP stage involved**
Aim, implementer, and user	For a *data analyst* to allow *learning through knowledge processing* for *KG*	
Context	In the system of intelligent industrial maintenance assistant	
Mechanisms	Employ inferencing algorithms involving data analysts	Knowledge processing, Knowledge storage
Rationale	Because it enables finding hidden information in data.	

Secondly, when service specialists use the assistant, they will offer valuable feedback about the applicability of the knowledge. The use case may also change in time, changing the information needs. Considering these changes is a new aspect of the KG construction and must be investigated in the design and building process.

Hence, the second design objective, described in Table 2, deals with knowledge utilization, which is a new component in the KGCP. The search behavior must be gathered and stored within the KG to allow learning via knowledge utilization actions. These actions may include but are not limited to search strings, communication with the assistant, feedback from the user, and user characteristics.

A Five-star rating system can be used for user feedback in the industrial maintenance case. Firstly, this can be used to order recommendations by the given rating. Secondly, the rating can be used for learning algorithms to recognize features for good advice. The recommendations with low ratings can be removed or returned to the author to fill in low-quality/missing information.

Table 2. Describing the second objective of learning

Title: Objective of Learning Through Knowledge Utilization		
Components	**Structure**	**KGCP stage involved**
Aim, implementer, and user	For a *software developer* and a *data analyst* to allow *learning through knowledge utilization* for *KG*	
Context	In the system of intelligent industrial maintenance assistant	
Mechanisms	Employ *search behavior and feedback gathering by representing humans* in the KG involving *service specialists/engineers* and *data analysts*	Knowledge utilization, Knowledge acquisition, Knowledge storage
Rationale	Because doing so will enable tacit knowledge gathering from maintenance engineers and maintain timely knowledge in the KG.	

Evolvability is an essential characteristic for a KG to enable learning in the system. Evolvability means the ability to process and execute (incremental) changes in the KG. These changes consist of the ability to process new types of entities and relations. In the case of evolvable KG, new kinds of entities and links are only added occasionally and in small amounts proportional to the size of the KG. However, new entities and relations are inevitable as the KG is constantly used. Evolvability also means the ability to handle differences between cases and users by recognizing similar cases and characteristics of users (such as security clearance) and the ability to remove redundant information to save storage and processing costs and time. The objective toward evolving KG is detailed in Table 3.

Table 3. Detailing the overall goal for evolvable KGs

Title: Objective of Evolvable KG		
Components	**Structure**	**KGCP stage involved**
Aim, implementer, and user	For a software developer to achieve evolvable KG for the intelligent assistant	
Context	In supporting maintenance engineers	
Mechanisms	Employ changing characteristics of relations and entities, adding new types of them and automatic data cleaning	Knowledge acquisition, Knowledge processing, Knowledge storage
Rationale	Because doing so will support achieving the objectives of learning through knowledge processing and utilization.	

Mechanisms to achieve evolvable KG involve tasks in knowledge acquisition especially, although the design aspects of all the KGCP stages (see Figure 1) must be considered at least to some extent. The knowledge extraction algorithms must be designed to allow these incremental changes. It may mean that there should be another layer of AI tools to recognize changes. When utilization is a fundamental part of the KGCP, there must be ways to handle feedback and other user-initiated changes in the KG.

The knowledge fusion allows the integration of knowledge from external sources, and the integration activities are affected by changing knowledge, especially knowledge structures. Furthermore, from an evolvability perspective, new KBs may be integrated into the KG, necessitating a carefully designed KG for interoperability.

For other KGCP stages, the fundamental idea is the importance of planning and design, although not all possibilities would be utilized when KG is first constructed. For instance, visualization could contain various options so that assistants can learn to pick the right ones for the user, and feedback should be designed to include meaningful aspects per use case.

In the last paper [13], we discussed design challenges in the industrial maintenance case. We recognized that the iterative nature of the KGC process would set challenges throughout the process stages. This is now concretized through the objective of evolvable KG. Although aspects of evolvability have come up in other research [25], our objective identifies the mechanisms to achieve evolvable KG. More established design principles are expected to emerge during the continuing investigation.

In industrial maintenance, changing characteristics could mean changing the strength of relations that lead to solving an issue or changing attributes of links and entities that help recognize similar cases. Maintenance engineers working on the machines have various levels of access to documents. Therefore, this characteristic should be stored in the KG so that only relevant, accessible information is delivered to the user. First, rule-based solutions for automatic data cleaning are planned because they seem the most reliable and feasible. Ensuring that removed data is not added again in updates is also important.

Evaluation of the design objectives has been a continuous process throughout the study. By engaging in ongoing discussions and seeking feedback and input from industry stakeholders, we were able to refine the objectives, ensuring their practical feasibility in real-world scenarios. Identifying design challenges lets us concentrate on core issues to be addressed during the design process. The challenges are encapsulated within the objectives, ensuring their relevance and effectiveness in solving the identified challenges.

6 Discussion

Our study contributes to KG construction in general and industrial maintenance specifically. The primary outcome of the research is the novel KG construction process model applicable across various fields already described in [13]. This process model is a valuable guide for knowledge graph construction and maintenance initiatives. It offers a comprehensive and integrated framework outlining the stages and tasks of the iterative process.

Previous literature has predominantly focused on details of varying stages and tasks of the process. Our research goes beyond this fragmented approach by consolidating and synthesizing existing literature on the KG construction and maintenance process, providing a cohesive understanding of the critical elements. By examining multiple sources and drawing insights from traditional KE principles and modern data science, the research achieves a comprehensive process model of KG. Compared to the hitherto most complete model that recognized the KG creation tasks in a waterfall model [43], our study suggests the iterative nature of KG creation and maintenance and recognizes the need for evolvable KGs.

The process model's contributions lie in its ability to bring clarity and structure to the complex KGCP. It provides a systematic approach to knowledge acquisition, fusion, processing, storage, and utilization, ensuring that all essential aspects are considered and addressed. Moreover, the model highlights the iterative nature of KGCP, emphasizing the need for continuous learning and adaptation.

In addition to the process model, this article describes further results of the continued design science research. The study categorized different learning aspects in the KGCP and identified design objectives towards evolvable KG. The study has two main elements of contributions that enhance the understanding and application of KGs.

Firstly, categorizing the learning aspects helps to highlight their specific requirements and challenges, enabling practitioners and researchers to address them more effectively. The categorization guides the design and implementation of KGCP, ensuring that the learning processes are appropriately accounted for and managed throughout the construction of KGs.

Moreover, by identifying the three design objectives, our study contributes to the KG construction research by describing clearly defined goals that can drive the future direction of KG research. By defining these objectives, we can focus our efforts on developing solutions that address the real-world needs of organizations, ultimately leading to more impactful and relevant research outcomes.

Our research emphasizes the importance of considering the long-term sustainability and scalability of KGs, enabling organizations to effectively manage and update their knowledge repositories over time and facilitating better decision-making, knowledge sharing, and collaboration. By formulating concrete objectives, our research provides practical guidance for practitioners seeking to leverage KGs in industrial maintenance. The design objectives provide mechanisms for achieving an evolvable KG and facilitate learning through knowledge processing and utilization. Following the DSR approach, we aim to uncover more thorough design principles for an evolvable KG that allows learning within intelligent systems.

The industrial maintenance case confirms the necessity of continuously changing the KG to accommodate evolving visions for utilizing intelligent systems in the field. Therefore, designing the KG construction with an iterative and continuous process is vital, acknowledging the need to learn from the constant interactions of the workers utilizing knowledge. The developed process facilitates a more systematic and efficient approach to construction that leads to more accurate KGs that capture the relevant knowledge needed for decision-making and problem-solving.

The evaluation of the model has been somewhat limited in scope because we have focused on the design aspect thus far. We gained valuable insights through interaction with an industrial partner, but the evaluation process is confined to this specific collaboration. We will address the need for a more complete process model through a systematic literature review focusing on learning in the KGCP. We recognize the importance of evaluating the process model usability for KG development in a broader setting and plan to do so after creating the next iteration of the process.

Overall, the contributions of our research advance the knowledge and application of KGs in industrial maintenance. Furthermore, the insights gained from our study can inform the design and implementation of KGs in other domains, facilitating the effective organization, retrieval and utilization of knowledge.

7 Conclusion

This research contributes to the existing knowledge with specified design objectives for evolvable KG in industrial maintenance. It continues our research that presented a comprehensive process model for the knowledge graph construction and maintenance process (KGCP). The process model consolidated and synthesized existing literature, providing a systematic framework that guides practitioners and researchers in building KGs beyond the IM context.

The process model encompasses five stages and fourteen tasks, covering the entire KG construction and maintenance lifecycle. It emphasizes the iterative nature of KGCP, recognizing the need for continuous learning and adaptation. By addressing the gaps in previous literature and integrating insights from traditional knowledge engineering principles and modern data science, the model offers a holistic view of KGCP.

Furthermore, this research identifies design objectives connected to the KGCP stages, enabling the development of an evolvable KG and facilitating learning through knowledge processing and utilization. This contributes to the ongoing advancement of intelligent systems and their effective utilization in various domains. The process model provides a foundation for KGCP initiatives, empowering organizations to construct, maintain, and utilize knowledge graphs effectively. By embracing this comprehensive and iterative approach, organizations can enhance their knowledge representation, retrieval, and utilization, ultimately leading to improved decision-making, efficiency, and innovation.

Future research should focus on validating the proposed model and design objectives in different contexts. Additionally, continuing our DSR process and discovering design principles will be valuable. By addressing these areas, researchers can further advance the field of KG construction and facilitate the development of intelligent systems in various domains. Our future work involves continuing with action-oriented collaborative design research with the reported case, unfolding design principles with reported solution experiences.

Acknowledgements

This research has been partially funded by the ITEA project Oxilate (18023) and Business Finland.

References

[1] E. Turban, J. E. Aronson, T.-P. Liang, and R. V. McCarthy, *Decision Support Systems and Intelligent Systems*, 7th edition. Upper Saddle River: Prentice Hall, 2004.

[2] J. Chicaiza and P. Valdiviezo-Diaz, "A Comprehensive Survey of Knowledge Graph-Based Recommender Systems: Technologies, Development, and Contributions," *Information*, vol. 12, no. 6, Article 6, 2021. Available: https://doi.org/10.3390/info12060232

[3] A. Ktob and Z. Li, "The Arabic Knowledge Graph: Opportunities and Challenges," in *2017 IEEE 11th International Conference on Semantic Computing (ICSC)*, 2017, pp. 48–52. Available: https://doi.org/10.1109/ICSC.2017.22

[4] P. Hitzler, "Semantic Web: A Review of The Field," *Semantic Web*, 2020.

[5] X. Zou, "A Survey on Application of Knowledge Graph," vol. 1487, p. 012016, 2020. Available: https://doi.org/10.1088/1742-6596/1487/1/012016

[6] E. Rajabi and K. Etminani, "Knowledge-graph-based explainable AI: A systematic review," *Journal of Information Science*, 2022. Available: https://doi.org/10.1177/01655515221112844

[7] M. Yahya, J. G. Breslin, and M. I. Ali, "Semantic Web and Knowledge Graphs for Industry 4.0," *Applied Sciences*, vol. 11, no. 11, Article 5110, 2021. Available: https://doi.org/10.3390/app11115110

[8] N. F. Noy and D. L. McGuinness, "Ontology Development 101: A Guide to Creating Your First Ontology," 2001.

[9] L. Tesfaye, "How to Build a Knowledge Graph in Four Steps: The Roadmap from Metadata to AI," Enterprise Knowledge. Available: https://enterprise-knowledge.com/how-to-build-a-knowledge-graph-in-four-steps-the-roadmap-from-metadata-to-ai/. Accessed on June 08, 2023.

[10] G. Fenza, M. Gallo, V. Loia, D. Marino, and F. Orciuoli, "A Cognitive Approach based on the Actionable Knowledge Graph for supporting Maintenance Operations," in *2020 IEEE Conference on Evolving and Adaptive Intelligent Systems (EAIS)*, May 2020, pp. 1–7. Available: https://doi.org/10.1109/EAIS48028.2020.9122759

[11] L. Kohl, F. Ansari, and W. Sihn, "A Modular Federated Learning Architecture for Integration of AI-enhanced Assistance in Industrial Maintenance – A novel architecture for enhancing industrial maintenance management systems in the automotive and semiconductor industry," in *Competence development and learning assistance systems for the data-driven future*, Goto Verlag, 2021, pp. 229–242. Available: https://doi.org/10.30844/wgab_2021_14

[12] H. Hossayni, I. Khan, M. Aazam, A. Taleghani-Isfahani, and N. Crespi, "SemKoRe: Improving Machine Maintenance in Industrial IoT with Semantic Knowledge Graphs," *Applied Sciences*, vol. 10, no. 18, Article 6325, 2020. Available: https://doi.org/10.3390/app10186325

[13] A. Teern, M. Kelanti, T. Päivärinta, and M. Karaila, "Knowledge graph construction and maintenance process: Design challenges for industrial maintenance support," in *BIR 2022 Workshops and Doctoral Consortium*, CEUR-WS.org, vol. 3223, pp. 1–13, 2022. Available: https://ceur-ws.org/Vol-3223/paper1.pdf

[14] K. Peffers, T. Tuunanen, M. A. Rothenberger, and S. Chatterjee, "A Design Science Research Methodology for Information Systems Research," *Journal of Management Information Systems*, vol. 24, no. 3, pp. 45–77, 2007. Available: https://doi.org/10.2753/MIS0742-1222240302

[15] M. H. Karray, F. Ameri, M. Hodkiewicz, and T. Louge, "ROMAIN: Towards a BFO compliant reference ontology for industrial maintenance," *Applied Ontology*, vol. 14, no. 2, pp. 155–177, 2019. Available: https://doi.org/10.3233/AO-190208

[16] C. Sonnenberg and J. vom Brocke, "Evaluations in the Science of the Artificial – Reconsidering the Build-Evaluate Pattern in Design Science Research," in *Design Science Research in Information Systems. Advances in Theory and Practice*, K. Peffers, M. Rothenberger, and B. Kuechler, Eds., Lecture Notes in Computer Science. Springer, vol. 7286, 2012, pp. 381–397. Available: https://doi.org/10.1007/978-3-642-29863-9_28

[17] J. Z. Pan, G. Vetere, J. M. Gomez-Perez, and H. Wu, *Exploiting Linked Data and Knowledge Graphs in Large Organisations*. Springer, 2017. Available: https://doi.org/10.1007/978-3-319-45654-6

[18] X. Zhao, Y. Jia, A. Li, R. Jiang, and Y. Song, "Multi-source knowledge fusion: a survey," *World Wide Web*, vol. 23, no. 4, pp. 2567–2592, 2020. Available: https://doi.org/10.1007/s11280-020-00811-0

[19] J. Yan, C. Wang, W. Cheng, M. Gao, and A. Zhou, "A retrospective of knowledge graphs," *Frontiers of Computer Science*, vol. 12, no. 1, pp. 55–74, 2018. Available: https://doi.org/10.1007/s11704-016-5228-9

[20] J. Euzenat and P. Shvaiko, *Ontology Matching*. Springer, 2013. Available: https://doi.org/10.1007/978-3-642-38721-0

[21] E. Thiéblin, O. Haemmerlé, N. Hernandez, and C. Trojahn, "Survey on complex ontology matching," *Semantic Web*, vol. 11, no. 4, pp. 689–727, 2020. Available: https://doi.org/10.3233/SW-190366

[22] A. Bento, A. Zouaq, and M. Gagnon, "Ontology Matching Using Convolutional Neural Networks," in *Proceedings of the Twelfth Language Resources and Evaluation Conference*, Marseille, France: European Language Resources Association, 2020, pp. 5648–5653. Available: https://aclanthology.org/2020.lrec-1.693. Accessed on Jan. 11, 2023.

[23] M. A. Islam, "Leveraging knowledge graph for UK environmental legislation related to building construction and maintenance," M.S. thesis, School of Engineering Science, Lappeenranta University of Technology, Lappeenranta, Finland, 2019.

[24] E. Filtz, S. Kirrane, and A. Polleres, "The linked legal data landscape: linking legal data across different countries," *Artificial Intelligence and Law*, vol. 29, pp. 485–539, 2021. Available: https://doi.org/10.1007/s10506-021-09282-8

[25] A. Hogan *et al.*, "Knowledge graphs," *ACM Computing Surveys*, vol. 54, no. 4, pp. 1–37, 2021. Available: https://doi.org/10.1145/3447772

[26] X. Li, Y. Chen, B. Pettit, and M. De Rijke, "Personalised Reranking of Paper Recommendations Using Paper Content and User Behavior," *ACM Transactions on Information Systems*, vol. 37, no. 3, pp 1–23, 2019. https://doi.org/10.1145/3312528

[27] H. Bomström *et al.*, "Digital Twins About Humans – Design Objectives from Three Projects," *Journal of Computing and Information Science in Engineering*, vol. 22, pp. 1–10, 2022. Available: https://doi.org/10.1115/1.4054270

[28] A. Müller, I. Grangel-González, and C. Lange, "An Ontological View of the RAMI4.0 Asset Administration Shell," in *Proceedings of the 12th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management*, Budapest, Hungary: SCITEPRESS – Science and Technology Publications, 2020, pp. 165–172. Available: https://doi.org/10.5220/0010111101650172

[29] FieldComm Group, "Process Automation Device Information Model – PA-DIM," FieldComm Group. Available: https://www.fieldcommgroup.org/technologies/pa-dim. Accessed on June 26, 2023.

[30] I. Campos, "ChatGPT LangChain Example for Chatbot Q&A," Sopmac AI. Available: https://medium.com/sopmac-ai/chatgpt-langchain-example-for-chatbot-q-a-a8b6ef40bbb6. Accessed on June 26, 2023.

[31] G. Pili, "Toward a Philosophical Definition of Intelligence," The International Journal of Intelligence, Security, and Public Affairs, vol. 21, no. 2, pp. 162–190, 2019. Available: https://doi.org/10.1080/23800992.2019.1649113

[32] R. J. Sternberg, "Theories of intelligence," in *APA handbook of giftedness and talent*, Washington, DC, US: American Psychological Association, 2018, pp. 145–161. Available: https://doi.org/10.1037/0000038-010

[33] J. R. López-Robles, J. R. Otegi-Olaso, I. Porto Gómez, and M. J. Cobo, "30 years of intelligence models in management and business: A bibliometric review," *International Journal of Information Management*, vol. 48, pp. 22–38, 2019. Available: https://doi.org/10.1016/j.ijinfomgt.2019.01.013

[34] P. Parpola, "Inference in the SOOKAT object-oriented knowledge acquisition tool," *Knowledge and Information Systems*, vol. 8, pp. 310–329, 2005. Available: https://doi.org/10.1007/s10115-004-0181-6

[35] M. Niknam and S. Karshenas, "A shared ontology approach to semantic representation of BIM data," *Automation in Construction*, vol. 80, pp. 22–36, 2017. Available: https://doi.org/10.1016/j.autcon.2017.03.013

[36] X. Chen, S. Jia, and Y. Xiang, "A review: Knowledge reasoning over knowledge graph," *Expert Systems with Applications*, vol. 141, 2020. Available: https://doi.org/10.1016/j.eswa.2019.112948

[37] Y. Ma, D. Hong, F. Dan, X. Yang, and X. Li, “Research on the Construction Method of Knowledge Graph for Power Grid Education Resources,” in *2021 IEEE 3rd International Conference on Computer Science and Educational Informatization (CSEI)*, IEEE, 2021, pp. 99–103. Available: https://doi.org/10.1109/CSEI51395.2021.9477728

[38] J. Zhao, X. Yang, Q. Qiao, and L. Chen, “Personalized Learning Design of Ideology and Politics of Distance Education Courses Based on Big Data,” in *2020 IEEE International Conference on Progress in Informatics and Computing (PIC)*, 2020, pp. 135–139. Available: https://doi.org/10.1109/PIC50277.2020.9350748

[39] J. Vanschoren, “Meta-Learning,” in *Automated Machine Learning: Methods, Systems, Challenges*, F. Hutter, L. Kotthoff, and J. Vanschoren, eds., in The Springer Series on Challenges in Machine Learning, Springer, 2019, pp. 35–61. Available: https://doi.org/10.1007/978-3-030-05318-5_2

[40] R. Vilalta and Y. Drissi, “A Perspective View and Survey of Meta-Learning,” *Artificial Intelligence Review*, vol. 18, pp. 77–95, 2002. Available: https://doi.org/10.1023/A:1019956318069

[41] S. Gregor, L. C. Kruse, and S. Seidel, “Research Perspectives: The Anatomy of a Design Principle,” *Journal of the Association for Information Systems*, vol. 21, no. 6, pp. 1622–1652, 2020. Available: https://doi.org/10.17705/1jais.00649

[42] L. He, B. Dong, and P. Jiang, “A Heuristic Grafting Strategy for Manufacturing Knowledge Graph Extending and Completion Based on Nature Language Processing: KnowTree,” *IEEE Access*, vol. 9, pp. 90847–90862, 2021. Available: https://doi.org/10.1109/ACCESS.2021.3092019

[43] J. Hou, R. Qiu, J. Xue, C. Wang, and X.-Q. Jiang, “Failure Prediction of Elevator Running System Based on Knowledge Graph,” in *Proceedings of the 3rd International Conference on Data Science and Information Technology*, 2020, pp. 53–58. Available: https://doi.org/10.1145/3414274.3414283