Title	Building semantic models and knowledge graphs for intelligent smart manufacturing applications
Author(s)	Yahya, Muhammad
Publication Date	2024-04-04
Publisher	NUI Galway
Item record	http://hdl.handle.net/10379/18131

Doctoral Thesis

Building Semantic Models and Knowledge Graphs for Intelligent Smart Manufacturing Applications

Muhammad Yahya

April 2, 2024

Supervisors
Prof. John G Breslin, Dr. Muhammad Intizar Ali

CONFIRM SFI Research Centre for Smart Manufacturing
College of Science and Engineering
University of Galway, Ireland

ABSTRACT

Industry 4.0 (I4.0) (or smart manufacturing) is a new era in the industrial revolution that emphasises machine connectivity, automation, and data analytics. This revolution has led to the creation of production lines that produce machine-generated process data through sensors, leading to increased efficiency and productivity. Ontologies have been used to integrate the data from various formats into a single, unified form. However, most of these ontologies have overstudied the essential concepts related to the I4.0 production line that are of key importance in building a knowledge graph for smart manufacturing. This thesis aims to propose a framework that can be adopted by any I4.0 production line with minimal modifications to build its knowledge graph. The framework has been tested using realistic data from two separate industrial use cases.

The existing ontologies in the manufacturing domain have limited depth and expressiveness due to their scope and purpose mapping for application specificity. As a result, this hinders the stakeholders in constructing their knowledge graphs. The **First Contribution** of this thesis is to address this challenge of application specificity. We provide *Reference Generalized Ontological Model (RGOM)* based on the Reference Architecture Model for I4.0.

The I4.0-based Knowledge Graphs, or Knowledge Graph (KG) have been receiving significant attention over the past few years, and many researchers are involved in building them in the form of manufacturing production lines KG. However, most of the time, they have limitations when applied to a specific use case. These use cases are based on two possibilities: (1) if the researchers are using synthetic data, or (2) if the use case is coming from an industry based on their private company data. The **Second Contribution** of this thesis is to address this challenge related to data being real or synthetic. We provide one of the first datasets based on the realistic data collected from a football production line. We have proposed an automated approach for mapping the data into RGOM to build a KG that is made publicly available for experiments by the I4.0 community. Moreover,

the dataset enables the demonstration of RGOM adaptability with minimal modification in a manufacturing environment.

The current techniques used to build KGs focus on integrating data from heterogeneous sources and often result in missing links between the entities. As a consequence of the missing links within the KGs, they cannot be exploited by the applications. We observe some missing links in the developed football production line KG. The **Third Contribution** of this thesis is to solve this challenge related to missing links. We address the challenge of KG missing links by utilizing state-of-the-art KG embedding models, namely *ComplEx*, *DistMult*, *TransE*, *ConvKB*, and *ConvE*, on football manufacturing production line datasets.

The current ontologies are not publicly available and therefore cannot be accessed by other users for reuse purposes. Such a lack of availability often requires that users build their ontologies from scratch, is a time-consuming task. The **Fourth Contribution** of this thesis is the employment of a use case from Bosch to determine how RGOM can serve as a domain manufacturing ontology, facilitating integration among various data sources. In relation to this, we developed the *Resistance Spot Welding Ontology (RSWO)* and align it with the RGOM.

This research has introduced the Reference Generalised Ontological Model (RGOM) as a flexible framework for manufacturing production lines, which can be applied to any production line with minimal modifications. It can also be employed as a manufacturing domain-level ontology by aligning ontologies at the application level for enhanced interoperability. The results on the benchmark dataset (I40KG) have demonstrated more efficient production processes and improved overall performance. Furthermore, the process of predicting missing links in the I40KG indicated that translational models demonstrated better performance on manufacturing-based KGs compared to neural network models. This distinction can be attributed to the hierarchical structure of the KGs.

DECLARATION

I declare that this thesis, titled "*Building Semantic Models and Knowledge Graphs for Intelligent Smart Manufacturing Applications*", is composed by myself, that the work contained herein is my own except where explicitly stated otherwise in the text, and that this work has not been submitted for any other degree or professional qualification.

Galway, April 2, 2024

Muhammad Yahya

DEDICATION

This thesis is dedicated to the memory of Ghat Abbai and Haji Baba (both of them rest in peace), and to my supportive parents, Daji and Abbai. Their love, wisdom, and unwavering faith guided my journey and inspired each word of this work. This accomplishment is yours as much as it is mine.

ACKNOWLEDGEMENTS

First and foremost, I want to express my gratitude to Allah (SWT) for giving me the stamina to do this work.

I am deeply grateful to my Ph.D. supervisors, Prof. John G. Breslin and Dr. Muhammad Intizar Ali, for their unwavering support throughout my doctoral journey. Their guidance, insightful discussions, valuable suggestions, and constructive feedback have been actively involved in shaping my research work and enhancing my critical thinking.

I would also like to sincerely thank Dr. Evgeny Kharlamov, and Dr. Baifan Zhou for providing me with a research intern position at the Bosch Centre of Artificial Intelligence which helped me enhance my research skills in ontology engineering.

I would also like to thank my family members, who have been my sources of strength and inspiration. My father, Mr. Fazli Wahid, and my mother, Sahib Zadgai, always pushed me towards success with the values of hard work, perseverance, and determination. My uncle Dr. Sheroz Khan, my brothers Iftikhar Ahmed, Hilal Muhammad, and Muhammad Asim Khan, my sisters Salma Bibi and Najma Bibi, my lovely wife Bushra Zaman, and my daughter Anabia Yahya have been my constant support during my Ph.D. journey.

Furthermore, I would like to acknowledge the role of my close friends Dr. Faisal Khan, Dr. Muhammad Ali Farooq, Dr. Qaiser Mehmood, Yasar Khan, and Nouman Khan, who have been a constant source of motivation and encouragement. Their fruitful discussions, constructive criticism, and unwavering support have helped me.

Finally, I sincerely thank my colleagues, especially Dr. Lan Yang, Dr. Abdul Wahid and Dr. Jaleed Khan, for proofreading my work, their valuable career insights and their unwavering support. Their contributions to my research and professional development have been invaluable, and I owe them a debt of gratitude.

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

AI	Artificial Intelligence
CQs	Competency Questions
CPS	Cyber Physical system
DL	Description Logic
FAIR	Findable Accessible Interoperable Reusable
I4.0	Industry 4.0
I4.0KG	Industry 4.0 Knowledge Graphs
ISO	International Organization for Standardization
IoT	Internet of Things
KG	Knowledge Graph
KGs	Knowledge Graphs
LOD	Linked Open Data
LOT	Linked Open Terms
OOPS	Ontology Pitfall Scanner
RAMI4.0	Reference Architectural Model Industrie 4.0
RGOM	Reference Generalized Ontological Model
RSW	Resistance Spot Welding
RSWO	Resistance Spot Welding Ontology
RDF	Resource Description Framework
RDFS	Resource Description Framework Schema
OWL	Web Ontology Language

LIST OF PAPERS

Following are the peer-reviewed scientific articles produced from this thesis for getting published in repute journals, and international conferences:

1. Yahya, Muhammad, John G. Breslin, and Muhammad Intizar Ali. "Semantic web and knowledge graphs for industry 4.0." Applied Sciences 11, no. 11 (2021): 5110.

2. Yahya, Muhammad, Baifan Zhou, Zhuoxun Zheng, Dongzhuoran Zhou, John G. Breslin, Muhammad Intizar Ali, and Evgeny Kharlamov. "Towards generalized welding ontology in line with ISO and knowledge graph construction." In The Semantic Web: ESWC 2022 Satellite Events: Hersonissos, Crete, Greece, May 29–June 2, 2022, Proceedings, pp. 83-88. Cham: Springer International Publishing, 2022.

3. Yahya, M., Ali, A., Mehmood, Q., Yang, L., Breslin, J.G. and Ali, M.I., A benchmark dataset with Knowledge Graph generation for Industry 4.0 production lines. Semantic Web, (Preprint), pp.1-19.

4. Yahya, Muhammad, Baifan Zhou, John G. Breslin, Muhammad Intizar Ali, and Evgeny Kharlamov. "Semantic Modeling, Development and Evaluation for the Resistance Spot Welding Industry." IEEE Access (2023).

5. Muhammad Yahya, Abdul Wahid, Lan Yang, John G. Breslin , Evgeny Kharlamov and, Muhammad Intizar Ali. "Harnessing Knowledge Graphs for Advanced Manufacturing: A Performance Analysis of Embedding Models for Link Predictions" Under Review in Expert Systems With Applications Journal.

CONTENTS

1 INTRODUCTION

Manufacturing has undergone a significant transition as a result of Industry 4.0 (I4.0), which is characterised by the growth of smart manufacturing [153]. This transition has opened up new opportunities for greater efficiency, flexibility, and decision-making capabilities. Cutting-edge technologies like the Internet of Things (IoT) harnessed by Artificial Intelligence (AI), and controlled by the Cyber-Physical Systems (CPS) are now deeply interlinked with traditional manufacturing processes [153]. This advancement has initiated a fundamental shift towards an interconnected, intelligent ecosystem. Within this setting, the interactions between human-machine and machine-machine have become increasingly commonplace, largely due to the semantic web's heavy reliance on ontologies [55]. However, the potential of this transformation has not yet been fully realised due to the challenges in managing and leveraging the substantial amounts of data generated by smart manufacturing systems [150]. The crucial issue lies in integrating data of various formats emanating from diverse sources and with an ultimate goal of inter-operation within a manufacturing environment for improved decision-making, proactive maintenance, quality control, and resource optimisation, ensuring data availability for sharing [170].

This thesis aims to explore the challenges of data integration, data interoperability, and knowledge discovery and propose solutions to build effective semantic models and knowledge graphs for intelligent smart manufacturing applications, ultimately enhancing our ability to understand, integrate, and utilise data more universally in the manufacturing domain to realise human-machine and machine-machine communication more concisely.

1.1 BACKGROUND AND MOTIVATION

Over the past decade, the manufacturing industry has been experiencing a significant transformation driven by advancements in technology and the widespread adoption of intelligent systems [37]. This transformation, commonly referred to as smart manufacturing or I4.0 has revolutionised traditional manufacturing processes by integrating their production facilities with disruptive trends of cutting-edge technologies such as the Internet-of-Things (IoT), artificial intelligence (AI), cloud computing and data analytics, and autonomous and cyber-physical systems [77] throughout their operations. As a result, the manufacturing landscape has changed, offering new opportunities for improved efficiency, enhanced quality control, increased flexibility, and capabilities for better decision-making.

Smart manufacturing enables the digitisation and interconnection of machine-machine, human-machine, and various components across the manufacturing ecosystem, including processes, products, and supply chains [52]. By harnessing the power of intelligent technologies, manufacturers can optimise production processes, minimise downtime, and reduce costs. Furthermore, they can deliver customised products and services to meet the dynamic and ever-changing demands of the market by adapting to user requirements that play an important role in the acceptance of emerging applications.

While smart manufacturing offers immense potential, it also brings forth several challenges that need to be addressed for its fully functional implementation [128]. One of the significant challenges lies in effectively utilising the vast amounts of data generated by smart manufacturing systems consisting of sensors, machines, machine parts, processes, and other interconnected devices for capturing valuable information about processes, product quality, resource utilisation, and so on [26]. However, the challenge lies in the integration and interoperability of data sources of a heterogeneous nature within the manufacturing environment [59]. Data is generated and stored in various formats, structures, and locations across the systems in a manufacturing production line, making it available for sharing and transmission as semantic knowledge. Achieving seamless integration and meaningful collaboration between these diverse data sources is essential for a comprehensive understanding of the manufacturing processes by facilitating effective decision-making and enabling proactive maintenance, quality control, and resource optimisation. Furthermore, the manufacturing domain lacks standardised models and frameworks to represent and share knowledge effectively [95]. Existing ontologies in the manufacturing domain are often limited in their scope and purpose and designed for specific applications in the manufacturing domain [139]. This restricts the stakeholders' ability to construct comprehensive Knowledge Graphs (KG) that can holistically capture and represent the manufacturing domains' rich semantics [1]. As a result, interoperability and information exchange in a manufacturing production line between different systems become challenging, hindering the integration and utilisation of relevant information across the system landscape.

To overcome these challenges, the application of semantic models and KGs in smart manufacturing has gained great significance [176]. Semantic models provide a structured representation of knowledge, enabling a common understanding of concepts, relationships, and semantics across different systems in a manufacturing production line. KGs capture and connect knowledge elements in a graph-like structure, facilitating efficient knowledge organisation and integration [79].

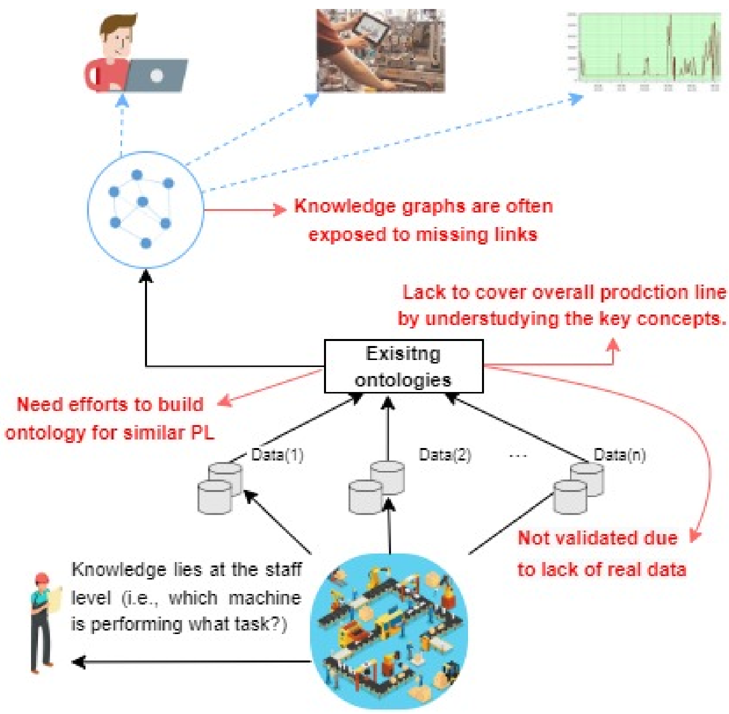
By leveraging semantic models and KGs, intelligent smart manufacturing applications can be embedded with enhanced interoperability, knowledge sharing, and decision-making capabilities. They enable a holistic and realistically shared view of integrating information from diverse sources, such as sensors, machines, and processes, to generate heterogeneous and unstructured data. This, in turn, empowers manufacturers to gain valuable insights, optimise processes, detect anomalies, predict maintenance needs, and make informed decisions to promote operational efficiency, product quality, and customer satisfaction.

Moreover, the existing semantic models developed for smart manufacturing face several challenges. Firstly, they suffer from limited semantic expressiveness due to a lack of alignment with industry standards, hindering stakeholders from constructing comprehensive KGs and integrating relevant information. Within the manufacturing industry, there is a prevalent tendency to create new ontologies from scratch, disregarding the *Reuse* principle of Linked Open Data (LOD) [69]. Next, due to the unavailability of real-time industrial datasets, semantic models are yet to be evaluated, thus restricting them to only proposals. Moreover, there are often missing links in the KGs constructed from manufacturing production line data. The missing links are the edges that are considered valid but are not included in the knowledge graph. Lastly, the ontologies are not aligned with the domain-level ontologies and thus suffer from interoperability issues.

In this thesis, we aim to explore and address the challenges of constructing semantic models and knowledge graphs for intelligent, smart manufacturing production lines. By developing novel methodologies and leveraging State-Of-The-Art (SOTA) techniques [17], [106], [157], we strive to enhance the level of understanding, integration, and utilisation of data and knowledge in the field of manufacturing. Furthermore, our objective is to demonstrate the practical application of the semantic model by utilising real-world data obtained directly from the manufacturing industry to represent knowledge for subsequent extraction and relevant applications.

1.2 PROBLEM DEFINITION AND CHALLENGES

The digitisation and automation of manufacturing processes characterise I4.0. It is driven by disruptive trends because of potential human-machine and machine-machine interactions, aiming to enhance production and re-



(a) Challenges

Figure 1.1: Overview of the challenges faced by production lines in the I4.0 landscape.

lated services. A defining feature of I4.0 factories is the equipping of assets and machinery with sensors for effective monitoring of production line resources. This comprehensive monitoring allows for the early identification of potential failures, affecting the performance, energy consumption, and reliability of the manufacturing processes in a smart manufacturing environment to produce products for intelligent applications such as digital twins, smart robotics, quality control systems, and predictive maintenance. Proactive decisions can then be made to prevent production downtime. However, interpreting the data collected by the sensors is a complex task due to the need to integrate and process heterogeneous data from various sources, each produced in different formats. The building of semantic models[1] and knowledge graphs have emerged as a vital solution to this problem [76]. Despite this, existing semantic models and KGs have the following challenges:

Challenge I: Adherence to the Reuse of Existing Ontologies and Industry Standards. Ontologies have become crucial tools in the realm of smart manufacturing, playing a substantial role in the integration and interoperability of data acquired from production machinery, processes, sensors, and others [26, 42]. These ontological models are composed of components like classes and properties that together collectively establish a shared vocabulary. However, while they are increasingly utilised, it is crucial to examine these ontological models critically to identify potential areas for improvement. The current ontologies in smart manufacturing have overlooked some of the crucial concepts important to the domain. Additionally, they are not reusing the existing vocabulary and are not tailored to industrial standards for semantic representation, posing several challenges being addressed by both the industrial and academic research communities. The first challenge is that it lacks the re-usability of the existing vocabulary, and it needs customised data for a specific use referred to at low-level application specificity.

Challenge II: Real Industrial Data for Evaluating Manufacturing Ontologies. Ontology evaluation is an important component of ontology engineering ensuring that ontology has the quality, relevance, and effectiveness necessary to meet the domain's use case requirements [22]. Unfortunately, in smart manufacturing, many ontologies have not been thoroughly evaluated and remain as mere proposals of academic pursuits lacking the prop-

1 Semantic models and ontology(ies) are alternatively used representing the same notion.

erties of industrial implementation due to scarcity of data [30]. The few that have been assessed typically are based on either synthetic data or proprietary level industry data, which is not available for reuse purposes, and hence shun the properties of broader applications.

Challenge III: Missing Links in the I40KG. Real-world knowledge graphs have seen extensive use across a broad spectrum of applications, such as question answering [16], recommender systems [180], and dialogue systems [98]. Despite their utility, a common challenge faced by most KGs is their incompleteness, marked by missing links that limit their full potential [36]. This challenge presents a compelling case for Knowledge Graph Completion (KGC), also known as link prediction, aimed at predicting these missing links, which is a KG analysis task used to predict missing or future connections between nodes in a target network [167]. In the context of this thesis, the research has experienced this as a first-hand issue while constructing the I4.0 Knowledge Graph (I40KG) using football production data. The encountered missing links lead to an incomplete I40KG. This incompleteness poses a hindrance to its full utilisation for the intended intelligent applications [59, 164].

Challenge IV: Alignment of Application Ontologies to the Domain-Level Ontologies. The terms within an ontology are defined to address a particular scope and purpose; for example, the authors in Ramirez-Duran et al. have the knowledge represented at the ontology schema level through the use of *rdfs:subClassOf* [126]. The lack of *rdfs:subClassOf* makes it impossible to express capability and represent hierarchical relationships between classes, which limits its potential applications. More specifically, there is a notable gap when the ontology is created without being aligned to the domain-level ontology that helps to achieve modules' level interoperability for broader applications. A terminological alignment between the concepts occurs when ontologies refer to the same real-world entity but use different names (for example, in the welding domain, a product produced during the manufacturing process can be represented as a "weld spot" or a "weld nugget"). A semantic alignment is thus achieved only when relations and axioms used in the ontologies are correlated with concepts in an alignment extending to the domain level. Additionally, the methodology involved in creating ontologies, particularly those associated with smart manufacturing, is overlooked [30].

1.3 RESEARCH QUESTIONS (RQS)

The following research questions are defined based on the discussion in the previous sections:

RQ1: *How can the limitations such as re-usability, missing concepts, and missing re-adaptability of industry reference architecture of current ontology be addressed to better understand and utilise the vast amounts of data generated by resources in the manufacturing production line?*

In response to RQ1, the RGOM ontology is proposed that represents the generic components of a production line with adaptability for any manufacturing industry by identifying the terminology with the existing missing vocabulary and making necessary modifications.

RQ2: *Can the challenge of scarcity in realistic data for production lines be addressed, and what are the implications of transforming a use-case dataset using RGOM into a benchmark for applications that require production line data?*

To answer RQ2, we collaborated with production line supervisors and engineers from the football industry. The aim was to obtain actual production line data. We acquired a real data instance from the production line in real-time, which served as a basis for generating synthetic data. The engineers in charge of the real-time production line validated the generated data. The RGOMs was used to represent the generated data as a KG. Thus implying the benefits of using ontologies and semantic annotations of data to showcase how the I4.0 industry can benefit from KGs and semantic datasets.

RQ3: *What is the comparative performance of various knowledge graph embedding models, including TransE, DistMult, ComplEx, ConvKB, and ConvE, for the link prediction task in the football I4.0 KG?*

To address this question, five SOTA KG embedding models are trained and tested on the football manufacturing production line I40KG. The models are evaluated with the help of two metrics known as Mean Reciprocal Rank (MRR) and Hits@N, which are widely used to evaluate KG embedding models [172].

> RQ4: *How can the generic semantic model (RGOM) and Knowledge Graphs (KGs) for production lines be adapted and integrated across different domains in the manufacturing industry?*

RSWO ontology is proposed with the help of an ontology development process that is functional for this precise production scenario (while simultaneously aligning with the RGOMs, a domain-level ontology), which is publicly available for reuse purposes. The alignment encourages interoperability, facilitating seamless data exchanges and communication across various systems within the domain.

1.4 MAIN CONTRIBUTIONS

The major contributions of this thesis are outlined as claimed in Papers I through IV, each of which corresponds to achieving objectives I through IV. An overview of the contributions is presented as below:

1.4.1 Contribution I: Propose Reference Generalized Ontological Model to Represent Manufacturing Production Line Domain

To address **RQ1**, the existing ontological models for I4.0 are comprehensively reviewed to identify key limitations and room for improvement. The current models, their uses, and their shortcomings are critically reviewed and have formed the solid foundation for this research. Additionally, the Reference Architectural Model Industrie 4.0 (RAMI4.0), an accepted standard for I4.0, is thoroughly analysed to understand the overall domain knowledge of I4.0 [64]. This has enabled us to answer if any current ontologies follow RAMI4.0. Then, the Reference Generalized Ontological Model (RGOM) is proposed which is intended to address these identified limitations, offering a more extensive and reusable model for I4.0. RGOM is designed while keeping the reuse principle of Linking Open Data (LOD) in mind [100]. It does so by formalizing knowledge such as time, location, and sensor data, along with a multitude of important-to-industry feature attributes like product creation, process management, machine operations, and warehouse operations. The RGOMs leverage reusing existing vocabularies to ensure seamless integration while incorporating missing concepts to bridge the gaps identified in the study. This contribution aims to address **RQ1**.

1.4.2 Contribution II: Benchmarking Dataset and an Automated Approach to Populate its Instances into RGOM to Build I40 knowledge graph

To accomplish Research Question **RQ2**, RGOM's adaptability is demonstrated by benchmarking the dataset from the real data industry manufacturing production line. Like many other models, RGOM adaptability is not demonstrated due to the unavailability of manufacturing data. In this thesis, our goal is to benchmark a dataset to facilitate the generation of knowledge graphs for I4.0 production lines. We further aim to highlight the advantages of employing ontologies and semantic annotations of data, illustrating how Industry 4.0 can benefit from the so-obtained knowledge graphs and semantic datasets. This work was made possible through collaboration with production line managers, supervisors, and engineers in the football industry, which enabled us to gather realistic production line data. Furthermore, the data is automatically mapped and populated to the classes and relationships of the RGOMs using a solution based on JenaAPI [11], resulting in an I40KG. This KG comprises over 2.5 million axioms and approximately 1 million instances. The creation of this extensive KG serves to exhibit the adaptability and practicality of the RGOMs. This address the contributions highlighted in **RQ2**.

1.4.3 Contribution III: Analysing the Effectiveness of SOTA KG Embedding Models on I40KG

To achieve Research Question **RQ3**, this thesis addresses the challenge of predicting missing links by applying SOTA KG models such as ComplEx [157], DistMult [177], TransE [17], ConvKB [106], and ConvE [41]. These models have been utilised on football manufacturing production line datasets I40KG for predicting missing links on the unseen data. The performance of these models has been critically assessed using two essential metrics: Mean Reciprocal Rank (MRR), and Hits@N (Hits@10, Hits@3, and

Hits@1). Our analysis indicates that the TransE model demonstrates superior performance with an average accuracy of 0.91%, closely followed by ComplEx and DistMult with an accuracy of 0.87% and 0.84% respectively. On the contrary, the ConvKB and ConvE models have shown lower performance levels, with an accuracy of 0.79% and 0.76% respectively. Additionally, a remarkable variance in MRR values among the models has been identified, with TransE yielding the highest mean MRR value, and ConvE, the lowest. Significantly, this part of the study enriches both scholarly research and industrial methodologies by identifying the most efficient KG embedding model for predicting missing links in the hierarchical KGs within the field of manufacturing. The results analysed in this contribution support **RQ3**.

1.4.4 Contribution IV: Propose Resistance Spot Welding Ontology (RSWO) and aligned it with RGOM

For addressing Research Question **RQ4**, the RSWO is developed. This ontology formally presents the resistance spot welding operations, equipment, individual machine parts, and software systems. By integrating the domain knowledge of RSWO and harmonising it with the ISO standards (ISO-14327 and ISO-14373) and RGOM ensures that first-hand knowledge terminology is strictly followed. The RSWO offers a comprehensive understanding of RSW welding concepts, extending beyond the narrow application focus of existing ontologies. Next, this research illustrates a systematic ontology development process based on a real-world industrial scenario involving expert knowledge and data from a globally recognised industrial partner. The process entails domain analysis for knowledge collection, formalisation of concepts, and subsequent implementation, validation, and publication including its maintenance. The ontology is practically implemented in the first phase with authentic data from the Bosch welding production line. Moreover, an evaluation of RSWO is carried out using the O'FAIRe methodology to ensure it follows FAIR principles [6]. The OOPs! tool has been employed to assess the structural and functional quality dimensions, such as clarity, completeness, consistency, and conciseness of the proposed ontology [121]. Lastly, the OntoMetrics tool is used to measure the richness of RSWO's attributes, ensuring it is well-populated with real data items from the production lines [87]. This contribution addresses responding to **RQ4**.

1.5 COMPLETE LIST OF PUBLICATIONS

This section presents the complete list of articles and conference papers.

(i). Journal Publications as a first author

1. Yahya, Muhammad, John G. Breslin, and Muhammad Intizar Ali. "Semantic web and knowledge graphs for industry 4.0." *Applied Sciences 11, no. 11* (2021): 5110.
2. Yahya, Muhammad, Aabid Ali, Qaiser Mehmood, Lan Yang, John G. Breslin, and Muhammad Intizar Ali. "A Benchmark Dataset with Knowledge Graph Generation for Industry 4.0 Production Lines." *Semantic Web Journal, IOS press*, 2023
3. Yahya, Muhammad, Baifan Zhou, John G. Breslin, Muhammad Intizar Ali, and Evgeny Kharlamov. "Semantic Modeling, Development and Evaluation for the Resistance Spot Welding Industry." *IEEE Access (2023)*.

(ii). Conference Publications as a first author

1. Yahya, Muhammad, Baifan Zhou, Zhuoxun Zheng, Dongzhuoran Zhou, John G. Breslin, Muhammad Intizar Ali, and Evgeny Kharlamov. "Towards generalized welding ontology in line with ISO and knowledge graph construction." In *The Semantic Web: ESWC 2022 Satellite Events: Hersonissos, Crete, Greece*, May 29–June 2, 2022, Proceedings, pp. 83-88. Cham: Springer International Publishing, 2022.

(iii). Conference Publications as a co-author

1. Zhou, Baifan, Zhipeng Tan, Zhuoxun Zheng, Dongzhuoran Zhou, Yunjie He, Yuqicheng Zhu, Muhammad Yahya et al. "Neuro-Symbolic AI at Bosch: Data Foundation, Insights, and Deployment." *ISWC (2022)*.

1.6 THESIS OUTLINE

Chapter 1 presents the conceptual outline of the entire thesis, starting with the background and motivation to identify the research challenges, followed by defining the research questions and research objectives. Contributions to address the solutions of the research questions as challenges are then provided.

Chapter 2 presents the background knowledge and critical literature review of semantic modelling approaches to look for grey areas. The relative insights are appropriately cited in support of the title under research and are highlighted in order to develop a platform where research work is ultimately launched.

Chapter 3 begins by laying out the steps for constructing the RGOM with encoding concepts to develop the ontology in the domain of manufacturing. The ontological terminology i.e., concepts and their relations are reused into RGOM from existing ontologies and RAMI4.0.

Chapter 4 discusses benchmarking the dataset and approach to build an I40 knowledge graph to demonstrate the RGOMs adaptability through competency questions provided by the production line engineers.

Chapter 5 explores the SOTA knowledge graphs incorporating model performance on the I40 KG built from concepts defined for football datasets and experimentally obtains model prediction results alongside statistical results.

Chapter 6 presents the alignment of Resistance Spot Welding Ontology (RSWO) to the RGOMs and the details of the ontology development process. Moreover, the ontology evaluation is discussed from four different dimensions.

Finally, **Chapter 7** provides a conclusion and additional future directions for continuing research in exploring further avenues.

2 LITERATURE REVIEW

This chapter presents a review of the technologies, manufacturing companies, and service systems that have been undergoing a transformation by embodying disruptive concepts that cause a profound break with the existing concepts, leading to the creation of heterogeneous data with a wide variety of data types and formats, followed by the introduction of the semantic web as a platform for sharing knowledge and data modelling to support applications built upon it.

This chapter provides an overview of the technologies, manufacturing companies, and service systems experiencing transformation through the adoption of disruptive concepts. These concepts significantly differ from traditional practice, resulting in the generation of heterogeneous data across a broad spectrum of types and formats. Additionally, the chapter introduces the semantic web as a platform for knowledge sharing and data modelling, which supports the development of applications upon it. Furthermore, it presents ontologies as a structure for encoding knowledge into well-defined relationships as knowledge graphs for I4.0 applications, finally introducing the Reference Generalised Ontological Model (RGOM).

2.1 INDUSTRY 4.0 AND ITS REFERENCE ARCHITECTURES

Industry 4.0 (I4.0) is one of the emerging topics coined by Germany [185]. Other manufacturing countries like Japan [107] and Korea [114] have also been influenced by the concept of I4.0 by launching their related programs. I4.0 refers to the fourth industrial revolution, which is characterised by the digitization of manufacturing processes and the use of advanced technologies such as the Internet of Things (IoT), Cyber-Physical Systems (CPS), Artificial Intelligence (AI), big data analytics, and cloud computing. It involves the integration of physical and digital systems, resulting in smart factories where machinery and equipment can improve processes through automation and optimization for predictive maintenance, better quality control, and self-adaptive process optimisation with quality monitoring.

The term *reference architecture* in the context of I4.0 refers to a blueprint or a standard framework for orchestrating various components of an industrial system. It provides guidelines and standards for developing and integrating industrial systems. Reference architectures intend to solve several issues in the overall production line including the predictive maintenance [20] through sensor fusion used for the collection of data from multiple sensors to create a more complete and accurate picture of a production process in the industry environment, interoperability for system integration and system fine-tuning for the availability of data up and down the domain [30], improve the production efficiencies [12], and provides efficient scheduling services [159]. Table 2.1 shows reference architectures developed by different countries.

Table 2.1: Reference architectures and their publishers

Reference Architecture	Publisher
Reference Architecture Model for Industry 4.0 (RAMI4.0) [64]	Deutsches Institut für Normung (DIN)
National Smart Manufacturing Standards Architecture [91]	Ministry of Industry and Information Technology (MIIT) and the Standardization Administration of China (SAC)
Smart Manufacturing Systems (SMS) [95]	National Institute of Standards and Technology (NIST)
Industrial Internet Reference Architecture (IIRA) [94]	Industrial Internet Consortium Architecture (IICA)

Reference architecture comprises standards and guidelines for system development and solutions, and application architecture that are intended to furnish a plan for the world-wide utilization of standards in I4.0. Several industrial communities and enterprises such as platform I4.0 of Germany, the industrial internet consortium (IIC), and advanced manufacturing partnership 2.0 of the American Government are working together to support and develop standards to landscape reference architectures [7, 166].

2.1.1 Reference Architecture Model for Industry 4.0 (RAMI4.0)

RAMI4.0 stands for Reference Architecture Model Industries 4.0 [64], which was formed by the German Electrical and Electronics Manufacturer Association (ZVEI, VDMA, BITKOM) with the joint efforts of countries including India, Japan, and China. It is published by Deutsches Institut für Normung (DIN) as the extended version of the Smart Grid Architecture Model

(SGAM) that was designed to model the communication network entities in the field of renewable energy. RAMI4.0 is gaining worldwide agreement and is being adopted by many manufacturers to modernise their industry. The focus of RAMI4.0 is on industrial production involving discrete manufacturing to process industry as an application area.

RAMI4.0 provides a framework for companies to develop business models and future products. The main objective of RAMI4.0 is to ensure that all members follow a common framework for understanding each other in I4.0 activities and discussions. It combines I4.0 into a three-dimensional model that shows how to systematically deliver I4.0 implementation.

Each dimension of RAMI4.0 is as illustrated in Figure 2.1 as a unique part of these domains partitioned into distinct layers [117]. The corresponding dimensions of the model are described below:

1 Hierarchy Level: The Factory – The right horizontal axis is built on a standard IEC 62264 that represents four layers (from bottom to top) of system control integration known as **Control Device, Station, Work Centers,** and **Enterprise**. The two layers at the bottom are **Product** (it considers the similarity of product and production resources with their inter-dependencies during manufacturing), and **Field Device** (operating machine or devices in intelligent operation plus their sensors), and the **Connect World** at the top (partner factories collaborating via service networks) are then added to support smart factory [117]. This hierarchical structure allows for an organised approach towards data management and control across an industrial organization.

2 Life Cycle and Value Stream - The left horizontal axis represents the life cycle of products and facilities. This axis is based on the IEC 62890 standard to manage the value life cycle management of components such as orders, machines, products, and plants. It presents a difference between **Type** and **Instance**. The design and prototyping of the product are termed as a type while the completion of the type becomes an instance and the product is manufactured.

3 Layers - The left vertical axis decomposes the physical and machine assets to allow their virtual mapping. The entities in the Layers axis represent physical assets/hardware, assets integration, communications behaviour, and functional descriptions. It describes the ICT structure that demonstrates the I4.0.

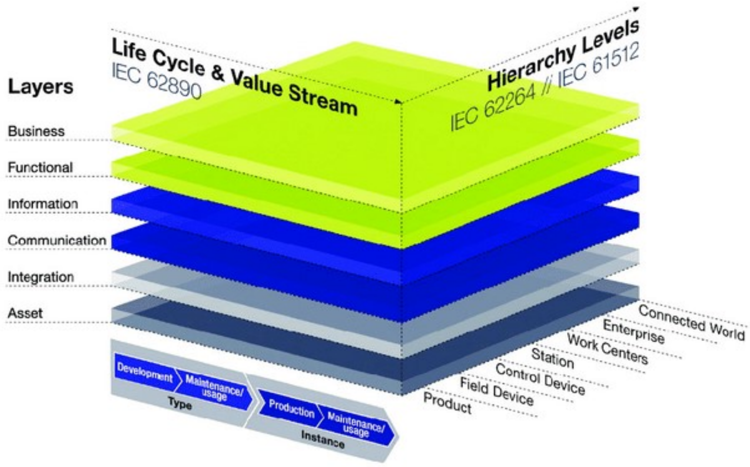


Figure 2.1: RAMI4.0 [64]

The interplay of these three axes in the RAMI4.0 model helps in developing and implementing flexible, effective I4.0 concepts. It allows the classification of industrial components, enabling the migration from traditional industrial practices to a digital, interconnected, and smart manufacturing setup [117].

2.1.2 Intelligent Manufacturing System Architecture (IMSA)

Partially influenced by German I4.0 technology, China's Ministry of Industry and Information Technology published an article defining the architecture of the National Smart Manufacturing Standards in collaboration with the Standardization Administration of China (SAC) [91].

The Intelligent Manufacturing System Architecture (IMSA) describes a 3D model intended to define the extent and application of various intelligent manufacturing technologies. The three dimensions [92], of this model are:

- **Life Cycle:** This dimension covers the entire life span of a product, from its conception and design, through its manufacturing and usage, to its eventual decommissioning and disposal.
- **System Level**: this dimension refers to the hierarchical levels within a manufacturing organization, similar to the concept described in the IEC 62264 standard. These levels can include field devices, control

systems, manufacturing processes, and business planning systems.

- **Intelligent Functions**: This dimension covers the intelligent, automated functions that these technologies bring to the manufacturing process.

The Intelligent Manufacturing System Architecture (IMSA) describes a 3D model intended to define the extent and application of various intelligent manufacturing technologies. The three dimensions of this model are: **Life Cycle:** This dimension covers the entire life span of a product, from its conception and design, through its manufacture and use, to its eventual decommissioning and disposal. **system level**: This dimension refers to the hierarchical levels within a manufacturing organization, similar to the concept described in IEC 62264. These levels can include field devices, control systems, manufacturing processes, and business planning systems. **intelligent functions**: This dimension covers the intelligent and automated functions that these technologies bring to the manufacturing process.

The IMSA has demonstrated an industrial robot as shown in Figure 2.2 and mentioned within the product life cycle manufacturing stage dimension, the equipment level of the system level, and is characterised by resource factors in Intelligent function.

The IMSA model includes an intelligent manufacturing standardization architecture landscape to guide standard classifications. It identifies the five basic types of standards required to support intelligent manufacturing. This model is proposed in the "National Intelligent Manufacturing Standards Architecture Construction Guidance," marking a crucial step in the development of a comprehensive standardization process for intelligent manufacturing. [178].

2.1.3 Smart Manufacturing Systems (SMS)

The National Institute of Standards and Technology (NIST) has defined a landscape of standards based on the Smart Manufacturing Systems that organise standards according to their functions [95]. NIST-SMS underlines the manufacturing capabilities line up with the enterprise's economic strategy. These manufacturing capabilities are categorised into four groups comprising agility, productivity, sustainability, and quality (discussed in detail [80]).

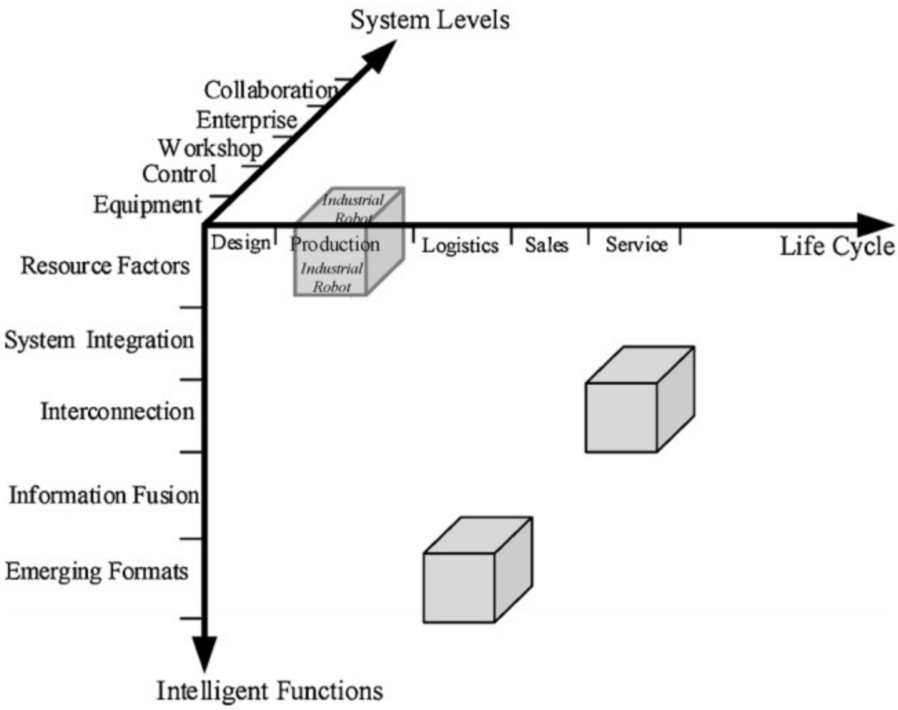


Figure 2.2: Intelligent Manufacturing Standards Architecture landscape. [92]

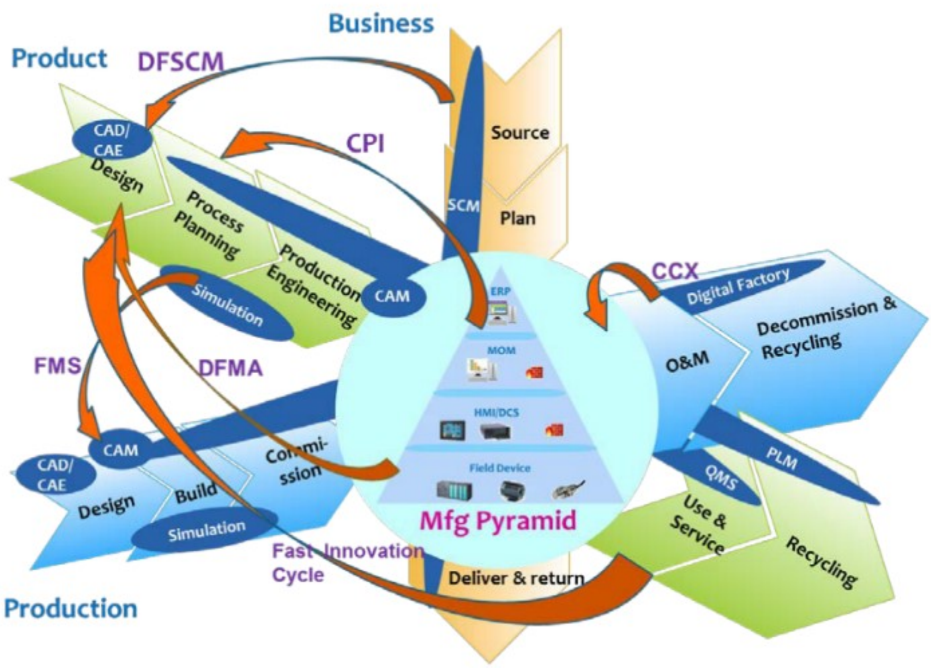


Figure 2.3: NIST Manufacturing Standards Landscape [95].

The Smart Manufacturing system arranged the standards into three dimensions product, production, and business lifecycle indicated in green, blue, and orange in Figure 2.3, respectively. The standards in these dimensions are aligned with the levels of the manufacturing pyramid of ISA95 i.e. Device to Enterprise [95].

2.1.4 Industrial Internet Reference Architecture

The Industrial Internet Reference Architecture (IIRA) is a comprehensive framework for understanding and facilitating discussions about the different components of an Industrial Internet of Things (IIoT) system [94]. The IIRA is typically split into four viewpoints as shown in Figure 2.4.

Business Viewpoint: This focuses on the business aspects of the IIoT system, including the business model, value proposition, objectives, and KPIs. This is where the broad alignment of an IIoT solution with business objectives is defined.

Usage Viewpoint: This deals with the functionality of the IIoT system, including its use cases, scenarios, and user interactions. This is essentially the operational aspect of the IIoT solutions.

Functional Viewpoint: This is where the logical architecture of the IIoT system is designed, including the functional components and their interactions. It includes design considerations for control, data communication, and applications.

Implementation Viewpoint: This involves the physical aspects of the IIoT system, including hardware, software, and network elements. This is where the physical construction of an IIoT solution comes into play.

The life cycle process, on the other hand, often relates to the various stages an IIoT solution goes through, from initial perception and design, through requirements, implementation, operation, maintenance, and ultimately decommissioning. These stages typically interact with all four viewpoints of the IIRA since business, usage, functional, and implementation considerations - all change and evolve over the life cycle of the system.

In terms of industrial sectors, the IIRA can be applied to any sector where IIoT solutions are relevant. This could include manufacturing, transportation, energy, healthcare, agriculture, and many others.

The specific details of the IIRA and its implementation will vary between sectors due to differences in business requirements, operational needs, technical constraints, and regulatory considerations, among other factors. However, the overall structure of the IIRA, with its four viewpoints and life cycle processes, remains consistent across sectors, providing a common framework for understanding and implementing IIoT solutions.

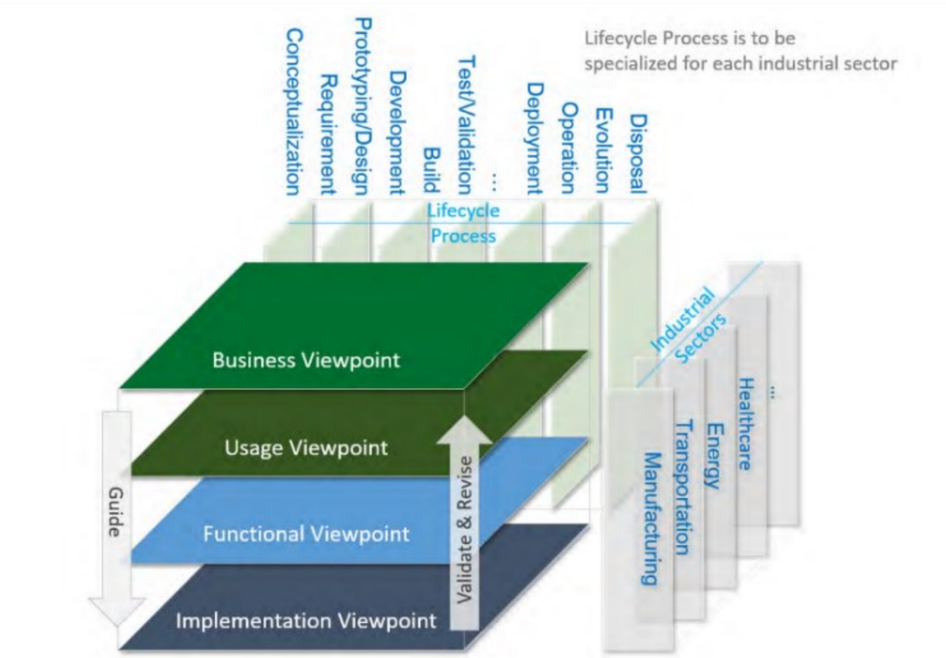


Figure 2.4: The relationship amongst viewpoints of IIRA, process lifecycle of the system, and System life cycle [94].

2.1.5 Analysis of Reference Architectures

A comparative analysis of RAMI4.0 with other emerging standard reference architectures such as IMSA, NIST-SMS, and IIRA for I4.0 is discussed in this section.

The RAMI4.0 emphasises the cyber-physical system while not addressing several characteristics such as the location of Enterprise Resource Planning (ERP), Manufacturing Execution System (MES), and the absence of digital agents and twins. However, each of the dimensions follows different standards such as IEC, ISO, and VDMA etc.; therefore, the RAMI4.0 application/implementation procedure is quite challenging[1] 4.0/Intelligent Manufacturing System Architecture; German Federal Ministry of Economic Affairs and Energy: Berlin, Germany, 2018[2].

1 German Federal Ministry of Economic Affairs and Energy. Alignment Report for Reference Architectural Model for Industry

2 https://www.dke.de/resource/blob/1711304/2e4d62811e90ee7aad10eeb6fdeb33d2/alignment-report-for-reference-architectural-model-for-industrie-4-0-data.pdf.

In general, both the IMSA and the RAMI4.0 models provide a common understanding of the structure of life-cycle, automation equipment and product by folding them into smart production. However, technically a strong differentiation between the life cycles of the two models exists. As a result, the IMSA is a variant of RAMI4.0 and thus further explores the type and instance sub-parts of life and value-stream phase by mapping them into design, manufacturing, logistics and services respectively. Additionally, it introduces a new functional element called the market[3]. Considering the exploration of the hierarchy and layers dimensions of RAMI4.0 and at the life-cycle phase of IMSA, it can be concluded that the RAMI4.0 has been closely focusing on the manufacturing lines while the IMSA has sharply dealt with life-cycle of the product.

Comparing NIST-SMS with RAMI4.0, the categorization of standards is more general. The standard published report fails to address the details of digital twins. In order to fill up the gap, the working group of ISO 23247 is currently focusing on the implementation use-cases of digital twins [96].

The interoperability among the elements tackled by IIRA at the functional level is different from RAMI4.0 [46]. In contrast to RAMI4.0, the Digital twin is more influential in the IIRA model. RAMI4.0 business layer emphasises handling the entire business life cycle, whereas the IIRA business viewpoint describes the IIoT business systems as providing interaction services among the various entities during the overall manufacturing process. The RAMI4.0 assets could either be virtual (software, agents, etc.) or physical (machines, materials, products, or personnel), which participate in the business process to provide smart production. On the other hand, the same term in IIRA refers to tangible objects of physical entities that are only being observed and controlled.

The aforementioned discussion portrays the moldability and maturity level of RAMI4.0 compared to other reference models. This model provides a baseline framework and is well aligned with other references, thus supporting the standardised protocol stacks solving the issues caused by the divergence of data [118]. However, the current I4.0 reference architecture is not able to provide enough information because it is dynamic and meets all the requirements, so a bigger reference model is needed [102].

3 https://www.plattform-i40.de/IP/Redaktion/EN/Downloads/Publikation/hm-2018-manufacturing.html

2.2 DISRUPTING TRENDS IN INDUSTRY 4.0

Industry 4.0 supports the collection of real-time data, which offers valuable insights for making intelligent operational and strategic decisions [31]. The concept of Industry 4.0 is expected to revolutionise the conventional machine-based industrial manufacturing process into a more adaptable and digitally-driven production approach [82]. In the present day, which is known as the information age, a vast quantity of data is being generated to enhance our everyday existence [39]. Every day, an enormous amount of data and communications, up to trillions, can be created and shared. So, it takes a lot of computing power to find and understand the important messages hidden in the data and service interactions made possible by disruptive technologies [85]. Disruptive technologies, collectively known as Industry 4.0, motivate innovation and enable us to transform our work practices [85]. The underlying technologies in the era of Industry 4.0 provide competitive benefits in terms of cost reduction, improvement of product quality, flexibility in operations, and better efficiency [158]. Various emerging technologies, such as artificial intelligence (AI) [90], robotics [21], blockchain [10], Big Data [49], and IoT [93], have been applied across various industries. These technologies and systems are promoting closer collaborations between humans and machines as well as machine-to-machine interactions. Thus promoting a new era of intelligent, interconnected, and dynamic industrial systems, as shown in Figure 2.5.

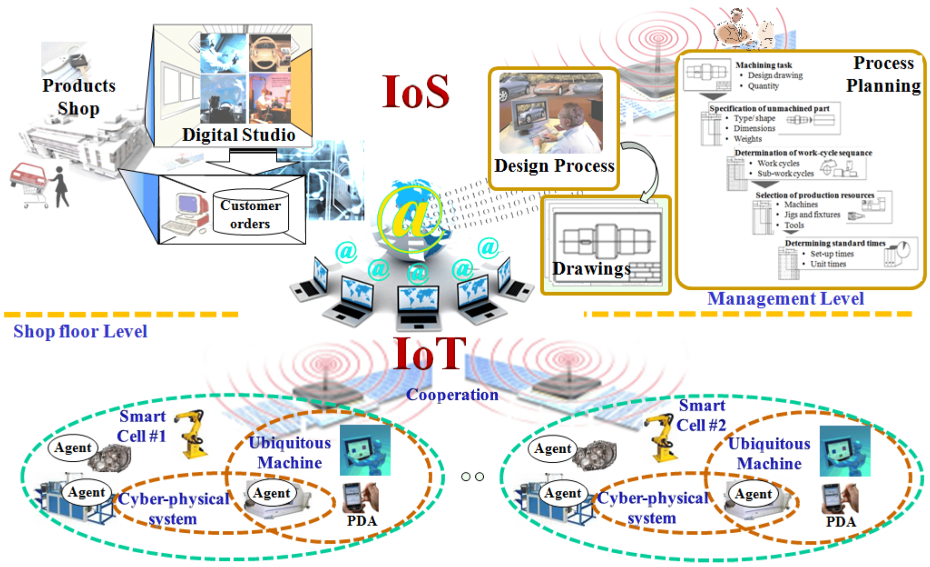


Figure 2.5: An illustration of the smart manufacturing technologies [155].

The impact of these technologies extends beyond enhancing the production capabilities of I4.0. They are redefining the very fundamentals of industrial operations and manufacturing processes. The implementation and advancement of these technologies have led to the proliferation of a multitude of field devices with diverse functionalities. These heterogeneous devices, capable of real-time communication, serve as vital nodes in the data-rich landscape of I4.0. They play a critical role in generating valuable data during the manufacturing process, presenting an opportunity to harness this data for optimising various aspects of industrial operations. However, the data generated by these heterogeneous devices is often varied in format, structure, and semantics. This presents a significant challenge to its efficient integration and utilisation. On the other hand, when harnessed effectively, this data holds the potential to enhance product life cycles, streamline on-time and on-demand productions, optimise resources, customise products, maintain machines, and reform logistic styles [123]. This underlines the importance of developing effective strategies for data integration and management that are capable of transforming this vast and varied data into actionable insights. Hence, the disruptive trends in technology are not only revolutionising the industrial landscape but also bringing forth new challenges and opportunities.

2.3 HETEROGENEOUS AND UNSTRUCTURED DATA

One of the main challenges introduced by a diverse field of devices in I4.0 is the generation of heterogeneous and unstructured data. Heterogeneous data refers to the different types, formats, and structures of data that these devices generate. For example, a sensor may generate numerical data regarding temperature or pressure, while an embedded system may produce more complex data like performance logs o error messages, etc.

This diversity is further compounded by the unstructured nature of much of the data. The unstructured data format does not follow a standard format or a predefined model, making it difficult to analyze and interpret. Examples can include text-based logs data generated by a motor of a manufacturing plant in case of experiencing some sort of vibrations. The vast amounts of heterogeneous and unstructured data generated by I4.0 technologies present a significant challenge for its effective utilization. Without proper management and interpretation, much of this valuable data can go unused and unharnessed as it stays misinterpreted, leading to inefficient operations and, accordingly, missed opportunities for optimization.

However, if effectively harnessed, this data has the potential to provide a wealth of insights into various aspects of industrial operations [150]. From enhancing product life cycles and enabling on-demand production to optimizing resources, and fine-tuning maintaining machines, the data generated in an I4.0 setup can provide actionable information. Thus, developing effective strategies for managing and analyzing this heterogeneous and unstructured data is an essential task in realizing the full potential of the I4.0 architecture.

2.4 OVERVIEW OF SEMANTIC WEB TECHNOLOGIES

The semantic web has revolutionised the existing document-based web into more intelligent systems by integrating data and web content into a more structured web environment where software agents can perform tasks more autonomously for users [129, 143].

It defines the information with metadata and semantic annotations, which allows the intelligent applications to understand the content [115]. These applications can then carry out tasks more efficiently such as decision support and queries leading to smarter services. The main technologies of the semantic web are defined in the next sections.

2.4.1 Resource Description Framework (RDF)

The World Wide Web Consortium web (W3C)[4] recommends the Resource Description Framework (RDF) as a universal data model purposely developed for the exchange of data [99]. It characterises the data in the form of triples, which consist of subjects, predicates, and objects. These triples can be combined to conceive directed graphs wherein the vertices symbolise subjects and objects, while the edges stand for predicates. The formal definition of an RDF triple is as follows:

4 https://www.w3.org/RDF/

Definition 2.1: RDF Triple. [8] *Let I, B, L be disjoint infinite sets of URIs, blank nodes, and literals, respectively. A triple (s, p, o) ∈ (I ∪ B) × I × (I ∪ B ∪ L) is denominated an RDF triple, where s is called the subject, p the predicate, and o the object.*

Figure 2.6 presents an illustration of an RDF graph that signifies the information about an RSW machine. Here, the resource `RSWO:RSWMachine` is designated as a type of Welding machine, represented as the rdf:type property that connects to relate two resources, that is, the `RSWO:RSWMachine` acting as the subject, and the `RSWO:WeldingMachine` declared as the object. Similarly, the rdf:type property also represents the resource `RSWO:Electrode` as a type of `RSWO:RSWElectrode`. Furthermore, this RDF graph demonstrates that the RSW machine comprises a part named as RSW Electrode.

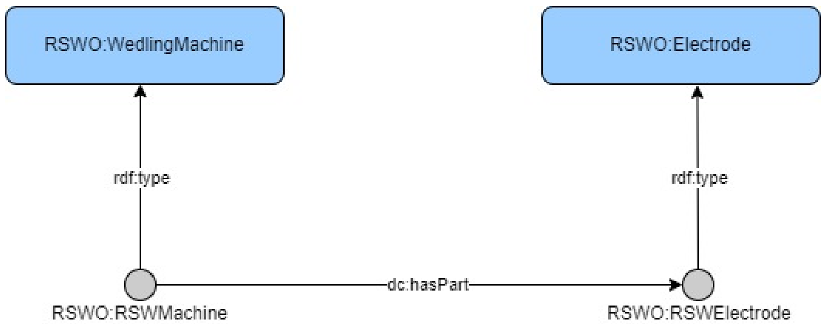


Figure 2.6: Example of an RDF graph representing the information about a Resistance Spot Welding (RSW) machine and its part Electrode.

Internationalised Resource Identifiers (IRIs) are employed to identify resources with absolute certainty. Literals, encompassing either a string with a language tag or a value with a datatype, delineate to describe specific data values. The examples employ the notation prefix:element for description; *prefix* pertains to the IRI's identification and *element* can refer to the components of RDF, namely, a subject, predicate or object. Formally, an RDF graph G is defined as a consortium of triples: $G \subset I \times I \times (I \cup L)$, where I signifies the set of IRIs and L represents the set of literals. Various formats can serialise RDF, such as RDF/XML[5], Turtle[6], RDFa[7], or JSON-LD[8]. Each characteristic format possesses unique advantages and disadvantages, that depends on the specific specific use case.

5 https://www.w3.org/TR/rdf-syntax-grammar/
6 https://www.w3.org/TR/turtle/
7 https://www.w3.org/TR/rdfa-syntax/
8 https://www.w3.org/TR/json-ld/

```
@prefix RSWO: <http://www.rswo.org/muhyah/ontologies/2022/7/
rswo> .
@prefix rdf:  <http://www.w3.org/1999/02/22rdfsyntaxns> .
@prefix dc:  <http://purl.org/dc/terms/> .
RSWO:RSWMachine rdf:type RSWO:WeldingMachine .
RSWO:RSWMachine dc:hasPart RSWO:RSWElectrode .
RSWO:RSWElectrode rdf:type RSWO:Electrode .
```

Listing 2.1: Illustration of Turtle serialization of the RDF graph in Figure 2.6

2.4.2 Ontology, RDF Schema and Web Ontology Language

An ontology is a formal specification used to describe a set of concepts and the relationships between them for a specific domain of interest [65]. It involves conceptualization, which is the simplified and abstract representation of the world [55]. When any knowledge-based system represents the world, it is committed to some conceptualization known as ontological commitment [55]. This commitment refers to recognising specific things and categories of entities that form the fundamental components of a conceptual model [55]. An ontological logical theory is a formal framework that uses logic to define and categorise the fundamental types of entities and relationships [23]. It constitutes a specific area of reality or knowledge and serves as a structured approach to understanding and representing the underlying nature and structure of a domain. The semantic web is developed to make use of an ontology that represents the information in a machine-processable structure [3]. An ontology can be defined as a structured and formal representation of a specific domain of knowledge, as follows:

Definition 2.2: Ontology. [62] *Consider* C *to be a conceptualization and* L *to be a logical language, equipped with a vocabulary V and ontological commitment* K*. The corresponding ontology* O_K *for conceptualization* C *has been defined with vocabulary V and ontological commitment* K*. This ontology* O_K *is a logical theory that comprises a collection of* L *formulas. The objective of this arrangement is to align the set of theoretical models closely with the set of intended models of the logical language* L*, consistent with the ontological commitment,* K*.*

In practice, the development of an ontology requires a consideration of the balance between the expressiveness and efficiency of language L. RDF offers a versatile language for expressing knowledge, it doesn't inherently make assumptions or establish the semantics concerning a specific application domain. The definition of domain semantics requires the use of an RDF schema, specifically, RDFS [101]. RDFS enables the development of standardised vocabulary for RDF data and defines the types of entities to which these attributes can be assigned. The RDF Schema [38] expands upon RDF by integrating constructs such as rdfs:Class, rdfs:subClassOf, rdfs:subPropertyOf, rdfs:domain, rdfs:range, to name the most significant ones. RDFS also introduces crucial annotation constructs, such as rdfs:label and rdfs:comment.

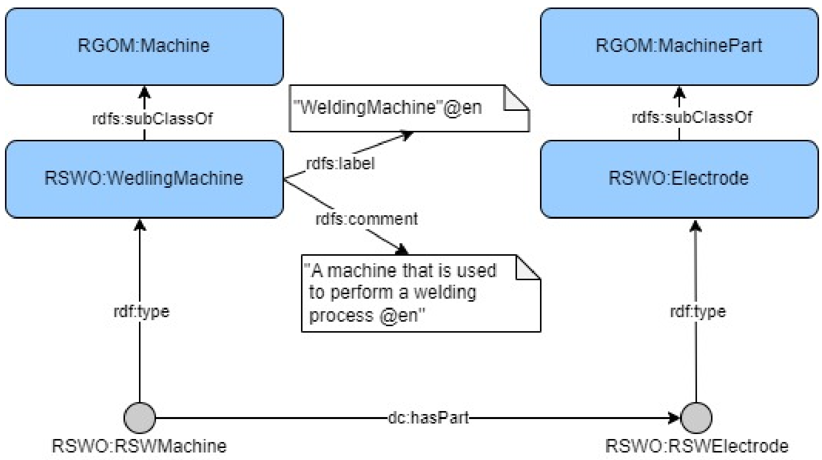


Figure 2.7: Example of an RDF graph representing the information about a Resistance Spot Welding (RSW) machine and its part Electrode.

For instance, the graph shown in Figure 2.6 can be further extended with such constructs and annotations to provide meaning to the RDF data[9].

9 https://www.w3.org/2000/01/rdf-schema

Figure 2.7 elucidates to make clear the creation of new classes, such as `RSWO:WeldingMachine`, which is a subclass of `RGOM:Machine`. This signifies that in the domain the example is modelled, all instances of `WeldingMachine` are regarded as machines. Similarly, the class `RSWO:Electrode` is a subclass of `RGOM:MachinePart`. The utilization of annotation properties such as `rdfs:label` and `rdfs:comment` can also be discerned and recognised from the figure. Moreover, the property `dc:hasPart` has the class `RSWO:WeldingMachine` as its domain (i.e., `rdfs:domain`) and the class `RSWO:Electrode` as its range (i.e., `rdfs:range`).

Thus, an ontology consists of two key components: terminological components (Tbox), which define concepts, and assertional components (Abox), which indicate the instances of these concepts. Ontologies also offer an automatic reasoning process that retrieves axioms that are not explicitly incorporated into the knowledge graph.

2.4.3 Knowledge Graphs

The term Knowledge Graph (KG) has recently gained significant attention through the tech industry such as Google, Facebook, Amazon, Netflix, and others [67]. The term KG was first introduced by Google in 2012 [63] to use semantic knowledge in web searching. It is also used to denote Semantic Web knowledge bases, including DBpedia, Wikidata, and YAGO. KGs make use of numerous knowledge representation styles, which span beyond solely RDF to embrace abstract modelling languages and probabilistic techniques [104]. In these graphs, the essence of the information is stored along with the data, usually taking the form of ontologies. This feature makes KGs self-explanatory, positioning them as a comprehensive source for both locating and understanding data.

In KGs, the semantics of the data are explicit and comprise of formal methods that assist in inferencing. KGs contain a large number of entities and provide definitions of key concepts and relationships. They also suggest ways to modify data to fit model specifications and provide the ability to draw conclusions and discover new information from existing data [28]. Furthermore, KGs have shown their effectiveness in resolving semantic interoperability challenges during the process of data integration in various fields, including health [27], agriculture [171], banks [1] and many other areas. In factory environments, KGs are viewed as the cornerstone for the next wave of enterprise information systems. A KG can be defined as:

Definition 2.3: Knowledge Graph. [62] *A labelled directed graph represented the RDF data model. Let I and V be the sets of URIs that correspond to the entities represented in the RDF documents and terms from ontologies, respectively; and L be a set of entities representing Literals.*

A knowledge graph thus acquires in order to integrate the information into an ontology that enables a reasoner to derive new knowledge. The knowledge graph is referred to as data organised by the following ontologies.

2.4.4 SPARQL Language

SPARQL, endorsed by the World Wide Web Consortium (W3C)[10], serves as a query language designed to retrieve and manipulate data encapsulated in RDF. Drawing its foundation from the RDF Turtle serialization and graph pattern matching, SPARQL queries data structured as RDF triples, incorporating variables for the subject, predicate, and object to form graph patterns. A typical SPARQL query thus comprises triple patterns, conjunctions, disjunctions, and optional patterns. These triple patterns bear a resemblance to RDF triples, where the subject, predicate, and object serve as variables in the query.

```
PREFIX RSWO: <http://www.rswo.org/muhyah/ontologies/2022/
7/rswo>
PREFIX dc:  <http://purl.org/dc/terms/>
PREFIX rdf:  <http://www.w3.org/1999/02/22-rdf-syntax-ns>
PREFIX rdfs:  <http://www.w3.org/2000/01/rdf-schema#>
SELECT ?machine ?part ?class
WHERE {
?machine dc:hasPart ?part .
?machine rdfs:subClassOf ?class . }
```

Listing 2.2: Example of a SPARQL Query 2.6

10 https://www.w3.org/TR/sparql11-query/

Within the query, variables function as placeholders, which construct the solutions when paired with RDF terms. Listing 2.2 showcases key aspects of the SPARQL language. The query utilises the prefixes to query a dataset. The goal of the query is to fetch the information about machines, their associated parts, and the class to which they belong. The `SELECT` clause specifies three variables: `?machine`, `?part`, and `?class`. In the `WHERE` clause, the query matches triples where a machine is linked to its parts using the `dc:hasPart` property, and where a machine is a subclass of another class using the `rdfs:subClassOf` property. The result set will include these variables, providing details about machines, parts, and associated classes.

2.5 SEMANTIC MODELING FOR INTEGRATING UNSTRUCTURED DATA

The semantic web has been extensively employed to represent the domain knowledge with an aim to address the semantic heterogeneity conflict in I4.0. The problem of integrating data has been addressed by several researchers. Ontology-Based Data Integration (OBDI) is a widely used technique for addressing this issue [38, 108]. OBDI methods are commonly employed for integrating semantic data due to the semantic representation provided by ontologies. The OBDI approach typically consists of three key components: i) an ontology that represents the domain's knowledge, ii) a data source containing the domain's data, and iii) mappings connecting the two components [38]. Cruz et al. [35] explore different perspectives on the use of ontologies for semantic data integration: i) the single ontology approach, where all sources are directly linked to a shared global ontology; ii) the multiple ontology approach, where each data source is described by its local ontology separately; and iii) the hybrid ontology approach, which combines the single ontology approach for describing each data source in the domain with mappings to a generally shared ontology. Other studies focus on the essential dimensions of mapping development. These dimensions include: i) discovering mappings among ontologies; ii) representing the mappings declaratively; and iii) reasoning with the mappings. Mappings play a crucial role in linking two ontologies that represent the same domain and address semantic heterogeneity conflicts between them [108]. The authors proposed an approach called SODIM, which combines semantic data integration techniques with service-oriented principles [2]. SODIM aims to improve the efficiency and effectiveness of data integration processes by leveraging semantic annotations and automated mapping generation. Rahm [124] discusses the need for a comprehensive data integration approach that can handle multiple sources, presenting six use cases where holistic data integration is applied. LDIF [142] introduces a large-scale framework for integrating Linked Data, using a mapping language and the R2R framework [15] to translate data from different vocabularies to a local target vocabulary. LDIF also utilises the SILK framework [68] to address heterogeneity conflicts and provides data quality assessment. Collarana et al. [33] propose MINTE, a framework that integrates data from diverse sources into a knowledge graph. MINTE utilises RDF molecules to represent data meaning and fusion policies to resolve semantic heterogeneity conflicts.

2.6 SEMANTIC MODELING FOR I4.0 DATA INTEGRATION

The manufacturing industry is transforming mechanization towards intelligent and digital processes. This shift is driven by the integration of technologies such as the IoT, sensors, and Cyber-Physical Systems (CPS), all of which are vertically integrated into a smart factory [88, 168]. Thus, these resources and processes generate a huge amount of heterogeneous and unconnected data which requires effort and time. Following sub-sections report the literature work in this regard.

2.6.1 Semantic Representations of I4.0 Resources using Onotlogies

Semantic modelling of smart factories, manufacturing production lines, and manufacturing systems interoperability is the crucial feature established in the I4.0 production between tangible assets including systems, devices, sensors, etc., connected to each other over the internet. In the I4.0 context, stress has been placed on the alignment of manufacturing systems, processes, and reconfiguration of resources in the production line. In the last decade, there have been numerous efforts to represent the domain knowledge of I4.0 in the form of modular ontologies, that is, resource, device ontology, process ontology, predictive maintenance ontology, etc., to meet manufacturing production requirements [30, 53]. There have been rigorous efforts to develop ontologies that aim to semantically model the manufacturing production line very clearly. Buchgeher et al. conducted a survey on the role of knowledge graphs in production and manufacturing [24]. They have reported the biblio-metric facts, type of research, statics and application scenarios of the knowledge graphs in manufacturing and production.

I4.0 for pharmaceutical products in re-configurable form has been proposed to adjust the increasing requirement of flexibility, agility, and low cost in the health sector [162]. The re-configurable form of I4.0 is comprised of three layers, namely: 1) executing, 2) deployment, and 3) perception layer. The knowledge graph as employed in the perception layer is representing the semantics of manufacturing based on the MASON Ontology responsible for scheduling the production plan. In the deployment layer, IEC61499 standard is implemented for modelling functionality and controlling machines. The feasibility of the proposed approach is validated by taking a use case of drug packing based on demand. Kovalenko et al. proposed AutomationML ontology to represent the semantic modelling of cyber-physical systems covering data exchange in an I4.0 scenario [84]. The semantic-based representation of I4.0 devices in the administration shell provides the integration, identification, data availability, and so on, of the devices [58, 57].

The Semantic Manufacturing ontology highlights the sequence of processes and machines required for an ordered workpiece product [116], Turtle file is available online[11]. Mazzola et al. proposed CDM-Core[12] ontology by re-using the existing domain and core ontologies [103]. The authors claimed it to be the largest publicly available global ontology. However, they have focused more on the service-oriented architecture and monitoring of the manufacturing services. There is no explicit information regarding the modelling of the manufacturing factory and the main concepts such as type of processing, type of machine, etc., are missing.

Manufacturing systems should be able to incorporate and assist humans (operators, technicians). Humans are participating in the environment of automated systems and it is necessary to consider the role of the operator in such an environment. Ferrer et al. proposed the addition of the skills

11 http://i40.semantic-interoperability.org/smo/smo.ttl
12 http://sourceforge.net/projects/cdm-core/

and tasks performed by humans in manufacturing ontology that is using the CPS knowledge repositories [48]. Their work presented a semantic model that allows the operations modelling achieved by human operators. However, they focused more on the service of orchestration during the production plans. Ahmad et.al proposed the integration of manufacturing domain data such as Product, Process, and Resource (PPR) using the ontology approach for matching the product requirements in assembly automation [47]. The mapping information of PPR helps in deriving the processes and resources required to manufacture the designed product.

The authors proposed an ontology by merging five ontologies which are base, product, process, device, and parameter ontologies to represent the manufacturing production process beginning from order to completion of the product [30],. The ontology is built on top of the product, process, device, and parameter ontologies to provide interaction with each other. Additionally, the order concept is modelled as a separate ontology that is linked with the product. Service-oriented architecture has been built on top of this ontology model to discover, select, organise, and consume semantic web services dynamically [29].

Seyedamir et al. utilised the concepts of manufacturing resource, process, and product from the ISA-95 standard [144]. They adopted the approach of semantic rules to infer implicit knowledge to allow inspecting the machines needed to produce product variants. Saeidlou, S. et al. designed an ontology model for the manufacturing domain and developed a semantic query algorithm to investigate the semantic richness of the queried keyword return by the ontology model [134]. Kalaycı et. al proposed a Semantic Integration of Bosch data (SIB) framework to integrate Bosch manufacturing data to analyse the surface mounting process pipeline [76]. To experiment with their framework, they have developed surface mounting (SMT) to map the production line data.

Some of the most renowned ontologies in the manufacturing domain are process specification language (PSL) [60], ONTOlogy for Product Data Management (ONTO-PDM) [113], MAnufacturing Semantic ONtology (MASON) [89], ADAptive holonic COntrol aRchitecture (ADACOR) [18], etc., ontology. MASON ontology has been developed to estimate the production cost of mechanical components. The design of PSL ontology emphasises enabling the exchange of process information in manufacturing systems accurately and comprehensively. Panetto, H. el. al modelled the product concepts based on two standards ISO-10303 and IEC-62264 to facilitate the interoperability between software applications exchanging product life cycle information. PSL ontology represents the concepts of process modelling, planning, scheduling, simulation, etc. in axioms of first-order logic theories. ADACOR ontology has highlighted the knowledge related to customer work orders, production plans, and model operations. These ontologies are helpful to recreate an ontology model to cover the notion of the whole production line from customer order to the product life cycle. There is a great amount of literature available for ontology-based agent systems such as CORA [122], ROA Ontology [109], ORArch, and O4I4 Ontology [86] that perform main tasks.

Overall, the current research lacks common desirable features for I4.0 manufacturing production line ontology. Firstly, they are tailored to represent a specific resource of the production line and lack comprehensive coverage of the I4.0 reference architecture. This in turn limits the depth and semantics of the manufacturing ontology to a specific use case. Secondly, they are typically developed without adherence to ontology design methodologies as the best-used practices, such as incorporating design patterns or reusing established vocabularies. Lastly, these ontologies are often not readily accessible or available for comprehensive reuse. Thus, we note here a room for improving semantic representations of manufacturing production lines by proposing improvements using the Reference Generalized Ontological Model (RGOM) framework as stated in the **RQ1**

2.6.2 Integrating Production Line Data into Knowledge Graphs: Data Availability based Perspective

Ontology evaluation plays a crucial role in assessing the quality, usability, and effectiveness of ontologies in achieving data integration and interoperability across different resources. Once an ontology is developed, the data is populated within it to evaluate its performance and suitability for the specific domain it was created for [4]. This evaluation helps ensure that the ontology effectively represents and organises the data, enabling its seamless integration and utilization.

Considering the domain of the I4.0 resources, Ramírez Durán et al. developed a semantic model (ExtruOnt) to describe the knowledge of a manufacturing machine known as an extruder machine that executes the extrusion process [126]. Though the scope of ExtruOnt is confined to a specific domain, and provides information about extruder components, three-dimensional representations of components and spatial connections, features, and sensors capturing data about machine performance. The authors thanked Urola Solutions for providing them with real data to evaluate the ontology. Grangel-González et. al concatenated domain ontologies on top of SMT ontology to accomplish the interoperability issue in manufacturing data [59]. The ontology developed is more focused on the mounting process and is evaluated on data from Bosch.

The authors examine the standards landscape of I4.0 from a semantic integration perspective and developed Standards Ontology (STO) [56]. It focuses on the integration of standards within the context of I4.0 and analyzes their semantic interoperability. They provide insights into the current state of standards and highlight the importance of semantic integration in achieving interoperability in I4.0. No information is available about the dataset used for their ontology evaluation.

Wan et al. proposed a resource configuration-based ontology describing the domain knowledge of the reconfiguration of sensible manufacturing resources using Web Ontology Language (OWL) [163]. The objective of their work is to integrate the CPS equipment through ontology-based resource integration architecture. The generated data are stored as a relational database and are associated and mapped into the model instances of the manufacturing ontology. The proposed ontology for resource reconfiguration is examined using an intelligent manipulator as a use case that verified its manufacturing feasibility. Manufacturing Resources Capability ontology has been proposed to describe the capabilities of the production system resources [71]. The ontology development process followed the five stages of ontology engineering methodology that are feasibility study, kick-off, refinement, evaluation, and usage and evolution. According to Jarvenpaa et al. Manufacturing resource capability ontology (MaRCO) is used by resource vendors to represent the capabilities of resources they are offering and publish it in the digital marketplaces or global resources list and is browsed by production companies or systems integrators when reconfiguring existing or designing new manufacturing systems [71]. MaRCO aims to provide the matchmaking between the required capabilities of a resource and the production requirements of a product.

Teslya et al. proposed an ontology-based approach to describe the industrial components merged from four different scenarios in order to form upper-level ontology [152]. Such a union will enable change in the created

Table 2.2: List of the ontologies with their research focus and datasets being used for evaluations.

Paper	Ontology	Research Focus	Dataset
[126]	ExtruOnt	Describing extruder components, 3D representations, and spatial connections, features, and sensors capturing data.	Data were taken from the extruder manufacturing factory.
[59]	SMT ontology combined with Domain ontologies	To achieve interoperability in I4.0.	Data taken from Bosch, no information available.
[56]	Standard Ontology (STO)	Solving interoperability issues between the analogous standards used by reference architectures.	X
[163]	Resource reconfiguration ontology	Integration of intelligent manufacturing equipment using resource configuration ontology.	Populated the ontology with the data produced by the manipulator using raspberry pi.
[71]	Manufacturing Resource Capability Ontology (MaRCO)	Development of resources ontology to describe manufacturing resources capabilities.	Data were taken from the Industrial laboratory Demonstration setup
[13]	Product, Process, Resource	Integration of Product, Process, and Resource	Festo Modular System (a testbed for an industrial test.)
[75]	Process	Decomposed the sentences of RAMI4.0 standards, architectures , and models into concepts map to integrate the processes of I4.0.	X
[152]	Components of Socio-Cyber Physical systems	Establishing a specific information space to connect all the production components.	X
[58]	I4.0 components	Semantically represented the I4.0 devices in administration shell	\url{https://cdd.iec.ch/cdd/iec61360/iec61360.nsf}
[29]	I4.0 Demonstration Production line	Modelled the I4.0 production line	X
[116]	Semantic Manufacturing Ontology (SMO)	Modelling of Smart Factory	X
[144]	Modular Ontologies (ISA-95)	Modelling Smart Factory	Data produced on FASTory simulator \url{http://escop.rd.tut.fi:3000/fmw}
[76]	Surface Mounting Process (SMT Ontology)	Integration of Bosch Manufacturing Data for analysis	Data taken from Bosch, no information available

business process to boost product customization for the customer and reduce the cost for its producers. The semantic-based representation of I4.0 devices in the administration shell provides the integration, identification, data availability, etc., of the devices [58, 57]. Some other research works such as [29, 116, 144, 13, 76] are mentioned in the Table 2.2 that provide the information about their research focus and the dataset for ontology evaluation.

The current research has assessed their ontologies using datasets that are not publicly accessible for result reproduction or evaluation of other ontologies. These datasets are often specific to a particular use case, which is typically either based on synthetic generated data by researchers or derived from private company data in an industrial context that is not available online. Therefore, as defined in **RQ2**, a publicly available benchmark Resistance Spot Welding Ontology (RSWO) dataset is capable of demonstrating the adaptability and effectiveness of RGOM in real-world production data.

2.6.3 Alignment of Ontologies with Domain Level: Use-Case

The alignment of ontologies with the domain level is a crucial aspect when addressing the challenges in the integration of Resistance Spot Welding (RSW) data within the broader context of manufacturing or welding domains. Among the RDF resources related to RSW, there exists a significant variation in the vocabularies used and the underlying data models applied. This variation poses a hindrance to achieving seamless interoperability and effective integration of RSW data. However, this problem can be effectively tackled through the process of alignment, specifically through post-alignment techniques such as terminology mapping [136] or semantic transformation between different data models [111].

In recent works, several ontologies have been proposed for addressing the challenges in resistance spot welding (RSW) data integration and semantic inconsistency. Sarkar et al. [138] developed an ontology based on foundational concepts to characterise joining operations. Saha et al. [135] introduced the Core Domain Ontology for Joining Processes (CDJOP) to categorise joining processes and tackle semantic inconsistency in standardization documents. However, their work lacked input from domain experts, real industry implementation, and omitted important concepts like squeeze time and spatter occurrence. Kim et al. [81] utilised an automotive OEM dataset to extract decision rules and transform them into SWRL rules for improved knowledge domain shareability. Nevertheless, their reliance solely on data-derived rules limited the interpretability of semantics. Solano et al. [146] formalised welding process knowledge through ontological modeling but focused on a limited scope of welding categories and lacked representation of procedure details and machine settings. Other works have explored machine learning models enhanced with ontologies for predicting welding quality [148, 182, 149]. Dong et al. [43] transformed online unstructured data into machine-interpretable data using the WeldGalaxy Ontology, although their ontology had insufficient object properties and lacked proper class connections. Furthermore, none of the work has demonstrated the development of their ontology with an ontology development process.

In addition to the aforementioned works, it is important to highlight the significance of aligning the proposed ontologies with the domain level. Alignment with the broader domain ontology ensures interoperability and effective integration of resistance spot welding (RSW) data within the context of manufacturing. This alignment addresses the variation in vocabularies and data models used among RSW-related RDF resources, enabling seamless data exchange and knowledge representation. By aligning the ontologies, a common understanding can be established between RSW-specific concepts and the broader domain-level ontology, bridging semantic gaps and facilitating effective data integration and interoperability within the Industry 4.0 framework. Therefore, there is a need for the development of RSW using an ontology development process. Additionally, the ontology is required to be aligned with a domain-level manufacturing ontology to the application-level ontology RSWO, utilizing the RGOM as defined in **RQ3**.

2.6.4 I4.0 Based Knowledge Graphs Completion

Knowledge Graphs (KG) have attracted a lot of attention from the research community over the past few years. They are currently being adopted in many domains, such as question-answering systems, information retrieval and recommendations in different domains, for instance, the supply chain system [83], the automotive industry [125] and industries on the immediate list of industry 4.0. As reported in the literature, the current research on I4.0-based KG is carried out in two dimensions: (i) techniques for building KGs [51], [40], and (ii) applications of KG [72], [179], [154]. To be more specific, regarding the first dimension, the current techniques used to build KG focus on integrating data from heterogeneous sources, but most of the time, this results in imperceptible missing links between the graph entities [59]. As a consequence of the missing links within the KGs, it cannot be exploited for the aforementioned applications in conjunction with other powerful tools such as predictive maintenance, the prediction of the remaining useful life of complex systems, and product quality monitoring, among others. Moreover, the I4.0 data-based KGs are mostly prone to missing links. Analyzing and predicting the missing links in such KG is nearly impossible with human heuristics, and is highly dependent on the power of using relevant algorithms [97].

The term *link prediction* refers to determining the likelihood of identifying pairs of nodes in a graph that will form a link or will not establish a link in the future. Graph-based link prediction research has witnessed a number of prediction models proposed using different architectures and approaches [133]. The proposed models are based on learning the features of KG to predict links better than the previous ones [167]. Moreover, every model is built on different relational features such as relations, path information, and substructure information for training to improve the link prediction [165].

The first category among the link prediction models is the geometric-based (aka translation) model. It uses a spatial transformation for relation embeddings in the latent space. Provided a fact, a spatial transformation is used to represent the head embeddings where the values of relation embeddings are parameterised. Distance functions such as the L1 norm and L2 norm are employed to compute an offset between the resulting head and tail vectors. The additional constraints in spatial transformation make the geometric models unique from those of Tensor decomposition. Some of the examples of geometric models are TransE [17], CrossE [181], TorusE [44]. Deep learning models are the second category of link prediction models. Deep learning models employ convolutional neural networks (CNN) to learn features using weights and biases as estimators. These estimators are then combined with the input facts to extract features of significant importance. There are several different deep learning architectures reported in literature [61]. However, their fundamental components are very similar. A neural network layer consists of three basic layers, namely, convolutional, pooling, and fully connected layers. The input feature set is represented in the convolutional layer that consists of a number of convolution kernels that are used to calculate various feature maps. Each neuron in a feature map is specifically linked to an area of nearby neurons in the layer underneath it. In the previous layer, this area is known as the neuron's receptive field. By first convolution, the input with a learnt kernel and then using the convolutional results to apply an element-wise nonlinear activation function, the new feature map may be produced. The kernel is shared by all spatial locations of the input to generate similarly produced feature maps. Several different kernels are used to create the entire feature maps [61]. The final feature map is passed through a fully connected layer to compute the fact score. ConvE [41], ConvKB [106], ConvR [73], CapsE [160] are some of the deep learning models proposed on the aforementioned notion.

The design of industrial KGs is different from the benchmark datasets that are commonly used. The industrial KGs are Hierarchical structures and the nodes are densely connected with other. Therefore, as defined in **RQ4**, there is a need to explore state-of-the-art link prediction models such as TransE, DistMult, ComplEx, ConvKB, and ConvE on the KGs developed from the Benchmark dataset, in order to evaluate the effectiveness of RGOM using heterogeneous and unstructured data by standard metrics such as Mean Reciprocal Rank (MRR) and Hits@N.

2.7 RESEARCH GAPS

Digitalization when coupled with AI has been offering an amazing and unprecedented acceptance almost in every domain of life. The digitalization drive aims to connect humans and machines through the internet, resulting in generating huge amounts of data. Hence, data science is bound to assume a position central to many of the research challenges today.

These challenges include improved methods of gathering valuable machine data from across sources of heterogeneous and unstructured kinds, demanding implementation of advanced data analytic technologies and methods, in order to provide the right information to the right person at the right time by visualising data through associating it with relevant semantics.

Our proposed approach is conducting a comparative study of the current ontologies which existing vocabularies are being used before they are reused with the additional concepts that were missing, making the whole process performed in an iterative manner, self-adapting for adjusting to the current requirements. Furthermore, it is observed that none of the ontology is available that can be used as domain manufacturing ontology for alignment purposes. Moreover, another issue is the lack of availability of benchmarked manufacturing production line datasets that can be used to validate these semantic models. Furthermore, how a specific ontology can be aligned with RGOM in line with the ontology development process also needs to be studied. Finally, the SOTA algorithms can complete the KGs formed from industrial data that need to be investigated.

2.8 SUMMARY

In this chapter, the aspects related to I4.0 within the scope of this thesis and its reference architectures are presented. We have explored reference architecture models of RAMI4.0, IMSA, SMS, and the IIRA, providing an analysis of their key features. We then moved on to disruptive trends in technology that uses AI, the Internet of Things and cloud computing, highlighting the challenges posed by heterogeneous and unstructured data. An overview of semantic web technologies is presented, including RDF, ontologies, RDF Schema, Web Ontology Language, knowledge graphs, and the SPARQL language. The application of semantic modelling as a vision for linking data across web pages, applications and files for integrating unstructured data, for I4.0 data integration has been discussed. This includes semantic representations of I4.0 resources using ontologies, integration of production line data into knowledge graphs from a data availability perspective, alignment of ontologies with the domain level through a use case, and completion of I4.0-based knowledge graphs. Finally, we have addressed the research gaps in the field, identifying grey areas that need to be highlighted by requiring further exploration and investigation. The challenge of harnessing the vast amount of un-utilised data by proposing the Reference Guide for Ontology Development (RGOM) is being addressed in the next chapter.

3 REFERENCE GENERALIZED ONTOLOGICAL MODEL (RGOM)

This chapter provides details about the development of the Reference Generalized Ontological Model (RGOM). It includes the process of ontology development and the use of RAMI architecture with a reference to explaining the I4.0 concepts and the RGOM ontology itself.

3.1 STEPS TOWARDS BUILDING RGOM

The methodology for the proposed Reference Generalized Ontological Model (RGOM) is composed of the following steps (see 3.2).

- A detailed survey is conducted by analyzing recent literature for ontological models of I4.0 major ontologies to identify and shortlist terms related to the production line. The survey was conducted based on three steps (bottom to top) including (i) planning and scope of the review, (ii) filtering of the papers for review, and (iii) reporting the review [156]. In the first stage of the methodology, the scope of the review is set to determine the literature's relevance to the semantic web and knowledge graphs in I4.0. This stage involved the identification of the most suitable keywords to select the articles. As a result, this stage provided an initial step with searching different databases such as ACM digital library, IEEE Explore, Science Direct, Google Scholar and Scopus which resulted in almost 164 articles, in total. These articles include academic as well as industry publications containing conferences, workshops, letters, journals and peer-reviewed books. In reporting the literature, we included only full-text work based on an ontology proposal as well as the construction of a knowledge graph for smart manufacturing. In the second stage, an advanced filtration was adopted by considering the different versions of the selected ontologies in conjunction with the combinations of the titles and abstract which resulted in the selection of more specific articles of 110. In line with the ontologies selected version the titles and abstract of each research paper were studied to identify its relevance for inclusion. The filtration process was carried out using the following steps.

- The most relevant ontologies covering reference architectures, manufacturing production line, predictive maintenance and supply chain concepts of I4.0 were captured.
- The study elaborated all versions of the chosen ontologies for understanding their functional behaviour and its adaptation in the study.

The third stage is reporting the review and is composed of two steps. In the first step, a full-text reading approach was adopted to further narrow the search and obtain 87 articles. This step excluded all those papers summarizing the work on the Semantic Web or Knowledge Graph in Smart Manufacturing. In the second step of reporting the review, a total of 51 papers were found relevant to be included in the study. Each round contains articles that were affirmed to be relevant in the previous round. The overall survey methodology adopted in this work is summarized in Figure 3.1.

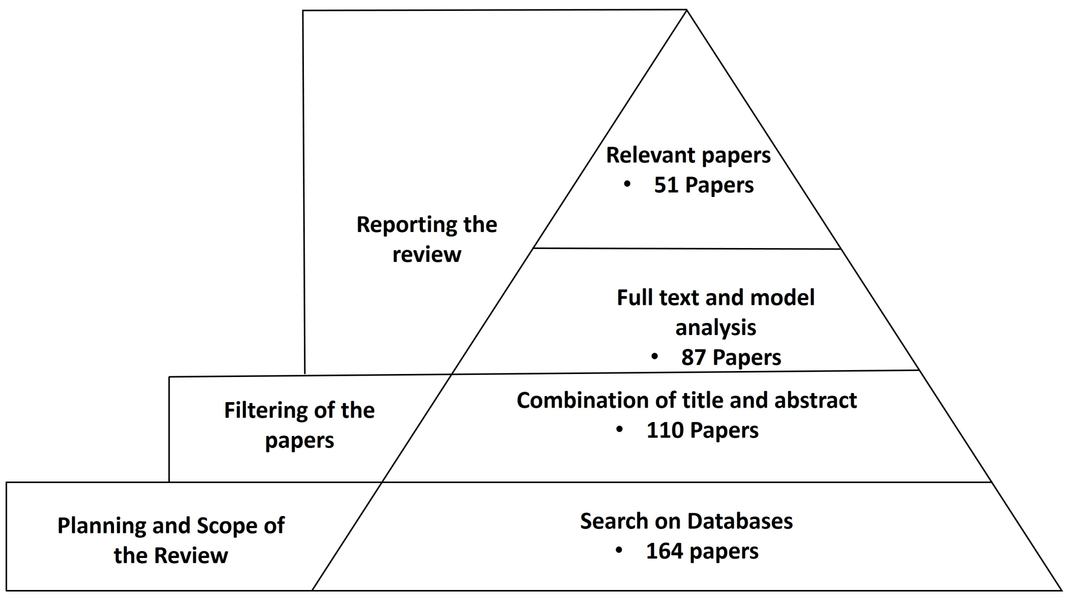


Figure 3.1: An illustration of the methodology adopted for conducting the survey.

- I4.0 architecture, such as the reference architectural model Industry 4.0 (RAMI 4.0), has been studied to find out the requirements needed

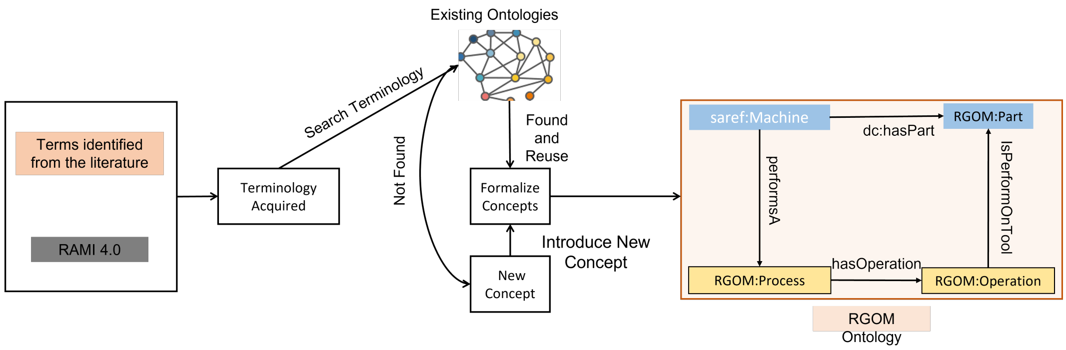


Figure 3.2: Steps towards building RGOM

for I4.0 production. Some of the RAMI requirements identified are detailed as follows:

1. **Connectivity and Networking:** Industry 4.0 systems require robust and reliable connectivity among machines, devices, and systems. This includes technologies such as the Internet of Things (IoT), wireless communication, and network infrastructure to enable seamless data and communication exchange.
2. **Data Collection and Analytics:** Industry 4.0 relies on the collection of vast amounts of data from various sources within the production environment. This data is analyzed using advanced analytic techniques, including machine learning and artificial intelligence, to derive actionable insights for process optimization, predictive maintenance, and decision-making.
3. **Interoperability and Integration:** To achieve the vision of Industry 4.0 in which systems and components must be interoperable, allowing seamless integration and communication across different machines, devices, and software platforms, while noting that Standards and protocols play a crucial role in ensuring compatibility and interoperability.

- A comparative study is then conducted to find out the gaps between the standards and the current state-of-the-art models. During this step, it is identified that the current ontologies do not follow the requirements of RAMI4.0 and are unable to follow the reuse principle of linked open data.
- The existing vocabularies have been reused with the additional concepts that are considered missing. The whole process has been performed iteratively.

3.2 RGOM DEVELOPMENT PROCESS

The process to develop the RGOM is adopted from the Linked open term methodology [120]. The reason to chose this methodology is due to its ability to offer a step-by-step refinement process during the creation of the ontology. This detailed refinement process is critical as it ensures that the ontology accurately encapsulates the domain-specific concepts prevalent in manufacturing industries. Figure 3.3 shows ontology development process steps which are described below.

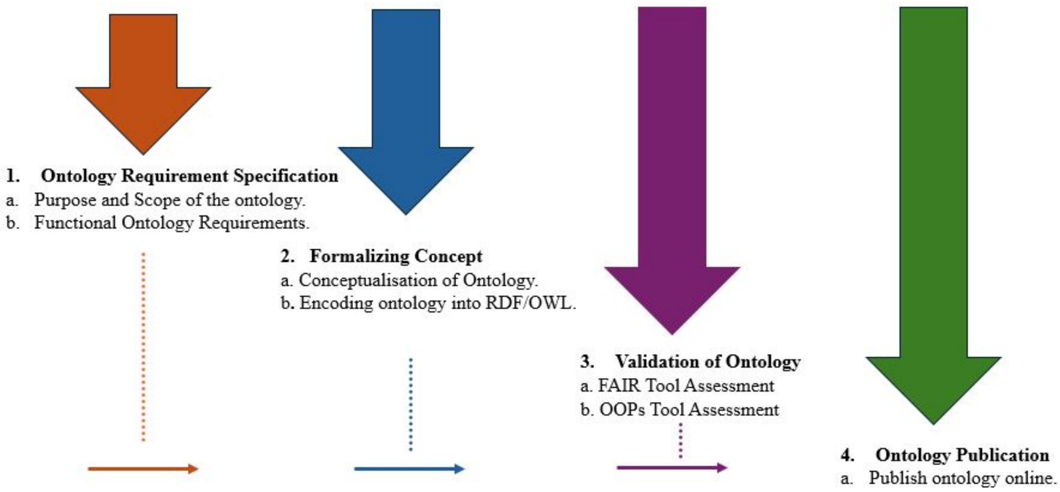


Figure 3.3: Ontology development process

3.2.1 Ontology Requirement Specification

It is the first step of the ontology development process. It is identified from the literature survey that the current ontologies are unable cover important concepts in a manufacturing industry as shown in Table 3.1

This limits the purpose and scope of existing onotlogies to specific use than generic. For example, the knowledge such as machine consume power can not be deduced [30, 56, 59]. Considering the scope of the RGOM, the functional requirements are defined that must be answered by the ontology. The natural language sentences are then transformed into related competency questions. Table 3.2 lists some of the competency questions.

Table 3.1: Manufacturing production line concepts covered by different research articles. The No and Yes in the rows indicate whether these concepts are absent or present, respectively, in the referenced articles.

Article	Sales	Manufacturing Production Line						
		Device	Operator	Process	Product	Time	Sensor	Material
[56]	No	No	No	No	No	No	No	No
[57, 58]	No	Yes	No	No	No	No	No	No
[70]	No	Yes	No	Yes	Yes	No	No	No
[75]	No	No	No	Yes	No	No	No	No
[89]	Yes	Yes	Yes	Yes	Yes	No	No	Yes
[126]	No	Yes	No	Yes	No	Yes	Yes	No
[127]	No	Yes	Yes	Yes	Yes	No	Yes	Yes
[152]	Yes	Yes	No	Yes	Yes	No	Yes	Yes
[161]	No	Yes	No	Yes	No	No	No	No
[30]	Yes	Yes	Yes	Yes	Yes	No	Yes	No
[116]	Yes	No	No	No	No	No	No	No
[140]	No	No	No	No	No	No	No	No
[53]	No	Yes	Yes	Yes	Yes	Yes	Yes	No
[144]	NO	Yes	NO	Yes	Yes	Yes	No	Yes
[76]	No	Yes	No	Yes	Yes	Yes	No	No
[59]	No	Yes	No	Yes	Yes	No	No	Yes
RGOM	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Table 3.2: List of competency questions

Serial No	Competency Question
1	What are the different types of machines used in the manufacturing production line?
2	What are the tools hosted on a particular machine?
3	What is the status of a motor at a certain time?
4	what is the temperature on different machines?
5	How many processes are performed by machines and count the tools used by them?
6	Which processes are performed by assembling machines?
7	what are the operating hours for a specific machine?
8	Which staff members are involved in specific manufacturing processes?
9	How much power is consumed in different manufacturing processes?
10	What products are produced by different machines?

3.2.2 Formalizing Concepts

The Formalizing Concepts step contains two sub-activities such as ontology conceptualization and ontology encoding. In order to conceptualise the knowledge of RGOM, the terminology is adopted from RAMI4.0 as it provides the basic representation of the manufacturing domain otherwise the terminology is reused from the existing ontologies that are relevant to the manufacturing domain.

For example, concepts named `Machine` and `Process` have been created to model a machine and the process it performs, with a property such as *performProcess* that defines the relationship between the Machine and the process. Upon formalizing this, it is implemented using the open-source ontology editor of Protege[1]. The RGOM is encoded into Resource Description Framework/Web Ontology Language (RDF/OWL).

3.2.3 Ontology Validation

This section discuss the validation steps of the ontology development process that is utilized to assess the RGOM through Competency Questions (CQs), FAIR assessment and OOPs assessment tool. CQs are important for the creation and verification of an ontology [14, 174]. These are questions that an ontology needs to be able to answer, and they serve as requirements or use cases for the ontology. For the development of the RGOM, the CQs in Table 3.2 are defined. Some of the CQs are demonstrated in Chapter 4 Section 5.3.4. FAIR (Findability, Accessibility, Interoperability, and Reusability) is assessed using O'FAIRe tool which stands for Ontology FAIRness Evaluator [6]. It is a tool that enables automatic assessment of the FAIRness of ontologies. The OOPS tools evaluate the structural, and functional dimensions of the ontology to analyze its clarity, conciseness and consistency. It has been widely used among researchers to identify flaws and pitfalls in ontology design [121]. The missing domain and range in the properties, creating unconnected elements in the ontologies are some of the pitfall example cases that are checked by the OOPS.

1 http://protege.stanford.edu/

3.2.4 Ontology Publication

The ontology is published on the industry portal[2] and contains metadata and information such as Uniform Resource Identifier (URI), license, and title. It also contains information such as creator, contributor, endorser, date of creation and others.

3.3 ENCODING RAMI4.0 CONCEPTS TO RGOM

Within the RAMI 4.0 framework, the hierarchy dimension plays a critical role in shaping the organisational structure of a company, specifically concerning how the various levels of the manufacturing system are meticulously organised and structured. It represents the highest level in the hierarchy and encompasses the entire physical manufacturing facility or site. It includes all the resources, equipment, and infrastructure located within the factory premises. This level focuses on the overall management and coordination of the manufacturing operations that take place within the facility. The focus of RGOM is to consolidate the generated data in this axis that can be reused with minimal effort into a single unified place. Figure 3.4 shows an illustration of encoding RAMI4.0 to RGOM.

Enterprise: The term **Enterprise** in RAMI4.0 refers to the highest level of organizational structure within the manufacturing system. It represents an entity that encompasses all other levels and entities involved in the production process. The Enterprise level typically includes multiple production plants or facilities. RGOM uses the same term `Enterprise` to represent it.

Work Center: In RAMI4.0 the **Work Center** represents a specific unit or area within the manufacturing system where manufacturing activities take place. It is a functional entity that consists of multiple cells or workstations and is responsible for executing specific tasks. RGOM represents the term Work Centers into `ProductionLine` which then can contain many `Cells`.

2 https://industryportal.enit.fr/ontologies/RGOM

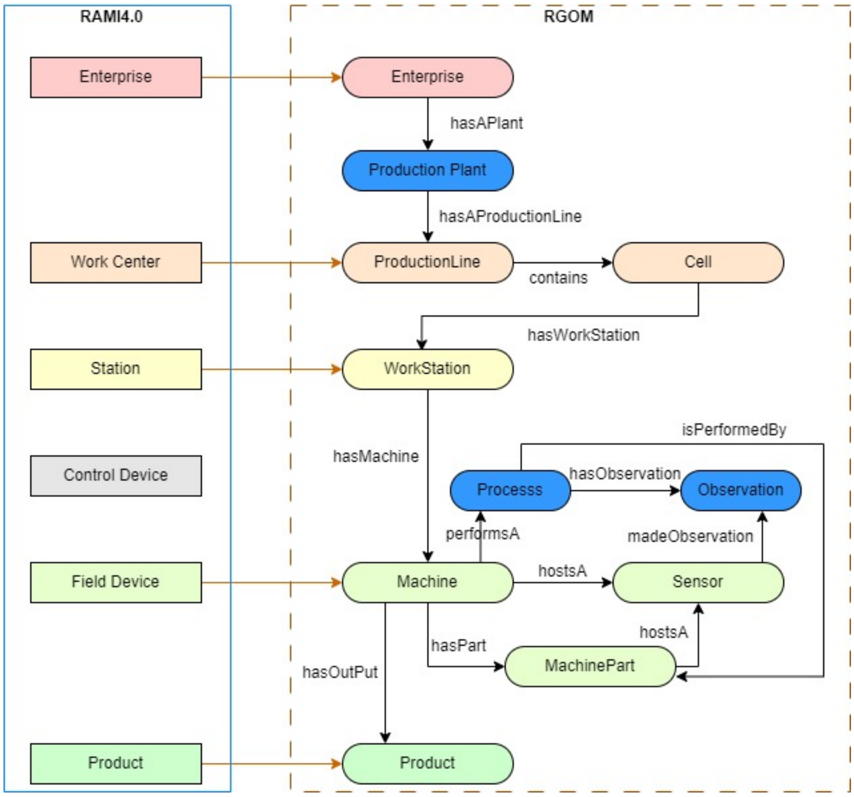


Figure 3.4: Encoding of RAMI4.0 concepts to RGOM

Station: In RAMI4.0 the **Station** represents a specific location or unit within a Work Center. It is a physical point where tasks or operations are performed. Stations are typically equipped with specific tools, equipment, or machinery required for carrying out the assigned manufacturing activities. RGOM adopted the term `WorkStation` to represent the Station.

Field Device: The term **Field Device** in RAMI4.0 refers to physical devices or components that interact with the manufacturing system at the operational level. This includes various devices such as machines, sensors, and other hardware elements. RGOM uses `Machine`, `MachinePart`, and `Sensor` to represent the field devices.

Product: In RAMI4.0, **Product** refers to the end result or output of the manufacturing process. It represents the tangible or intangible item that is being produced or manufactured. RGOM use the term `Product` to represent it.

RGOM provides a structured framework for organizing and understanding the various entities and components involved in the manufacturing sys-

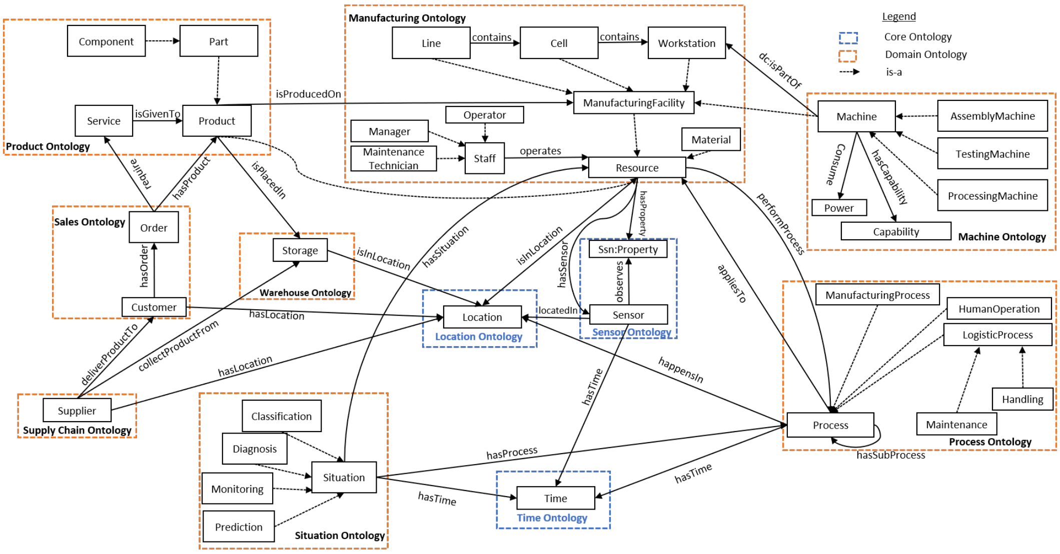


Figure 3.5: An overview of Reference Generalized Ontological Model

tem like the RAMI4.0. This allows for a clear delineation of responsibilities, control, and coordination of activities within the production environment.

3.4 DEVELOPED ONTOLOGY: RGOM

We have proposed a Reference Generalized Ontology Model (RGOM) that is developed by reusing the concepts of existing ontologies and defining new concepts that have been overlooked based on the reference architectural model Industry 4.0 (RAMI4.0). The proposed RGOM considers core areas such as time, location, sensor, and different domain attributes such as product, process, and machine along with the order, supply chain, warehouse, etc., and explores all the concepts along with the relationships among them.

This implies that the RGOM provides a detailed unified model that takes the I4.0 domain knowledge from raw material to finished product including supply to the customer as well as monitoring the different situations of machines and their processes. Machines and products are separated from the resource ontology to form a machine and product ontology for accommodating more concepts and relationships. For instance, the product ontology specifies the concepts such as product (production of product) and service (maintenance usage) adopted from RAMI4.0 and the identified concepts such as sales ontology are coupled. This helps to provide a full view that the order is placed for a service or the manufacturing of the product, depending on the order either the service or the resources in the manufacturing production line will be reconfigured. RGOM has reused the existing vocabulary, that is, the manufacturing facility machine associated with the workstation by reusing the *isPartOf* property from the Dublin Core vocabulary. The process(es) happening at different times and locations are linked to the manufacturing resources by process ontology using the *performA* property. It describes the basic taxonomy of all kinds of processes taking from manufacturing to human process(es) and logistic operations. Sales ontology defines customer order concepts for the product. The order can have various concepts such as design, quantity, delivery date, etc. The supply chain ontology can assist in monitoring the delivery of the manufactured products to the customer. Thus, the context of the core ontologies alone would not be able to answer why, where and what type of questions, but the RGOM can infer all the contextual information ranging from a particular entity situation to the complete production line. Figure 3.5 depicts the concepts and relations reused from relevant existing ontologies (ssn:*Sensor*, ssn:*Property*, sosa:*madeObservation*, sosa:*observes*, time:*Time*, time:*hadTime*, dc:*isPartOf*), while there are several newly defined concepts such as *Power*, *Tool*, *Part*, *Service*, etc. Furthermore, Table 3.3 provides an overview of RGOM with emphasis on key RGOM concepts and relationships as well as other reused ontologies and their references.

Table 3.3: An overview of RGOM highlighting key concepts and relations, along with references to other reused ontologies.

RGOM components	Reused concepts (References)	RGOM new concepts/relations
Manufacturing ontology	[30],[71], [144], [53]	Material, Container (Pellet), isProcessedby, isPlaceOn
Machine ontology	[30], [126], [144], [53]	MachinePart, Tool, Capabilities, Current, Power, consumesPower, hasInputMaterial, hasTool, useTool
Process ontology	[30],[53]	ManualProcess, MeasurementProcess, ConveyorOperation, FeederOperation
Product ontology	[53]	Service, isGivenTo
CoreOntology	[53], [112],[131], [34]	X

Moreover, the main modules related to the RGOM ontology are discussed as follows.

3.4.1 Manufacturing Ontology

The objective of manufacturing ontology is to semantically describe the resources in the manufacturing production line. The *Staff* concept represents all the people participating in the production activities, that is, technicians, operators, engineers, supervisors and managers. The *ManufacturingFacility* concepts characterize different physical entities and hardware modules in the factory. The concepts of the production line are decomposed into *Cell* is the combination of the workstation to perform a complex task; a *line* includes cells, and *Workstations* contains the physically integrated machines. This decomposition presents a potential reconfigurable processing line. Additionally, this taxonomy makes it possible to describe the manufacturing facility context at various next levels such as the characterization of a line or to illustrate the context of the cell that belongs to that line. The physical entities in the manufacturing facility are the resources linked to other ontologies via related object properties.

3.4.2 Machine Ontology

The machine is the main resource to process raw or refined material into semi or finished products on the production line. The machine performs the process with the help of tools by itself or with intervention from the human. It can be either a processing or assembly machine processing a raw or refined material or assembling the refined parts. The machine is a manufacturing facility that is part of the workstation.

3.4.3 Process Ontology

A set of tasks or operations completed by a resource is known as the *Process*. The process(es) performed by a resource(s) can be known as controlled operations as well as machining or assembly ones. The process ontology represents the fundamental taxonomy of all the processes executed in the manufacturing and is specified with contextual such as process *happensIn* location, process *appliesTo* a product, machine *performA* process, etc.

3.4.4 Product Ontology

Product ontology covers the basic taxonomy related to products based on RAMI4.0. The components are the parts assembled by an assembling machine into a finished product. The customers can place an order of one of the two, that is, for service to the bought product or buying a new product.

3.4.5 Core Ontologies

Creating the context for scientific manufacturing tasks is difficult and poses a major problem in the industrial domain because it includes many varying entities associated with time and locations. *Time*, *Location*, *Process*, *Machine* and *Resource* are the primary concepts to semantically represent the manufacturing knowledge in line with domain ontologies. In addition, the terminology reused from the sensor ontology, that is, the SSN enhances the semantic representation of the collected sensor data. The concepts for measuring the sensor data are reused from the Ontology of units of Measure (OM) [131]. Thus, the use of basic ontologies such as *Time* [112], *Location* (adopted from [53]) and *Sensor* [34] with the domain ontologies present useful information regarding various situations, that is, inquiring about the status of the motor at a particular time.

The RGOM contains a total of 84 classes, 74 object properties and 48 data properties. In the context of RGOM, the extent of class reuse is as follows: OM2 (8.3%), Saref (4.76%), SSN/SOSA (2.3%), MSDL (7.1%), and Extruont (1.19%). The namespaces used in the RGOM are given in Table 3.4.

Table 3.4: Namespace Prefixes and IRIs

Prefix	IRI
rgom	http://www.semanticweb.org/manufacturingproductionline#
extruont	http://bdi.si.ehu.es/bdi/ontologies/ExtruOnt/components4ExtruOnt#
msdl	http://infoneer.txstate.edu/ontology/
om2	http://www.ontology-of-units-of-measure.org/resource/om-2/
saref	https://saref.etsi.org/core/
sosa	http://www.w3.org/ns/sosa#
ssn	http://purl.oclc.org/NET/ssnx/ssn#
tm	http://www.w3.org/2006/time#
dc	http://purl.org/dc/elements/1.1#
owl	http://www.w3.org/2002/07/owl#
rdf	http://www.w3.org/1999/02/22-rdf-syntax-ns#
rdfs	http://www.w3.org/2000/01/rdf-schema#

3.5 ONTOLOGY EVALUATION

3.5.1 FAIR Tool Assessment

The FAIRness of RGOM is performed with O'FAIRe tool. It is integrated within the industryportal ontology repository and assesses the FAIRness of semantic resources or ontologies based on the FAIR Principles. The tool operates through a set of 61 questions/tests [6]. O'FAIRe provides both comprehensive and detailed scores, normalized against the 15 FAIR Principles, for individual ontologies or groups of semantic resources. These results help in understanding how well a semantic resource or ontology adheres to the FAIR Principles. Figure 3.6 shows an overview of the results returned for an individual evaluation of the RGOM in industryportal.

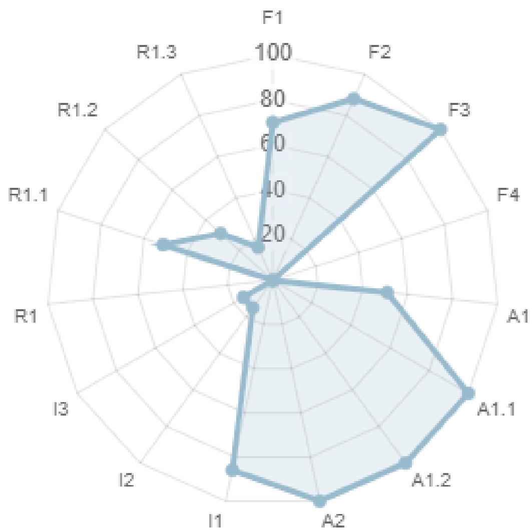


Figure 3.6: Overview of O'FAIRe evaluation of RGOM in industryportal.

3.5.2 OOPs Tool Assessment

The Ontology Pitfall Scanner (OOPS) evaluates ontologies for common flaws during their development. It identifies 41 potential pitfalls classified as minor, important, or critical. In the assessment of the RGOM ontology using OOPS, minor issues were found, such as unconnected elements, duplicated class labels, and missing domains or ranges. Figure 3.7 shows the OOPS pitfall detection result. OOPS also checks for clarity, conciseness, and consistency.

- **Clarity:** The ontological terms defined to represent the classes, concepts, and relations of all the modules, contain unambiguous names, and annotations. The annotations aid in the readability of humans to avoid uncertainty and difficulty during the insertion of data elements.
- **Conciseness:** The industry knowledge represented by the ontology that is gathered in line with the sources, particularly workstations,

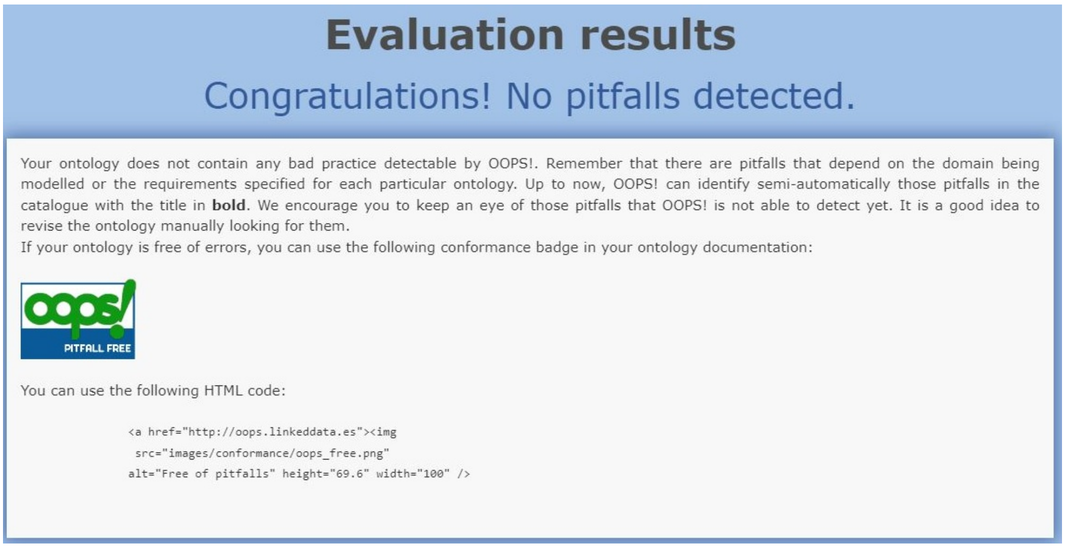


Figure 3.7: OOPs assessment shows RSWO does not contain any bad practice detectable by OOPS!.

machines, tools, materials and processes, and enterprise and their production lines.

- **Consistency:** The Hermit[3] reasoner has been applied to find inconsistencies in the RGOM. Accordingly, the reasoner has not found any inconsistencies in the developed ontology.

3.6 SUMMARY

In this chapter, the steps involved in constructing the RGOM are presented. Moving on forward, the encoding of concepts from the Reference Architecture Model for Industry 4.0 (RAMI4.0) into RGOM has been focused. Finally, we have presented the developed ontology RGOM. This ontology serves as a comprehensive knowledge representation framework, capturing the domain-specific concepts, relationships, and semantics required for effective ontology development and its subsequent integration within the Industry 4.0 context. In the next chapter, we will explain the ontology development process for RSWO in line with RGOM using a Bosch resistance spot welding use case.

3 http://www.hermit-reasoner.com/

4 RESISTANCE SPOT WELDING ONTOLOGY IN LINE WITH RGOM

This chapter describes the alignment of RGOM as a domain-level ontology to provide interoperability for the Resistance Spot Welding Ontology (RSWO). Followed by the ontology development process, the main steps are then discussed, which explain the ontology requirements gathered from the domain experts and ISO welding documents, concepts formalization, ontology validation and its publication and maintenance. Then, the domain knowledge encoded into the RSWO is presented. Finally, the ontology is evaluated from four dimensions, (i) Industrial Use-Case: Quality Monitoring in Resistance Spot Welding, (ii) Evaluation on Findable, Accessible, Interoperable, Reusable (FAIR), (iii) OOPS! too and (iv) OntoMetrics.

4.1 ALIGNMENT WITH THE DOMAIN LEVEL ONTOLOGY (RGOM)

The generation of data in various formats usually results in having associated with it the interoperability issue that hinders inter-communication. To enable interoperability between Resistance Spot Welding (RSW) resources the RSW ontology alignment with domain-level ontology is considered. The domain-level ontology is very important as it provides semantic interoperability across the domain. There exist several domain-level ontologies such as MASON [89], CDM-Core [103], RGOM [175] etc. RGOM is selected as a domain-level ontology as it is built on re-using the manufacturing ontologies with the terms being introduced that are considered to be overlooked in the previously existing vocabularies. In order to illustrate the alignment of RSWO with the domain level ontology RGOM, consider Figure 4.1, which shows the alignment of RSWO shown in orange color) to the RGOM ontology (shown in green color). For example, the class `RSWMachine` of RSWO is created as a subClass of `Machine` and the `Electrode` class in the RSWO is created as a subClass of `MachinePart` of RGOM. The `RSWMachine` class is linked to the `Electrode` class through the property *hasPartElectrode* which is the subProperty of *hasPart* of the RGOM.

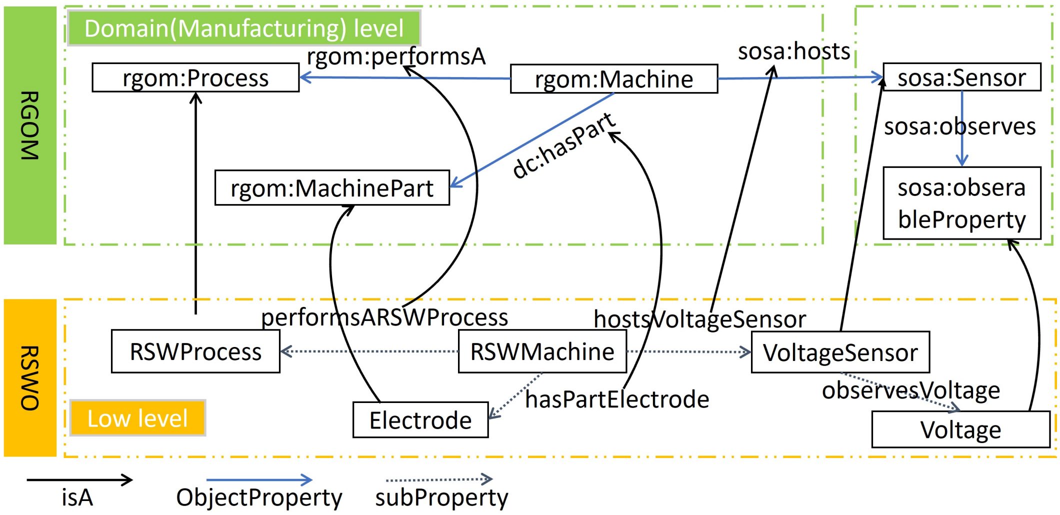


Figure 4.1: Alignment of the RSWO ontology with the domain level ontology.

4.2 ONTOLOGY DEVELOPMENT PROCESS

Ontology development provides a step by step guidelines for designing and developing ontology engineering by including the construction of classes, and their relationships. There exists a number of methodologies for ontology development in the literature. Some of the most popular ontology development methodologies are METHONTOLOGY [45], Common-KADS [141], and Linked Open Terms (LOT) [120]. In this work, the LOT methodology [120] is adopted which is a refined work based on the top for developing RSWO. The reason for selecting this methodology is that it provides a gradual refinement process throughout the ontology creation. This refinement ensures that the ontology captures the domain knowledge concepts as well as the low-level manufacturing data. The ontology development process is as shown in Figure 4.2.

4.2.1 Ontology Requirement Specification

With the help of experts in the Bosch manufacturing company, it is identified why there is a need for an RSW ontology. This is specified with the use case (Quality Monitoring in Resistance Spot Welding). In relation to this, several documents have been provided including the ISO welding standards, datasets description, and datasets itself. The aforementioned activity has thus helped in the identification of the purpose and scope of the RSW ontology (unified model for answering the questions related to RSW). Considering the scope of the RSW ontology, the functional requirements are collected in the natural language sentences (such as, *Resistance spot welding operation consume power*) from the Bosch welding experts, as they have zero knowledge about ontologies. The natural language sentences are then transformed into related competency questions. Some of the Competency questions (CQs) are listed in Table 4.1. These CQs are provided by welding experts. They are grouped into two categories:

(1) Data Inspection: (CQ1-CQ5) We have used RSWO-based Bosch welding process data to examine it from a variety of angles. We have inspected the data during and after the welding operations for the objectives of verification and quantification of welding quality.

(2) Diagnostics: (CQ6-CQ10) We have performed different diagnostic tasks such as dressing required, spot repetition occurred, the occurrence of any spatters, and many others. Besides, the diagnostics

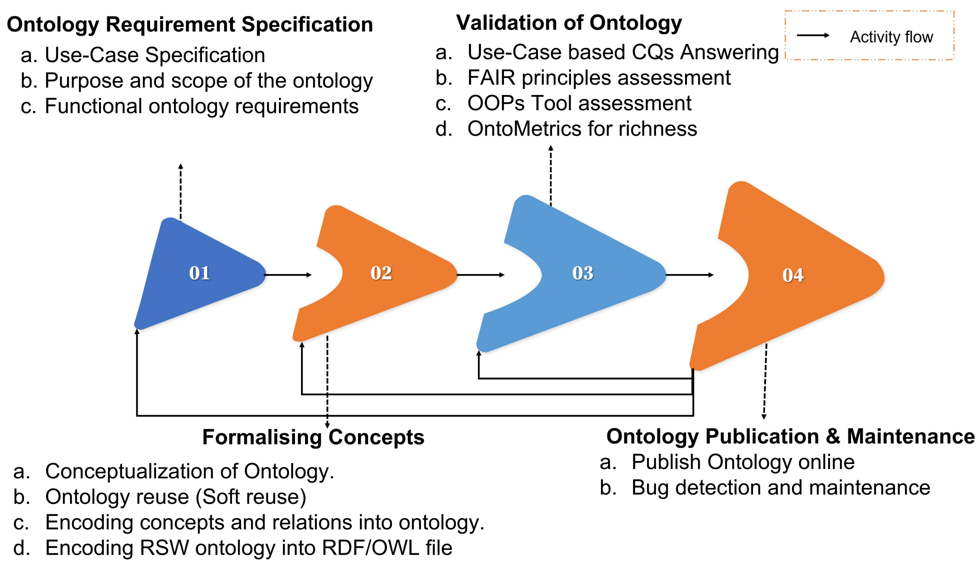


Figure 4.2: Steps used in the ontology development process

enable the user to learn more about the surrounding irregularities to comprehend what occurred nearby and identify potential root causes.

Finally, when these functional requirements are successfully approved we have moved on to the second activity of the ontology development.

Table 4.1: Competency questions provided by Bosch experts

Data Inspection	CQ1	How much weld force, voltage, current and power is utilised in an operation?
	CQ2	What machine parts are being used in a resistance spot welding operation?
	CQ3	How much force is utilised in the squeeze, weld, and hold step of the operation?
	CQ4	How much is the resistance between the bottom Electrode and bottom sheets?
	CQ5	How many cycles of weld time is utilised in an operation?
Diagnostics	CQ6	Find all those values of Q-Value higher than a threshold along with their voltage and power in an operation.
	CQ7	Is there any spatter that occurred during a particular time?
	CQ8	Does the electrode require dressing?
	CQ9	How many weld spots have spot repetition?
	CQ10	How much force is utilised in the squeeze, weld and hold steps of the operation?

4.2.2 Formalizing Concepts

The Formalizing Concepts activity contains the first three sub-activities such as ontology conceptualization, ontology reuse and ontology encoding. In order to conceptualise the knowledge of RSW, the terminology is adopted from the RGOM as it provides the basic representation of the manufacturing domain otherwise the terminology is introduced from the ISO welding documents that are relevant to the RSW domain.

For example, a concept named `WeldingMachine` has been created to model a welding machine with a property such as *dc:hasPart* that defines the relationship between WeldingMachine and the MachinePart. The `WeldingMachine` is further linked via the *performsA* property with the `AssemblyProcess` concept, and it is then linked to the *hasOperations* relationship with the concept *RSWOperation*. RSWOperation is shown to be linked with the concept Assembly through *hasRawProduct* property. The *isOperationProductOf* property is used to connect the concepts`WeldSpot` and `RSWOperation`.

Upon formalizing the concepts, it is implemented using the open-source ontology editor of Protege[1]. The RSWO is encoded into Resource Description Framework/Web Ontology Language (RDF/OWL).

1 http://protege.stanford.edu/

4.2.3 Ontology Validation

This section discusses the validation steps of the ontology development process that is utilised to assess the RSWO through several metrics: Usecase-based Competency Questions (CQs) answering, Findable, Accessible, Interoperable, Reusable (FAIR) principles, Ontology Pitfall Scanner! (OOPS) and OntoMetrics. The Use-case-based CQs answering is performed on the Bosch Production data to demonstrate the functionality and utilization of the RSWO. The CQs have been provided by the Bosch experts. The CQs have determined whether the RSWO ontology has captured the domain knowledge, for example, the diameter of the weld spot produced in a certain welding operation, the weld force applied to workpieces, and the resistance between the electrode and the workpiece.

Moreover, the O'FAIRe methodology assesses the FAIR principles, and the OOPS tools evaluate the structural, and functional dimensions of the ontology to analyze its clarity, completeness, conciseness and consistency. It has been widely used among researchers to identify flaws and pitfalls in ontology design [121]. The missing domain and range in the properties, creating unconnected elements in the ontologies are some of the pitfall example cases that are checked by the OOPS. Also, the ontology populated with data instances is uploaded to OntoMetrics[2] for advanced analytics.

4.2.4 Ontology Publication and Maintenance

The RSWO is available online[3] and is accessible. The metadata is published on the industry portal and contains information such as Uniform Resource Identifier (URI), license, and title. It also contains information such as creator, contributor, endorser, date of creation and others. Additionally, to maintain the ontology, the bugs can be reported on the GitHub page[4] that can be tracked.

2 https://ontometrics.informatik.uni-rostock.de/ontologymetrics/
3 https://w3id.org/def/mo-rswo
4 https://github.com/nsai-uio/RSWO

4.3 RESISTANCE SPOT WELDING ONTOLOGY

The RSWO ontology description has been provided here that has been implemented in OWL in the light of the aforementioned methodology. The primary purpose of using OWL is to provide a widely accepted information-sharing environment in order to improve RSW processes. Furthermore, we have exemplified our ontology with simplified Description Logic (DL) syntax in the following subsections. The developed RSWO ontology is comprised of the metrics listed in Table 4.2. Moreover, there are a total number of 112 classes, 98 object properties and 71 data properties in the RSWO.

Table 4.2: Ontology metrics

Metric	RSWO
Axioms	1164
Logical axioms count	469
Declaration axioms count	287
Class count	112
Object property count	98
Data property count	71
Individual count	0

4.3.1 Overview of RSW Process and the Ontology

Resistance Spot Welding (RSW) is frequently used in the automotive sector, for example, in the manufacture of vehicle bodies. This process is controlled by welding control systems that store weld configurations. In the resistance spot welding process, the welding gun is equipped with electrodes that end with caps to press two or three worksheets. Then, an electric current flows from one electrode, through the worksheets, to the other electrode, generating a large amount of heat due to electrical resistance. The material in a small area between the two worksheets, called the welding spot, thus melts, forming a solder mass that connects the worksheets.

In connection with the above description, the ontology of resistance spot welding has been developed that provides the automotive welding industries with a common knowledge architecture. Figure 4.3 depicts the main classes and characteristics of the RSW ontology. The ontology has reused the domain core concepts such as `Machine` and `Operation` to facilitate inter-domain data integration. The subclass of the `Machine` is the

`WeldingMachine` that contains a number of machine parts. The figure represents the parts of the machine with the class `MachinePart`, which is linked to the class `Machine` via the object property *dc:hasPart*. On the other hand, RSWOperation is a subclass of the domain core concept Operation. Axioms 1 and 2 define the subclass constraints in the RSWO and axiom 3 shows the constraints that every WeldingMachine performsA some RSWOperation in Description logics (DL).

axiom 1: $WeldingMachine \sqsubseteq RGOM : Machine$

axiom 2: $RSWOperation \sqsubseteq RGOM : Operation$

axiom 3: $RSWOperation \sqsubseteq \exists performsA^{-}.WeldingMachine$

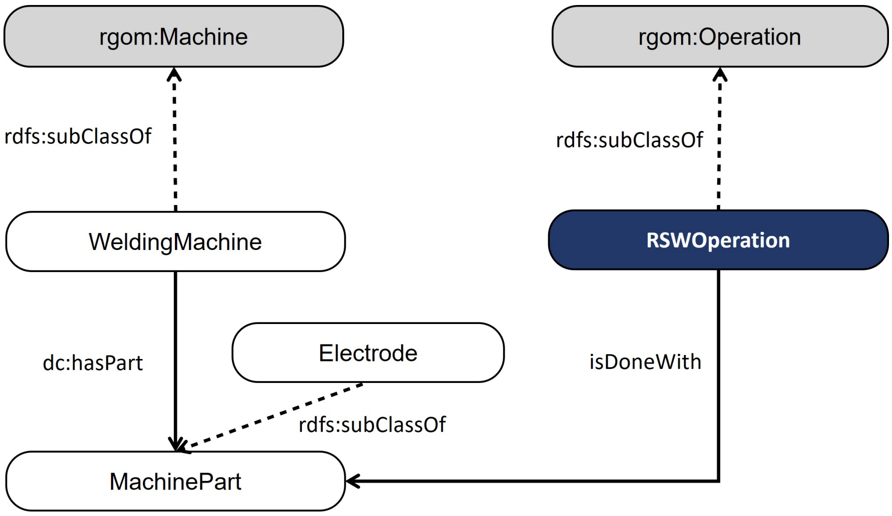


Figure 4.3: Overview of core concepts in the RSW ontology.

In the following subsections, we will go to the detailed modelling of these concepts.

4.3.2 Resistance Spot Welding Operation

Resistance Spot Welding (RSW) is a complex task that is widely used in diverse applications such as vehicle body parts, railway tracks, turbine blades, etc. [184]. It contains a number of activities that are performed to produce welding processes. An operation is an atomic process that takes in the worksheets as raw products in order to produce a product output, namely a welding spot.

RSW Operation modelling is as shown in Figure 4.4 that shows that class `RSWOperation` is linked to the `WorkPieceCombination` class through *hasRawProduct* object property. In the RSW operation, an assembly is the raw product which is the subclass of the `WorkPieceCombination`. The class `Assembly` has parts such as `TopWorkSheet` and `BottomWorkSheet` that are the subclasses of the `WorkPiece`. The property *dc:isPartOf* connects *WorkPiece* with *Assembly*.

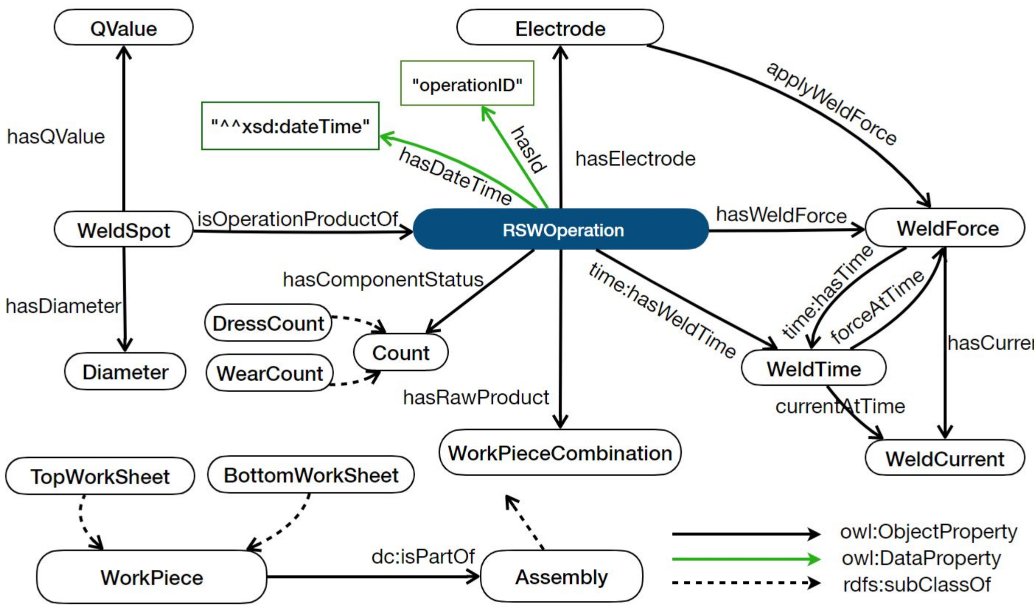


Figure 4.4: RSW operation modelling. The rounded rectangle represents the classes and the square rectangle represents the literals.

Furthermore, RSW operation has gotten an electrode that applies pressure to the aligned workpiece point of interest. It can be observed from the top of the figure that `RSWOperation` and `Electrode` classes are related via *hasElectrode* property. In addition to this, the `WeldForce` class is linked through the *applyWeldForce* relation to `Electrode`. After applying the pressure, a constant current is applied through the electrodes into the *WorkPieces*. Based on this description, the right side of the figure illustrates the welding conditions maintained during the RSW operation are that the `WeldCurrent` and `WeldTime` has a constant `WeldForce` or the `WeldCurrent` and `WeldForce` in the welding has a constant `WeldTime`. This brings out an internal resistance in the worksheets and in results produces a weld spot as its product. The `WeldSpot` and `RSWOperation` classes are linked through *isOperationProduct* relation (left side of figure). The RSW operation has the

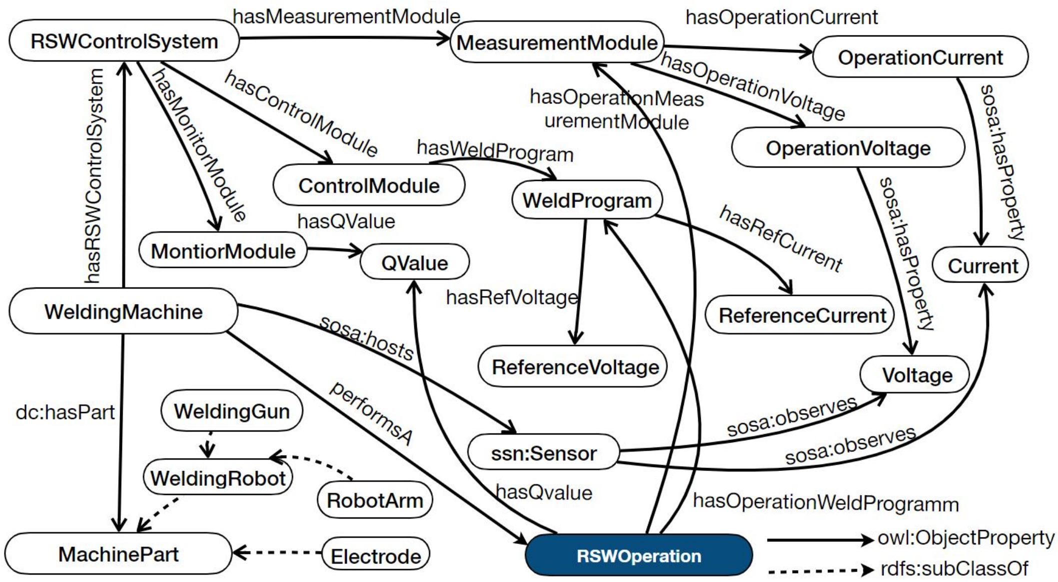


Figure 4.5: An overview of Machine and Software modelling

date time on which the operation is executed, so the *hasTime* property from the time vocabulary is reused to link the operation class to its date time data instance.

We exemplify the operation part of the ontology with an example about the weldspot as shown on the left side of 4.4 using DL. The weldspot diameter is also an operation quality indicator. Axiom 4 represents the axiom that for every `RSWOperation` there exists some `WeldSpot` and `WeldSpot` is the (operation) product of `RSWOperation`. Axiom 5 specifies that a `WeldSpot` has some value `Q-Value`[5]. Axiom 6 also specifies quality indicator (`Diameter`) that a `WeldSpot` has exactly one diameter.

axiom 4: $RSWOperation \sqsubseteq \exists isOperationProductOf^{-}.WeldSpot$

axiom 5: $WeldSpot \sqsubseteq \exists hasQValue.QValue$

axiom 6: $WeldSpot \sqsubseteq = 1 hasDiameter.Diameter$

5 The Q-Value is a quality indicator that is used to quantify the welding quality. The Q-Value is empirically developed by Bosch Rexroth in the Bosch labs with longtime experience and engineering expertise. A Q-Value of 1 indicates perfect quality.

4.3.3 Welding Machine and Software

The welding machine and its software are now shortly introduced. A welding machine performs an RSWOperation. The welding machine consists of several parts such as a welding robot, welding gun, electrode, sensors etc. that carry the commands of the software systems to carry out the required operation. The welding machine is controlled by a software system to perform the desired designed welding operation.

The software system known as the RSW control system has three modules, each of which has a specific task, that is, the monitoring module monitors the quality of the weldspot and operation, the control module provides the setpoints and reference programs for operation, and the measurement module collects voltage, energy, resistance, etc. and other observations.

An excerpt of the semantic representation of the welding machine and software system is as shown in Figure 4.5. The `WeldingRobot` and `Electrode` classes are the subclasses of `MachinePart` which is linked to the object property *dc:hasPart* to the `WeldingMachine` class, (axiom 7). Moreover, the `WeldingRobot` and `Electrode` are the disjoint classes. Furthermore, The `MachinePart` is connected to `RSWOperation` through *performsA* relation. The `WeldingMachine` and `RSWControlSystem` classes are linked via *hasRSWControlSystem* property.

axiom 7: $WeldingMachine \sqsubseteq \exists hasPart.MachinePart$

$$MachinePart \sqsubseteq WeldingRobot \sqcup Electrode$$

$$WeldingRobot \sqsubseteq \neg Electrode$$

The welding machine hosts sensors to collect the power, energy, and voltage reading observations for being recorded in the measurement module. The `WeldingMachine` and `Sensor` classes are linked by the *hosts* relationship. In an operation of RSW, a summary of useful information can be retrieved by using *hasOperationWeldingProgram*, *hasOperationMeasurementmodule*, and *hasQValue* relationships. The axioms 8a-8d represent that there exists a `MeasurementModule` on the welding control systems to collect operation current property readings.

axiom 8a: $WeldingMachine \sqsubseteq \exists hasRSWControlSystem.RSWControlSystem$

axiom 8b: $RSWControlSystem \sqsubseteq \exists hasMeasurementModule.MeasurementModule$

axiom 8c: $\mathit{MeasurementModule} \sqsubseteq \exists \mathit{hasOperationCurrent.OperationCurrent}$
axiom 8d: $\mathit{OperationCurrent} \sqsubseteq \exists \mathit{hasProperty.Current}$

4.3.4 Electrode

The electrode is an important component of the welding machine used in the `RSWOperation` because its condition as characterised by `WearCount` and `DressCount`) has a significant influence on the welding quality. The object property `hasElectrode` relates the `RSWOperation` with the electrode (Figure 4.6 and axiom 9(a,b)). The electrode has two subclasses, namely: `TopElectrode` and `BottomElectrode` that apply force to the workpiece and then pass current to produce resistance creating thus a welding spot. The workpiece has two subclasses of `TopWorkSheet` and `BottomWorkSheet` that interact with TopElectrode and BottomElectrode, respectively. The class `PiecePieceInteraction` is used to model interaction properties between the worksheets. For instance, between the two worksheets of RSW, there exist interaction properties such as adhesive, thermal conductivity and electrical conductivity. In this regard, such modelling provides useful information in the operation by representing the `PiecePiece Interaction` and `RSWOperation` through *hasInteraction* relation.

axiom 9a: $\mathit{Electrode} \sqsubseteq \mathit{MachinePart}$
axiom 9b: $\mathit{Electrode} \sqsubseteq \exists \mathit{hasElectrode}^{-}.\mathit{RSWOperation}$

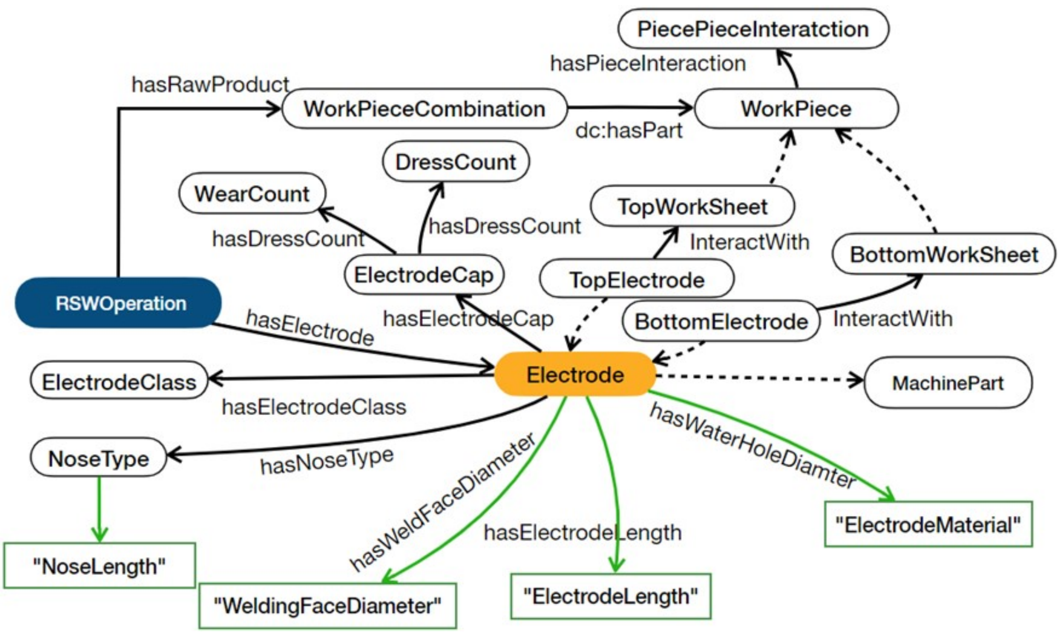


Figure 4.6: An illustration of electrode modelling

In the `RSWOperation`, the spot welding electrode cap wears and appears as a mushroom, which passes insufficient current and results in inconsistent welds. The electrode dressing procedure is used to restore the original shape of the electrode cap. In order to capture this information, the welding caps have system component status of `WearCount` and `DressCount` that are named as operation count and maintenance count. The electrode comes with a variety of nose configurations that are considered during the design phase of the welding. The electrode face is exposed to extremely high temperatures for a short period of time during the procedure. The electrode temperature is cooled down with water to prevent premature corrosion. The electrode has a water hole of a particular diameter that allows water to flow through it.

4.4 ONTOLOGY EVALUATION

The RSWO is evaluated in four dimensions: (1) the use-case use of the Bosch resistance welding process to monitor quality, (2) analyzing for FAIR principles, (3) structural and functional aspects of RSWO with OOPS, and (4) analyzing the attributes richness with OntoMetrics.

4.4.1 Industrial Use-Case: Quality Monitoring in Resistance Spot Welding

This subsection demonstrates the use-case of a quality monitoring task performed in the resistance spot welding process at Bosch in Germany. The purpose of the industrial use case is to assess the utility and function of RSWO in a truly intelligent manufacturing environment. The remainder of this section gives a comprehensive explanation of the Bosch welding experts, Bosch welding process, welding data and quality monitoring using RSWO.

Evaluation by Bosch Welding and Ontology Experts

This subsection discusses the evaluation of RSWO by domain experts and ontology Experts (OE). The ontology experts have specialised skills in creating and refining ontologies. They have extensive knowledge in designing semantic models ensuring knowledge representation.An industrial use-case-based workshop is carried out with the RSW domain experts. The domain experts lack the knowledge about ontology generation and query and as per recommendation, they provided the queries in natural language sentences (Section 4.2.1. The natural language sentences are transformed into competency questions (CQs) that are asked via

SPARQL query (Section 4.4.1). The CQs in terms of data inspection and diagnostics demonstrated that the ontology developed can be used for the defined use case.

We evaluated our ontology from ontology experts based on the criteria defined by [151]. We provided the ontology file, and documents defining the scope and functional requirements to the ontology experts. The experts gained RSW knowledge from the provided documents. They responded to a series of questions (shown in Table 4.3). The questions given in the table define the quality criteria related to clarity (1-4), accuracy (5-6), consistency (7), and completeness (8-10) to assess ontology context coverage, level of detail, relevance and semantic richness.

We received a score of 4 (Agree) from almost all the OEs. However, the OEs have different views on the clarity of the ontology (questions 1 and 3) and provided a score of 3 (Neutral) which means that there is still some space for improvement. Furthermore, OE 3 think there is some inconsistency in the ontology and gave some specific comments about ontology classes and proprieties (such as has hasControlModule, hasMeasurementModule and hasMonitorModule can be included as part of the RSW control system) and provided a score of 3. OE 2 strongly agreed with the completeness of the ontology and gave a score of 5.

Table 4.3: Ontology evaluation by ontology experts. OE indicate the ontology experts. A scale of 1-5 is used to evaluate the ontology criteria where 1 shows strongly disagree, 2 disagree, 3 neutral, 4 agree and 5 strongly agree.

Evaluation criteria	OE 1	OE 2	OE 3
1. Are the annotations of classes sufficient?	3	4	3
2. Are the annotations of classes unambiguous?	4	4	4
3. Are the annotations of properties sufficient?	3	4	4
4. Are the annotations of properties unambiguous?	4	4	4
5. Are [owl:Class]s and [owl:Properties]s well-structured for an RSW domain ontology and do they properly represent the entities?	4	4	4
6. Do the axioms adhere to annotations for the RSW domain?	4	4	4
7. Do the axioms employed convey the concepts intended meaning?	4	4	3
8. Does the ontology cover the necessary concepts required by RSW domain ontology?	4	4	4
9. Does the ontology align to domain-level ontology?	4	5	4
10. Does the ontology reuse term from other ontologies?	4	5	4

Bosch RSW Process Quality Monitoring

Bosch is one of the top world's top manufacturers in the automotive industry. Bosch uses the RSW process to join body parts to manufacture an automotive. In the RSW operation, the surfaces of metal sheets are bonded by the heat gained from the resistance generated by electric current. The body of a typical car can have up to 6000 welding points [182], where metal pieces are connected. Bosch offers a variety of welding solutions, including as software, service, development support, and welding equipment. Other than the Bosch welding plant, these solu-

Table 4.4: Examples of the dataset attributes, their datatype and short description

Data Attribute	Data Type	Description
WeldSpot_ID	Integer	Unique identifier of the weldspot
WeldSpot_Diameter	Float	It provide the diameter of the weldspot
WeldSpot_Repetition	Boolean	Indicates whether a weldspot is repeated or not
RSWOperation_ID	Integer	Unique identifier of the resistance spot welding operation
TimeStamp	Date	Timestamp of the date when the operation was performed
Machine_Name	String	Machine that performs the operation Machine_ID
Electrode1_WearCount	Integer	Wear count faced by an electrode during welding
Electrode1_DressCount	Integer	Count of dressing applied to an electrode
Electrode1_WeldForce	Float	Weld force applied by an electrode
Electrode1_SetPoint_Resistance	Float	Reference value of resistance set for Electrode 1

tions are also adopted by customers all over the world-wide such as BMW, Audi, Ford, and Daimler.

Bosch resistance welding machine and its parts are a tried-and-true way to quickly join hundreds of pieces every hour. To determine and ensure the weld quality (essential to many facets of the vehicle's performance and value) a Q-Value (described in Section 4.3.3) test of the selected components is performed in all operations. Usually, Q-Values are calculated using data collected from production lines. Besides, other important process parameters are monitored to ensure welding quality. Furthermore, the characteristics of the welding robots and their parts are monitored to avoid operation interruptions.

Currently, engineers in Bosch follow a human heuristic approach to monitoring the quality of the welding process for weldspot quantification to avoid RSW operation interruption. However, manually monitoring such a large number of parameters is a complex task. This motivates us to use RSWO to provide access to ontology-based data that facilitates the quality monitoring process for RSW. Using the semantics and domain knowledge of modelling, reasoning and inference of RSWO, the Bosch dataset enhances the quality monitoring process of RSW.

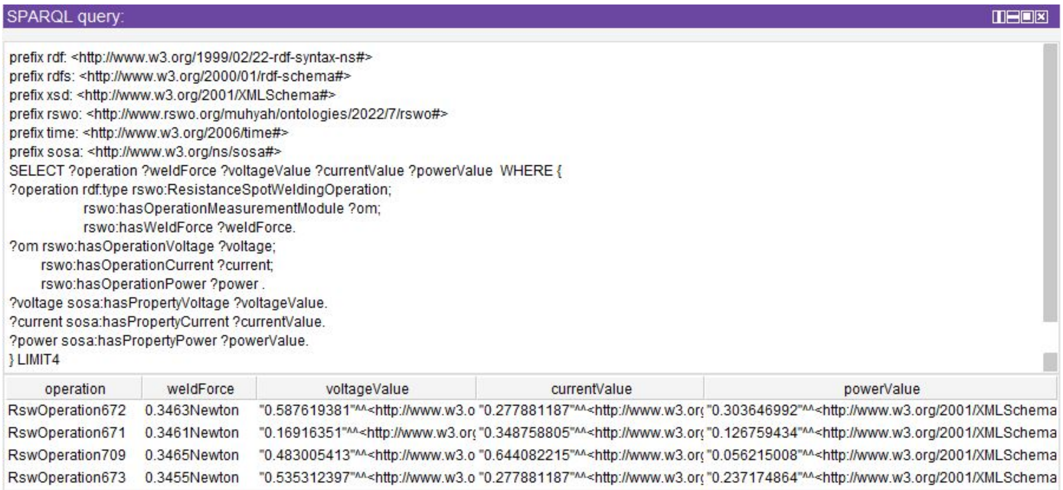
Bosch Welding Data

The datasets used in this use case are acquired from the resistance welding process at Bosch. The datasets are comprised of many formats, such as CSV tables, SQL databases, and text files and inconsistency in variable names and data formats. These datasets contain the attributes that are important to monitor the quality process. Additionally, the data is of two types: namely static and dynamic. As listed in Table 4.4, for instance, Machine_Name, Machine_ID are static variable its value remains the same in each operation while WeldSpotDiameter, TimeStamp, WeldSpotRepatition and others continue to be dynamic variables and their values change in each operation. The data attributes in Bosch datasets are not interconnected but are co-related with each other semantically. To utilise the data effectively for the monitoring of the welding quality, we have used RSWO to integrate and access the data. The data to ontology was manually mapped using the RSWO terms that integrated the data from different sources into uniform data format [183, 173].

Answering the Competency Questions: Bosch Data-based Monitoring the Welding Process

We now demonstrate several examples of using SPARQL queries to answer the competency questions provided by the experts for quality monitoring.

Example 1. Provided the CQs in Table 4.1, we considered CQ1 from the data inspection category to perform basic monitoring tasks for quality welding. This demonstrates the efficiency and usefulness of the RSWO in monitoring and tracking the critical to process parameters of the production resources and processes. The query retrieves the welding force, voltage, current and power values during the particular operations upon successful execution as is shown in Figure 4.7.



SPARQL query:

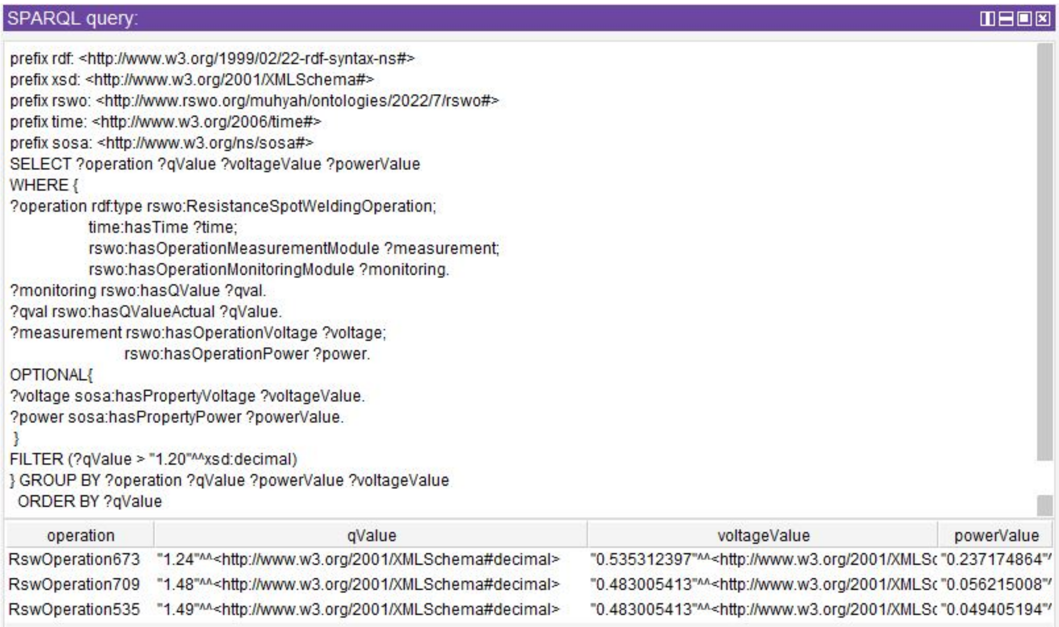
```
prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#>
prefix xsd: <http://www.w3.org/2001/XMLSchema#>
prefix rswo: <http://www.rswo.org/muhyah/ontologies/2022/7/rswo#>
prefix time: <http://www.w3.org/2006/time#>
prefix sosa: <http://www.w3.org/ns/sosa#>
SELECT ?operation ?weldForce ?voltageValue ?currentValue ?powerValue  WHERE {
?operation rdf:type rswo:ResistanceSpotWeldingOperation;
            rswo:hasOperationMeasurementModule ?om;
            rswo:hasWeldForce ?weldForce.
?om rswo:hasOperationVoltage ?voltage;
      rswo:hasOperationCurrent ?current;
      rswo:hasOperationPower ?power .
?voltage sosa:hasPropertyVoltage ?voltageValue.
?current sosa:hasPropertyCurrent ?currentValue.
?power sosa:hasPropertyPower ?powerValue.
} LIMIT4
```

operation	weldForce	voltageValue	currentValue	powerValue
RswOperation672	0.3463Newton	"0.587619381"^^<http://www.w3.o	"0.277881187"^^<http://www.w3.or	"0.303646992"^^<http://www.w3.org/2001/XMLSchema
RswOperation671	0.3461Newton	"0.16916351"^^<http://www.w3.or	"0.348758805"^^<http://www.w3.or	"0.126759434"^^<http://www.w3.org/2001/XMLSchema
RswOperation709	0.3465Newton	"0.483005413"^^<http://www.w3.o	"0.644082215"^^<http://www.w3.or	"0.056215008"^^<http://www.w3.org/2001/XMLSchema
RswOperation673	0.3455Newton	"0.535312397"^^<http://www.w3.o	"0.277881187"^^<http://www.w3.or	"0.237174864"^^<http://www.w3.org/2001/XMLSchema

Figure 4.7: Inspecting the critical to process parameters data CQ1.

Example 2. In the context of the monitoring welding process, the query (CQ 6) from the diagnostics is adopted. The query mentioned in Figure 4.8 is executed to reason about the Q-Values of the weld spot higher than the threshold in any operation.

In Bosch weld production, the increase in the Q-Value is due to the increase in the voltage and power that are considered critical to process parameters. This alternatively raises the spatters occurrence on the near parts of the worksheets. In this context, the query retrieved the critical to process parameters, and the results returned are shown in Figure 4.8. The query has used the `FILTER` keyword to monitor all the Q-Value of the RSW operations and *:hasQValueActual* greater than 1.20. The `OPTIONAL` keyword is binding in this query that enables us to query for data but prevents the query from failing when the requested data is not there.



SPARQL query:

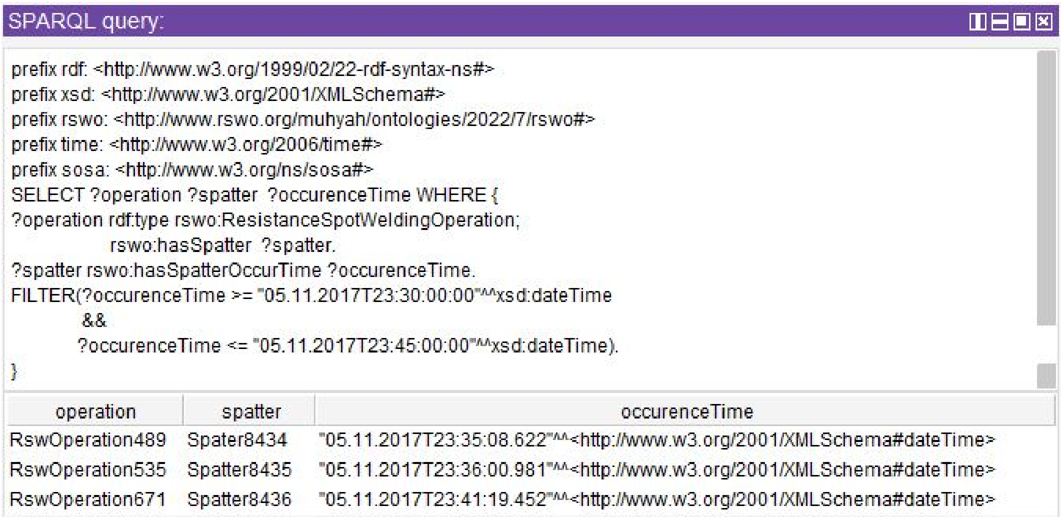
```
prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
prefix xsd: <http://www.w3.org/2001/XMLSchema#>
prefix rswo: <http://www.rswo.org/muhyah/ontologies/2022/7/rswo#>
prefix time: <http://www.w3.org/2006/time#>
prefix sosa: <http://www.w3.org/ns/sosa#>
SELECT ?operation ?qValue ?voltageValue ?powerValue
WHERE {
?operation rdf:type rswo:ResistanceSpotWeldingOperation;
            time:hasTime ?time;
            rswo:hasOperationMeasurementModule ?measurement;
            rswo:hasOperationMonitoringModule ?monitoring.
?monitoring rswo:hasQValue ?qval.
?qval rswo:hasQValueActual ?qValue.
?measurement rswo:hasOperationVoltage ?voltage;
               rswo:hasOperationPower ?power.
OPTIONAL{
?voltage sosa:hasPropertyVoltage ?voltageValue.
?power sosa:hasPropertyPower ?powerValue.
 }
FILTER (?qValue > "1.20"^^xsd:decimal)
} GROUP BY ?operation ?qValue ?powerValue ?voltageValue
  ORDER BY ?qValue
```

operation	qValue	voltageValue	powerValue
RswOperation673	"1.24"^^<http://www.w3.org/2001/XMLSchema#decimal>	"0.535312397"^^<http://www.w3.org/2001/XMLSc	"0.237174864"^
RswOperation709	"1.48"^^<http://www.w3.org/2001/XMLSchema#decimal>	"0.483005413"^^<http://www.w3.org/2001/XMLSc	"0.056215008"^
RswOperation535	"1.49"^^<http://www.w3.org/2001/XMLSchema#decimal>	"0.483005413"^^<http://www.w3.org/2001/XMLSc	"0.049405194"^

Figure 4.8: Result returned by CQ6 to analyze Q-Value, voltage value and power value.

After query processing, both optional and non-optional information is provided. The keyword `GROUP BY` grouped the query results where its order sequence is established by the clause `ORDER BY`.

The proposed query fetched three RSW operations that have a Q-Value greater than the threshold. The voltage and power parameters can be analysed at the same time and it can be observed from the returned results by 4.9 that as voltage and power raises, the Q-Value also increases. A weldspot produced with such high



SPARQL query:

```
prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
prefix xsd: <http://www.w3.org/2001/XMLSchema#>
prefix rswo: <http://www.rswo.org/muhyah/ontologies/2022/7/rswo#>
prefix time: <http://www.w3.org/2006/time#>
prefix sosa: <http://www.w3.org/ns/sosa#>
SELECT ?operation ?spatter  ?occurenceTime WHERE {
?operation rdf:type rswo:ResistanceSpotWeldingOperation;
              rswo:hasSpatter  ?spatter.
?spatter rswo:hasSpatterOccurTime ?occurenceTime.
FILTER(?occurenceTime >= "05.11.2017T23:30:00:00"^^xsd:dateTime
          &&
         ?occurenceTime <= "05.11.2017T23:45:00:00"^^xsd:dateTime).
}
```

operation	spatter	occurenceTime
RswOperation489	Spater8434	"05.11.2017T23:35:08.622"^^<http://www.w3.org/2001/XMLSchema#dateTime>
RswOperation535	Spatter8435	"05.11.2017T23:36:00.981"^^<http://www.w3.org/2001/XMLSchema#dateTime>
RswOperation671	Spatter8436	"05.11.2017T23:41:19.452"^^<http://www.w3.org/2001/XMLSchema#dateTime>

Figure 4.9: Retrieved results of spatters and its occurrence during particular time by utilizing CQ7.

values of critical-to-process parameters can halt the production line and thus can badly affect the parts of the welding machine.

Example 3. Furthermore, in relation to monitoring the welding quality, the spatters that occurred during the operation are usually observed. The occurrence of the spatters badly affects the quality of welding. The spatters occurrence indicates the production line engineers to adjust the critical to process parameters such as weld force level, squeeze time, voltage, current, etc. We utilised CQ7 to find the occurrence of spatters during a particular time. The query along with its result is shown in Figure 4.9. The `FILTER` keyword narrowed down our search between a given particular time. Thus, the above examples demonstrate the usability of RSWO modelling for retrieving the integrated data and information within the RSW domain.

4.4.2 Evaluation on Findable, Accessible, Interoperable, Reusable (FAIR)

We now discuss the evaluation of RSWO using the Findable, Accessible, Interoperable, and Reusable (FAIR) principles. FAIR principles are a set of guidelines [169] that facilitate to build of a coherent and machine-friendly data environment. [5] developed O'FAIRe[6] methodology to encourage ontologies, vocabularies and semantic artefacts compliance following the FAIR guiding principles. It includes fifteen foundational FAIR principles for ontologies and is harmonised with state-of-the-art FAIRness assessment initiatives. The first term *Findable* in the FAIR makes sure that ontology is described with sufficient metadata that can be searched in a registered repository using a persistent and unique identifier. The second term *Accessible* assesses that the ontology can be retrieved in an implementable protocol. The third term of *Interoperable* evaluates the ontology that can be processed in a standard way by other stakeholders. The final *Reusable* term assesses in terms of explicit licences and usage information of the ontology for humans and machines. The O'FAIRe methodology has assessed the RSWO using sixty-one FAIR questions. We have adopted the O'FAIR methodology as it is being used by AgroPortal[7] and IndustryPortal[8] to assess ontologies for FAIR score.

We have evaluated RSWO in line with the O'FAIRe methodology and thus have obtained a total FAIR score of two hundred and seventy one (271) out of 478 which is 56.0%. The RSWO FAIR score against the 15 foundational FAIR principles is as shown in Figure 4.10. Moreover, to make a comparison with other relevant

6 https://github.com/agroportal/fairness
7 http://agroportal.lirmm.fr/ontologies
8 Industry Portal (http://industryportal.enit.fr/) is an online portal for Industrial manufacturing ontologies. It is supported by the OntoCommons project that encourages the researcher to deploy their ontologies designed for Industries especially manufacturing.

ontologies, we have shortlisted all ontologies with FAIR scores greater than 230 from Industry Portal Table 4.5.

Table 4.5: FAIR score with the industry portal ontologies in descending order. The results show our RSWO has relatively high scores compared to other state-of-the-art ontologies.

Ontology	Findable	Accessible	Interoperable	Reusable	TotalScore
EXTRUONT	64(56.63%)	92(81.41%)	63.13(57.91%)	53(37.06%)	272.13(56.0%)
Ours (RSWO)	75(66.37%)	92(81.41%)	54.00(49.54%)	50(34.96%)	271.00(56.0%)
SAREF4INMA	60(53.09%)	92(81.41%)	51.13(46.9%)	48(33.56%)	251.13(52.0%)
SCOR	58(51.32%)	90(79.64%)	45.13(41.4%)	48(33.56%)	241.13(50.0%)
FUNSTEP	54(47.78%)	92(81.41%)	42.00(38.53%)	53(37.06%)	241.00(50.0%)
IOF-MAIN.	51(45.13%)	90(79.64%)	48.00(44.03%)	46(32.16%)	235.00(49.0%)
IMAMO	58(51.32%)	92(81.41%)	41.00(37.61%)	43(30.06%)	234.00(48.0%)
SIMPM	56(49.55%)	92(81.41%)	45.75(41.97%)	39(27.27%)	232.75(48.0%)

It can be observed from the table that EXTRUONT [126] is the only ontology that has a higher FAIR score (272.13) than RSWO (271.00). In comparison with RSWO, the rest of the ontologies have got lower FAIR scores. The RSWO has a high Findable principle score in contrast to other ontologies which is 75 out of 113 which is 66.37%. The Accessible principle score of RSWO is equal to other ontologies. The EXTRUONT has a high Interoperable principle score of 63.13 out of 109 while RSWO has an acceptably low Interoperable principle score of 54 and comes second in the list for interoperability score. Other ontologies have low Interoperable score than RSWO. The Reusable principle score of the RSWO ontology is 50 which is acceptably low than EXTRUONT and FUNSTEP[9]. The RSWO has a high Reusable principle score to that of SAREF4INMA [132] and SCOR. The IOF-Maintenance (IOF-MAINT.) [66], IMAMO [78] and SIMPM has a low score of Reusable principle in contrast to RSWO.

Due to the reason that the RSWO is not yet included in a specific community, therefore, it received Score:0.0 for the FAIR principles question (*"R1.3Q2": Is the ontology included in a specific community set or group?)*. The metadata of the RSWO provides rich information that gives a higher score than other ontologies.

4.4.3 OOPS!

The Ontology Pitfall Scanner (OOPS) assesses the ontology in its creation process by looking at the design imperfections from a list of 41 recurring flaws, which are categorised as minor, important and critical. The OOPS can identify the majority of them (33 out of 41 dangers) semi-automatically. The OOPS tool has been used frequently to find minor, important, and critical changes.

9 http://www.funstep.org/ontology/

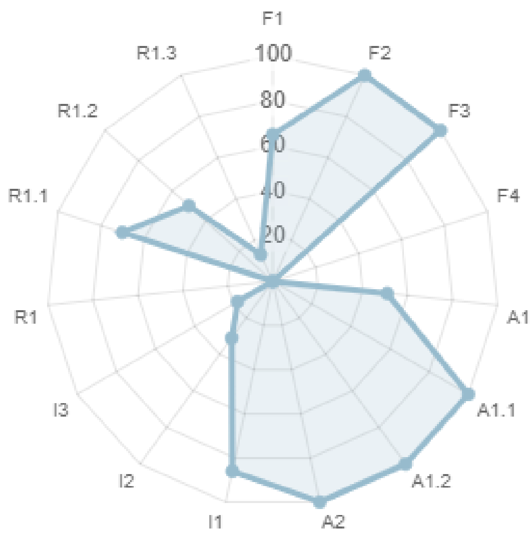


Figure 4.10: FAIR result of RSWO

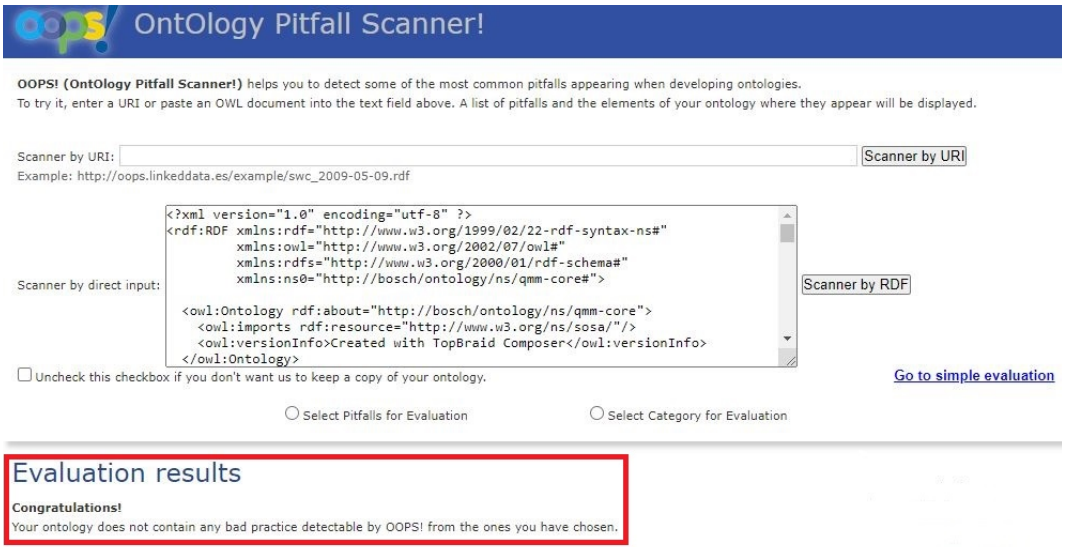


Figure 4.11: OOPs assessment shows RSWO does not contain any bad practice detectable by OOPS!.

The RSWO assessment with OOPS yields some minor pitfalls that have no bearing on the ontology reasoning, consistency or/and applications. The issue of *unconnected ontology elements*, *several classes with same labels*, and *missing domain and range* reported are mainly due to the inheritance of SSN ontology terms and relations. The assessment of the RSWO results by OOPS is as shown in Figure 4.11. Furthermore, the OOPs tool assesses criteria such as *clarity*, *completeness*, *conciseness* and *consistency*. The criteria of how the RSWO applies them in line with an explanation, are listed below.

- **Clarity:** The ontological terms defined to represent the classes, concepts, and relations of all the modules, contain unambiguous names, and annotations. The annotations aid in the readability of humans to avoid uncertainty and difficulty during the insertion of data elements.

- **Completeness:** The ontology is capable of answering all competency questions as defined by the industry experts, correctly describing the domain for which the ontology has been created.
- **Conciseness:** The industry knowledge represented by the ontology that is gathered in line with the sources, particularly those in the domains of electrodes, welding materials and processes, and enterprise and their production lines, thus eliminating the irrelevant information altogether.
- **Consistency:** The Fact++[10] reasoners have been applied to find inconsistencies in the RSWO. Accordingly, the reasoner has not found any inconsistencies in the accordingly developed ontology.

4.4.4 OntoMetrics

The RSWO has been assessed with the OntoMetrics tool to reflect some notions of ontology richness with five metrics. To the best of our knowledge, there is no publicly available resistance spot welding ontology with which we can directly compare the RSWO. However, we considered Library ontology [54] and EUCISE-OWL [130] that have used Ontometrics. Table 4.6 contains the metrics computed by OntoMetrics that highlight the ontology's most intriguing domain-level features.

Table 4.6: Evaluation of RSWO using OntoMetrics tool.

Metric	Library Ontology	EUCISE-OWL	RSWO
Attribute richness	2.692308	1.694805	1.613929
Inheritance richness	0.923077	0.967532	0.973007
Relationship richness	0.2	0.464029	0.495000
Average population	0.615385	5.603896	4.981892
Class richness	0.692308	0.558442	0.830357

- *Attribute richness:* It calculates the average number of attributes (slots) per class, indicating the quality of the ontology design and the quantity of information that can be included in the instance data. The RSWO has an attribute richness value of 1.613929 which is lower in comparison to Library ontology and EUCISE-OWL.
- *Inheritance richness:* The term inheritance richness refers to the average number of sub-classes per class that describes the distribution of information along the multiple levels of the ontology inheritance tree. The value of 0.973007 highlights that the RSWO covers a good range of concepts.

10 http://owl.man.ac.uk/factplusplus/

- *Relationship richness:* It indicates the variety of relational types and is calculated as the ratio of non-inheritance relationships to all the relationships in the ontology. OntoMetrics tool reported a value of 0.495000 for relationship richness which is higher than Library ontology and EUCISE-OWL.
- *Average population:* It provides information about the quality of the ontology population that corresponds to the ratio of instances to classes. The RSWO has an average population value of 4.981892.
- *Classes richness:* It represents the distribution of instances among classes. The overall number of classes is compared to the number of RSWO classes that have instances providing an overview of how well the knowledge-base uses the knowledge represented by the schema classes. The class richness of the RSWO value is 0.830357.

4.5 SUMMARY

In Chapter 5, we have discussed the development of RSWO ontology in line with the RGOM. Then we have highlighted the RSWO development process by outlining the steps covering the stages of ontology requirement specification, formalizing concepts, ontology validation, and ontology publication and maintenance. The RSWO ontology representing the knowledge of RSW has then been discussed. Moreover, the evaluation conducted from fours dimension is then provided. In the next chapter, we conclude the overall thesis and highlight the directions for this research in the future.

5 ADAPTABILITY OF RGOM ON REALISTIC DATA

This chapter presents the football manufacturing production line and its various machines. The acquisition of the dataset values and mapping them to the RGOM process are discussed. Through answering competency questions, we will evaluate the effectiveness of the proposed approach in capturing the essential knowledge required for understanding and managing the football manufacturing production line. This evaluation will ultimately help assess the efficiency of RGOM, with a focus on minimizing the need for extensive modifications.

5.1 FOOTBALL MANUFACTURING PRODUCTION LINE

This section explains the data acquisition and dataset construction. The production floor consists of several production lines, each consisting of nine machines with five operator personnel performing manual operations. A typical football construction requires a Thermoplastic polyurethane (TPU) roll, football cores, printing colours, glue, laser cutting machine, Oval Printing machine, high-frequency machine, glue spraying machine, heat activating conveyor machine, forming moulding machine, ball shaping machine, ball seam gluing machine, and heat drying machine. During a single production process, these machines perform different processes on different materials and produce four footballs as a finished product. Figure 5.1 depicts the single-process flow of football production and the flow of the sensor data in a manufacturing production line. The production includes 9 machines that are explained in the following subsections.

5.1.1 Laser Cutting Machine

Laser Cutting Machine (LCM) is a manufacturing machine that performs the first process in football production, known as the cutting TPU process. A laser-based cutting tool is hosted by the LCM to convert TPU rolls into patches. The laser

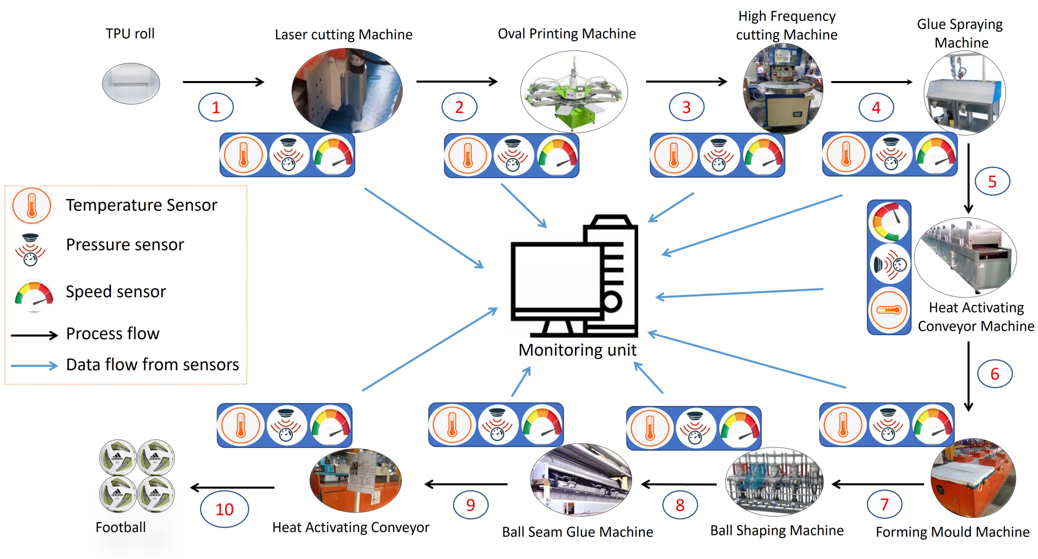


Figure 5.1: Flow of a single production process. The black arrows show the process flow in the production line and the blue arrows show the data flow from the sensors to the monitoring unit. ① In a single process, the TPU roll is fed into laser cutting machine. ② laser cutting machine converted the TPU roll into patches. ③ patches are printed via squeegee by the oval printing machine. ④ printed patches are cut into panels. ⑤ Back sides of panels and cores are sprayed with glue. ⑥ glued panels and cores are passed by the heated conveyor to form a moulding machine. ⑦ cores and panels are moulded. ⑧ Balling shaping machine gives football shape to the moulded cores and panels. ⑨ The gaps between the panels are sealed with glue via a Ball seam glue machine. ⑩ The glue is dried by the heat-activating conveyor and 4 footballs are produced.

is rotated with the help of a motor. The LCM produces six patches in a single process as a duplicate material, which is submitted as input to the Oval Printing Machine (OPM).

5.1.2 Oval Printing Machine

The OPM hosts nine tools, that is three beds, three squeegees, and three heaters. Each tool has a different function, for instance, the bed acts as a container to hold the patch, each squeegee prints different colours (color 1, color 2 and color 3), and the heater driers the printed patches with a temperature ranging from 55°C to 65°C. The squeegee has various attributes such as power consumption, pressure, hardness, etc. The six patches produced by LCM are passed as input material to the OPM. This machine performs a total of eight step-wise processes to print the colours on the patches in a single production process. In the first process, patch one is placed on the first bed, which is then forwarded to the squeegee for printing colour one. The printed patch placed on bed one is dried in the heat of heater one. In the second process performed by OPM, the same operation is repeated by squeegee one and heater one on patch two placed on bed two. Squeegee two prints colour two on patch one and heater two dry the printed patch. Now, the patch contains two colours. In the third process by OPM, bed three contains patch three, while bed two and bed one contain patch two and patch three, respectively. At the end of the third process, beds one, two, and three contain patches one, two, and three with printed colours one, one and two, one, two and three, respectively. The rest of the eight processes are performed in the same flow to print the three colours on the 4 to 6 patches. The output of the machine is passed to the High-Frequency Cutting Machine (HFCM).

5.1.3 High-Frequency Cutting Machine

The dry-printed patches are transferred to the bed of an HFCM. A die-cutting tool hosted by HFCM cuts off the printed patch into four panels. HFCM performs a single process for each patch, a total of 6 processes are performed, and 24 panels are produced in a single production process. An operator plucks the panels from the HFCM and matches the panels for a single football which is passed to the next machine known as the Glue Spraying Machine (GSM).

5.1.4 Glue Spraying Machine

The GSM in the production line receives two input materials, i.e., 6 panels and a rubber material inside the football known as the core. The glue is sprayed on the backside of the 6 panels and core with the help of a needle (the diameter of the needle is 0.5 millimetres) hosted on the GSM. The glue panels and core are sent to heat activating conveyor machine.

5.1.5 Heat Activating Conveyor Machine

Heat Activating Conveyor (HAC) is a conveyor machine. It has a heating tool that generates heat with a temperature ranging from 45 to 55 degrees Celsius. The function of the HAC is to dry the glue on the backside of the core and panels. The dried core and panels are sent to a ball-shaping machine.

5.1.6 Ball Shaping Machine

The panels attached to the core are provided as input to the ball-shaping machine. The core and panels are placed inside the ball-shaping machine, where pressure with a 60-degree Celsius temperature is applied to convert the panels on the core into a round shape. This manufacturing process results in the production of a semi-finished football. The semi-finished football is passed to the form moulding machine.

5.1.7 Form Moulding Machine

A form moulding machine is an assembly machine. It is used to assemble the panels on the core. It performs a total of 4 processes in a single production process. This machine output is provided to the ball seam glue machine as an input.

5.1.8 Ball Seam Glue Machine

The ball seam glue machine performs a manufacturing process. The ball seam glue machine hosts a needle with a diameter of 0.5 millimetres, aiming to fill the gap between panels with the glue. The filled gap of the product is then sent to the heat-drying conveyor machine.

5.1.9 Heat Drying Conveyor Machine

This is the final machine in the football production process. The function of the heat drying conveyor machine is similar to that of machine 5. The glue (wet) football is then passed through a Heat drying conveyor to become dry. After the process of machine 9, operators clean the ball, pack it in polybags, and then in the carton.

5.2 DATA DEFINITION AND DATA ATTRIBUTING

The sensors installed on the machines in the I4.0-based production line generate data that are sent to the monitoring unit. In order to collect the first real instance of the data, several meetings were held with production line managers and engineers of Forward Group Limited regarding the operations of the machines, resources, processes, and production. It generally involved recording the power consumption, temperature, pressure, location, and type of process performed at a given timestamp by the machines. Also, the working status and rotational speed of the motor and other attributes were also recorded. Table 5.1 depicts the tools and machines parameters which include machine name, timestamp, temperature, pressure, power, laser die, bed, squeegee, heater, and high-frequency die and many others.

Table 5.1: Overview of attributes for each machine in the production line.

Attributes	Machine1	Machine2	Machine3	Machine4	Machine5	Machine6	Machine7	Machine8	Machine9
Machine Name	Laser cutting	Oval Printing	High Frequency Cutting	Glue Sprayer	Heating activating panel.	Forming Mould	Ball Shaping	Ball Seam Glue	Heat Drying Conveyor
Process Name	Cutting TPU	Oval Printing	High frequency cutting	backside gluing process	Heating Glue process	Form moulding process	Ball shaping process	Seam Gluing process	Heating Glue process
Hosted Tools	Laser Die	Bed, Squeegee, Heater	High Freq. Die	Needle	Heater	Moulding Die	Die	Needle	Heater
Tool Attributes	Cutting speed, Laser Power	Power, Ink Viscosity, hardness	Temperature, Frequency	Needle Diameter, Glue quantity	Heater Temperature	Die diameter, Mould pressure	Die Diameter	Needle diameter Glue quantity	Heater Temperature
Input Material	TPU roll	Patch	printed Patch	Panel, core, glue.	glued unattached core, panels	Dried core and panels	Semi-finished Football	Semi-finished Football	Semi-finished Football
Output Material	Patch	printed patch	Panel	glued unattached panels, core	Dried attached core and panels	Semi-finished Football	Semi-finished Football	Semi-finished Football	Finished Football

Initially, the collected data were stored in a file comprising two types of attributes such as static attributes and variable attributes under the supervision of production line engineers and managers. Static attributes contain those attributes of the machine whose values remain the same in each process of manufacturing e.g. Machine model, process location, motor ID etc. On the other hand, variable attribute value changes in each process based on the performance condition of the machine, e.g. temperature, pressure, diameter, etc. Using the minimum and maximum values as well as the real value measured by the sensor, we are able to obtain the realistic data with the help of uniform probability distribution in each sub-processes. Uniform probability distribution takes input in a range bounded between the possible minimum and maximum value describing the possible likelihood and values of a variable [32]. The uniform probability distribution is utilised to generate new instances of temperature values for machines in the production line. For a given machine, say `machine1`, we compute a new temperature value ($\text{Temp}_{\text{synthetic}}$) from the real or actual temperature value ($\text{Temp}_{\text{real}}$) using a value generated by a uniform probability distribution within a specified range. The range is defined by the minimum and maximum reference values (a and b), as provided by production line supervisors and engineers.

The computation of $\text{Temp}_{\text{synthetic}}$ for `machine1` during a specific process like the TPU roll-cutting process can be formally represented as:

$$\text{Temp}_{\text{synthetic}} = \text{Temp}_{\text{real}} + \Delta T, \tag{5.1}$$

where ΔT is a variable temperature change determined by a uniform probability distribution within the range $[a, b]$. The uniform distribution is defined as:

$$\Delta T = \begin{cases} 0, & \text{if } x < a \\ \text{random.uniform}[a, b], & \text{if } a \leqslant x \leqslant b \\ 0, & \text{if } x > b \end{cases} \tag{5.2}$$

For instance, if the real temperature ($\text{Temp}_{\text{real}}$) of `machine1` is 41°C, and the range for ΔT is set between 1 and 5 (with $a = 1$ and $b = 5$), then ΔT is a random value obtained from the uniform distribution over the interval $[1, 5]$. This random value reflects the variability in temperature readings. Adding this ΔT value to $\text{Temp}_{\text{real}}$ gives the synthetic temperature value ($\text{Temp}_{\text{synthetic}}$) for `machine1`. This approach ensures the consistency and validity of the generated values, as verified by production line engineers. The same method is applied to other variable attributes in the production line.

5.3 AN APPROACH TO DATA INTEGRATION USING RGOM

This section describes the approach to building I4.0KG from the football production line data. As the acquired data becomes increasingly available in data storage, it is stored disregarding their semantics and relations. This restricts the usability of the data, e.g., querying information, data analysis, etc. Therefore, there is a pressing need to represent this data in a semantic representation, i.e., Linked Open Data [119]. Semantically enriched representation of data or KGs adds meaning and context to data through ontologies and vocabularies that make it more easily understood and interpreted by humans and machines [110]. This leads to several benefits including improved data integration, data understanding, data interoperability, and faster discovery of knowledge via more powerful data querying and analysis [50, 74, 137].

Figure 5.2 illustrates the workflow for constructing the KG which is comprised of four layers, Layer 1: Unstructured Data Sources, Layer 2: Knowledge Graph Construction, Layer 3: Football Production Line Knowledge Graph, and Layer 4:

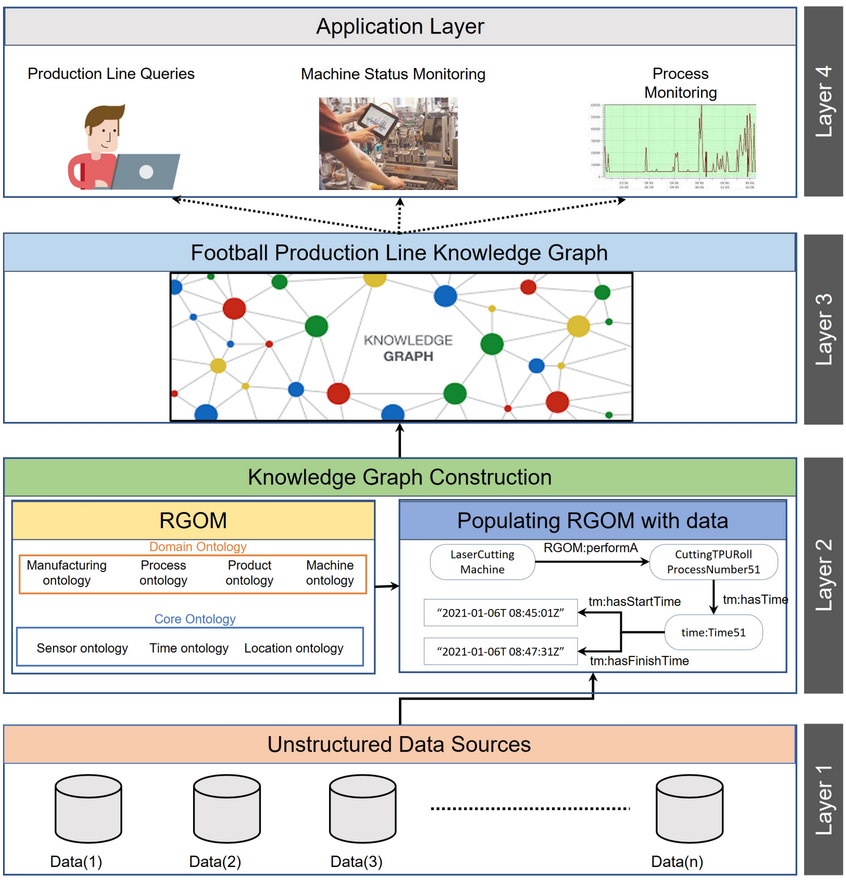


Figure 5.2: An illustration of the approach to integrate data

Users and Applications. The following subsection describes the layers and the interaction of the different components.

5.3.1 Layer 1: Unstructured Data Sources

The tools and sensors hosted by machines generate a huge amount of data at different timestamps in the manufacturing production line. The generated data is usually unstructured and is usually stored in different formats (e.g., TXT, CSV, XML, JSON, etc.) by the data storage. Accessing unstructured data in terms of information requires a lot of pre-processing and manual efforts. It is difficult for the production line staff to access information from unstructured data.

5.3.2 Layer 2: Knowledge Graph Construction

The goal of building an I4.0KG can be accomplished with the Reference Generalized Ontological model (RGOM), to which the data from Layer 1 is mapped to construct a KG. Figure 5.3 depicts the pipeline to construct I4.0KG. The RGOM and data sources are given as input to the reader component. The reader component read the ontology resources i.e., class, object and data properties from the RGOM and parsed data records from the data sources. The data instances are mapped with RGOM classes and properties by the mapping components.

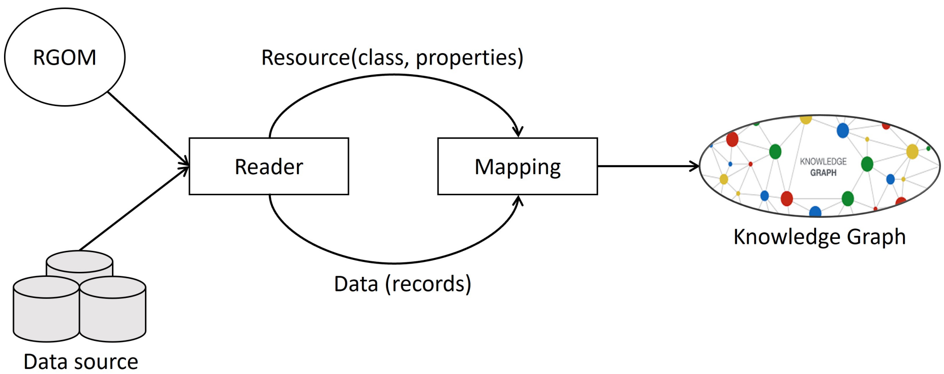


Figure 5.3: Pipeline for knowledge graph construction.

Algorithm 1 presents the process of populating data, beginning with the initialization of the ontology classes and properties for the data population. Initially, the algorithm eliminates any null or empty values. Subsequently, it checks each record in the dataset, parsing the values in the data columns into their corresponding datatypes. A resource individual of a specific class type is generated. The datatype value is converted into a literal and linked to the resource individual through a data property. This procedure is consistently applied to all the data.

Utilising the semantics built in the ontology model, an object property is selected to establish a triple that connects the subject and object individuals. For example, as can be seen from Figure 2 in Layer 2 of the data population into ontology terms, the algorithm gets the class type *ManufacturingMachine* and *ManufacturingProcess* from the ontology file, and iterates over the data records in the data file. Similarly, the columns *machine1* and *machine1_process* from the spreadsheets are created as a subject and object with the aforementioned class types, respectively in a single iteration.

Consequently, data is successfully populated to RGOM. An example is presented in Figure 5.4, which shows *machine_9* has individuals along with their features i.e *processMaterial*, *hasTools*, *consumesPower*, *performsProcess*, *isInLocation*.

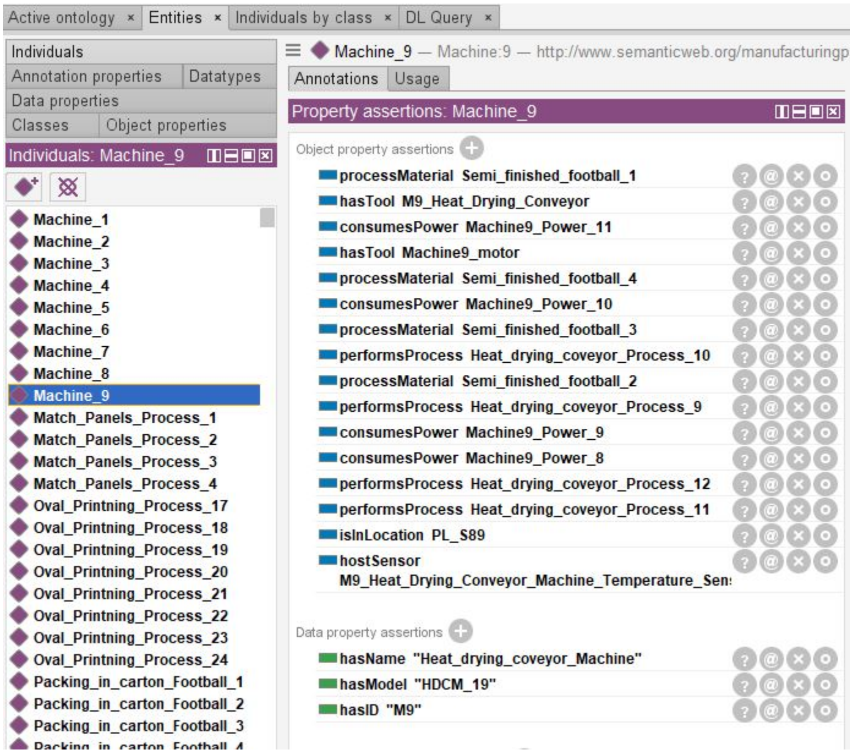


Figure 5.4: Illustration of machine 9 instances with their attributes instances.

5.3.3 Layer 3:Knowledge Graph

Layer 3 represents the KG generated from layer 2. It is often used to store interlinked descriptions of entities – objects, events, situations, or abstract concepts. Whilst the data is mapped into the ontology concepts and properties, it becomes a KG[1].

1 https://web.stanford.edu/class/cs520/

Algorithm 1 mapping data to rgom

```
Input: ontologymodel, data
Output: ontologymodel.write()
namespace ← IRI   // Define the namespace IRI for the ontology model
 resource ← ontologymodel.getResource(namespaces +"classtype"); // e.g., Machine, Process etc.
 r1← Null ; // Initialize r1 to null, storing the first individual created
 r2← Null ; // Initialize r2 to null, storing the second individual created
 ⋮
 rn← Null ; //Initialize rn to null, storing the n^th individual created
 objectProperty = ontologymodel.getProperty(namespace + "objectproperty"); // e.g., performProcess etc.
 dataProperty = ontologymodel.getProperty(namespace + "dataproperty"); // e.g., hasName, hasTime etc.
for each record in data do
    if (record.column{1} != Null && record.column{1} != ("")) // parse first column (index) of data
     then
        dataType value1 = parseDatatype(record.column{1}); // parsing string to datatype
        r1 ← ontologymodel.createIndividual(namespace+"resource_name", resource);
        r1.addProperty(dataProperty, model.createTypedLiteral(value1));
    end
    if (record.column{2} != Null && record.column{2} != ("")) //parse second column (index) of data
     then
        dataType value2 = parseDatatype(record.column{2}); // parsing string to datatype
        r2 ← ontologymodel.createIndividual(namespace+ "resource_name" ,resource);
        r2.addProperty(dataProperty, model.createTypedLiteral(value2));
        r2.addProperty(objectProperty, r1);
    end
    ⋮
    if (record.column{i^th} != Null && record.column{i^th} != ("")) //parse i^th column (index) of data
     then
        dataType valuen = parseDatatype(record.column{i^th}); //parsing string to datatype
        rn ← ontologymodel.createIndividual(namespace+ "resource_name", resource);
        rn.addProperty(dataProperty, model.createTypedLiteral(valuen));
        rn.addProperty(objectProperty, r2);
        end
    end
    KG = ontologymodel.write(path); //store the KG on the provided path.
```

In our case, the I4.0KG contains the football production line data that the engineers use to analyze the machines and process CTP parameters on a daily basis. The production process of a football contains nine machines and each performs different sub-processes. The data produced during this process contains the domain knowledge of the machines, tools hosted on the machines, processes performed on them, tools deployed on the machines, tool's critical parameters, and contextual data generated by the sensors hosted on the machines at some timestamp. The approach presented in Section 4.3 is followed to semantically integrate the data. At first, we gathered the data sources containing the data about all the machines which are then analyzed in line with the RGOM classes and relations. Next, the data is populated to the ontology terms, i.e., classes and relations with Jena API[2]. Upon the population of data into ontology terms, an RDF triple store is obtained, known as an I4.0KG.

To produce a single football, an average of 1730 triples, 1355 logical triples, and 233 declaration triples are produced. In one hour of the production line, a total of 9 main processes are executed, producing 36 footballs and 22150 triples on average from 2903 individuals.

Besides, three I4.0 KG-based datasets are produced to provide the researcher's community to evaluate their tools and techniques. These datasets are comprised of ten days, twenty days, and thirty days of data from a football production line. The types of machines and their parameters are explained in Section 4.1. The total number of axioms, logical axiom count, declaration axiom count, and individual count in each KG are illustrated in Table 5.2. The number of classes, object properties, and datatype properties are the same for each KG.

Table 5.2: Summary of the axioms in each Knowledge graph

KGs	Total number of Axiom	Logical Axiom Count	Declaration axioms Count	Individual Count
10 Days	525865	525503	225	166273
20 Days	1050535	1050173	225	332363
30 Days	1470280	1469918	225	465238

5.3.4 Layer 4: Users and Application Layer

After the construction of the KG, several queries are provided by the production line engineers and supervisors to find the usefulness of the KG. The SPARQL endpoint at the application layer of I4.0KG paves the way for users to access the required information embedded in the KG. Given a production line where the job at hand is to utilise the query drawn from listing 4.1 in order to access the type of machines and their names involved.

2 https://jena.apache.org

```
PREFIX smo: <http://www.semanticweb.org/manufacturingproductionline/>
SELECT *
WHERE {
        {?Machine a smo:ProcessingMachine;} UNION
        {?Machine a smo:AssemblingMachine;}
        ?Machine smo:hasName ?Name.}
```

Listing 4.1. Query to retrieve the machines involved in the production line.

Figure 5.5 shows the results returned from the listing 4.1 query. It can be seen from the figure that the production line consists of a single assembly line and eight processing machines, each with their name.

localhost:8890/sparql?default-graph-uri=&query=PREFIX+smo%3A+<http%3A%2F%2Fwww.seman

Machine	Name
http://www.semanticweb.org/manufacturingproductionline/data/Machine_1	"Laser_cutting_Machine"
http://www.semanticweb.org/manufacturingproductionline/data/Machine_2	"Oval_printing_Machine"
http://www.semanticweb.org/manufacturingproductionline/data/Machine_3	"High_frequency_cutting_Machine"
http://www.semanticweb.org/manufacturingproductionline/data/Machine_4	"Gluing_Spraying_Machine"
http://www.semanticweb.org/manufacturingproductionline/data/Machine_5	"Heat_Activating_conveyor_Machine"
http://www.semanticweb.org/manufacturingproductionline/data/Machine_7	"Ball_shaping_machine"
http://www.semanticweb.org/manufacturingproductionline/data/Machine_8	"Ball_seam_gluing_machine"
http://www.semanticweb.org/manufacturingproductionline/data/Machine_9	"Heat_drying_coveyor_Machine"
http://www.semanticweb.org/manufacturingproductionline/data/Machine_6	"Forming_Mold_Machine"

Figure 5.5: Listing 4.1 query provides the number of machines involved in the production line with their names.

Furthermore, the production line manager can utilise the Listing 2 query to find the tools present on a machine. The result of the listing 4.2 query is shown in Figure 5.6. It can be seen from the figure that machine 2 has a name and uses different types of tools such as one motor, three beds, three heaters, and three squeegees. The query fetches machine 2 name and different types of tools that reside in the KG.

```
PREFIX smo: <http://www.semanticweb.org/manufacturingproductionline/>
PREFIX d: <http://www.semanticweb.org/manufacturingproductionline/data/>
SELECT ?machine ?tool
WHERE { d:Machine_2 smo:hasName ?machine;
      smo:hasTool ?tool. }
```

Listing 4.2. Query to retrieve the tools hosted on machine 2.

Similarly, an engineer from the maintenance department wants to query the KG for CTP parameters to check the current observation of the sensor or the status of the motor. For instance, a maintenance engineer can retrieve the status of a motor at a particular period of time by using the query in Listing 4.3. The query fetches the status of the motor at different timestamps as illustrated in Figure 5.7.

localhost:8890/sparql?default-graph-uri=&query=PREFIX+smo%3A+<http%3A%2F%2Fwww.semanticweb.org

machine	tool
"Oval_printing_Machine"	http://www.semanticweb.org/manufacturingproductionline/data/Machine2_Bed_1
"Oval_printing_Machine"	http://www.semanticweb.org/manufacturingproductionline/data/Machine2_Bed_2
"Oval_printing_Machine"	http://www.semanticweb.org/manufacturingproductionline/data/Machine2_Bed_3
"Oval_printing_Machine"	http://www.semanticweb.org/manufacturingproductionline/data/Machine2_Heater_1
"Oval_printing_Machine"	http://www.semanticweb.org/manufacturingproductionline/data/Machine2_Heater_2
"Oval_printing_Machine"	http://www.semanticweb.org/manufacturingproductionline/data/Machine2_Heater_3
"Oval_printing_Machine"	http://www.semanticweb.org/manufacturingproductionline/data/Machine2_motor2
"Oval_printing_Machine"	http://www.semanticweb.org/manufacturingproductionline/data/machine2_Squeegee_1
"Oval_printing_Machine"	http://www.semanticweb.org/manufacturingproductionline/data/machine2_Squeegee_2
"Oval_printing_Machine"	http://www.semanticweb.org/manufacturingproductionline/data/machine2_Squeegee_3

Figure 5.6: Listing 4.2 query provides tools hosted by machine 2 (Oval Printing Machine).

```
PREFIX smo: <http://www.semanticweb.org/manufacturingproductionline/>
PREFIX d: <http://www.semanticweb.org/manufacturingproductionline/data/>
PREFIX tm: <http://www.w3.org/2006/time#>
SELECT DISTINCT ?Motor_Name ?Status ?Start_time
WHERE { d:Machine_1 smo:hasTool ?motor.
    ?motor smo:hasName ?Motor_Name.
    ?motor smo:hasMotorState ?state.
    ?process tm:hasTime ?time.
    ?state smo:hasState ?Status.
    ?time tm:hasStartTime ?Start_time.
     FILTER (?Start_time > "2021-06-01T 10:11:00Z"^^xsd:dateTime &&
     ?Start_time < "2021-06-01T 10:12:55Z"^^xsd:dateTime). }
```

Listing 4.3: Query to retrieve the status of machine 2 motors at a certain time period.

localhost:8890/sparql?default-

Motor_Name	Status	Start_time
"LCM_Motor"	"working"	2021-06-01T 10:12:13Z
"LCM_Motor"	"working"	2021-06-01T 10:11:57Z
"LCM_Motor"	"working"	2021-06-01T 10:12:28Z
"LCM_Motor"	"working"	2021-06-01T 10:11:32Z

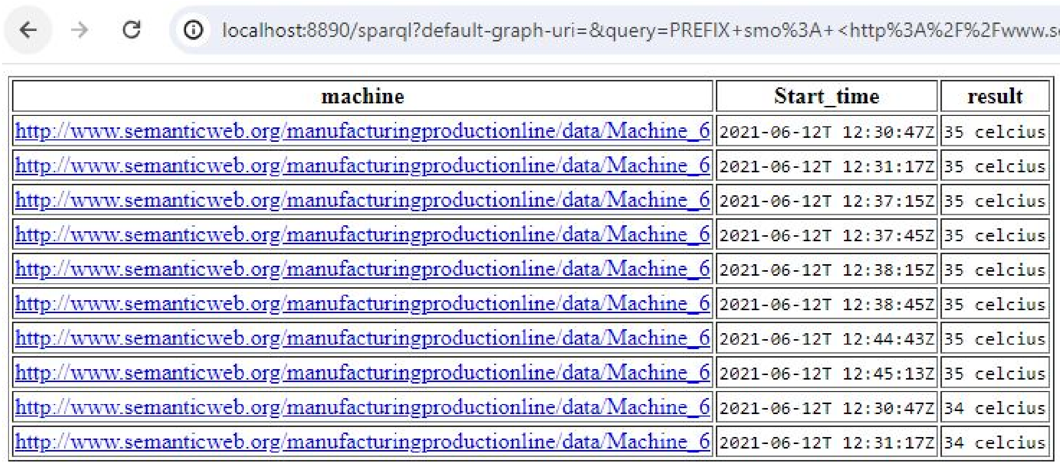
Figure 5.7: Result returned by the query in listing 4.3

In order to retrieve the temperature (a CTP parameter) query in listing 4.4 is utilised. The reuse principle of Linked Open Data has been followed by reusing the *SOSA* vocabulary as depicted in listing 4.4. Figure 5.8 shows the fetched results of the query in listing 4.4.

Furthermore, a query in Listing 4.5 is used to find the total number of processes performed by a machine and the total number of tools that each machine used during a given time. The query returns information about all the machines

```
PREFIX smo: <http://www.semanticweb.org/manufacturingproductionline/>
PREFIX d: <http://www.semanticweb.org/manufacturingproductionline/data/>
PREFIX tm: <http://www.w3.org/2006/time#>
PREFIX sosa:  <http://www.w3.org/ns/sosa#>
SELECT DISTINCT ?machine ?Start_time ?result
WHERE { ?machine  smo:hasTool  ?tool.
      ?tool  sosa:madeObservation  ?observation.
      ?observation  sosa:hasSimpleResult  ?result.
      ?time  tm:hasStartTime  ?Start_time.
      FILTER (?tool != d:M6_Folding_Mold_machine_Pressure_Sensor.)}
```

Listing 4.4 Query to retrieve the CTP parameter (Temperature) with time.



machine	Start_time	result
http://www.semanticweb.org/manufacturingproductionline/data/Machine_6	2021-06-12T 12:30:47Z	35 celcius
http://www.semanticweb.org/manufacturingproductionline/data/Machine_6	2021-06-12T 12:31:17Z	35 celcius
http://www.semanticweb.org/manufacturingproductionline/data/Machine_6	2021-06-12T 12:37:15Z	35 celcius
http://www.semanticweb.org/manufacturingproductionline/data/Machine_6	2021-06-12T 12:37:45Z	35 celcius
http://www.semanticweb.org/manufacturingproductionline/data/Machine_6	2021-06-12T 12:38:15Z	35 celcius
http://www.semanticweb.org/manufacturingproductionline/data/Machine_6	2021-06-12T 12:38:45Z	35 celcius
http://www.semanticweb.org/manufacturingproductionline/data/Machine_6	2021-06-12T 12:44:43Z	35 celcius
http://www.semanticweb.org/manufacturingproductionline/data/Machine_6	2021-06-12T 12:45:13Z	35 celcius
http://www.semanticweb.org/manufacturingproductionline/data/Machine_6	2021-06-12T 12:30:47Z	34 celcius
http://www.semanticweb.org/manufacturingproductionline/data/Machine_6	2021-06-12T 12:31:17Z	34 celcius

Figure 5.8: Listing 4.4 query provides the retrieval of the temperature (a CTP parameter).

with the total number of processes they performed during a given time and the total number of tools used by them, shown in Figure 5.9. For instance, in the list, machine_1 has performed a total of 100 processes and used a total of three tools during the time from `12:55:13` to `14:36:04`.

```
PREFIX smo: <http://www.semanticweb.org/manufacturingproductionline/>
PREFIX d: <http://www.semanticweb.org/manufacturingproductionline/data/>
SELECT DISTINCT ?machine (count(distinct ?process) as ?process_count)
(count(distinct ?tool) as ?tool_count)
WHERE { ?machine  smo:performsProcess  ?process.
      ?machine  smo:hasTool  ?tool.
      ?process  sosa:hasTime  ?time.
    ?time tm:hasStartTime ?start_time.
    ?time tm:hasFinishTime ?finish_time.
    FILTER (?start_time > "2021-06-08T 12:55:13Z"^^xsd:dateTime &&
    ?finish_time < "2021-06-12T 14:36:04Z"^^xsd:dateTime
} GROUP BY ?machine order by ?machine
```

Listing 4.5 Query to retrieve the count of processes performed by machines and the count of tools used by them during a time period.

localhost:8890/sparql?default-graph-uri=&query=PREFIX+smo%3A+<http%3A%2F%2Fwww.semanticweb.org%2Fmanufacturingproductionlin

machine	process_count	tool_count
http://www.semanticweb.org/manufacturingproductionline/data/Machine_1	110	3
http://www.semanticweb.org/manufacturingproductionline/data/Machine_2	1054	10
http://www.semanticweb.org/manufacturingproductionline/data/Machine_3	792	2
http://www.semanticweb.org/manufacturingproductionline/data/Machine_5	134	3
http://www.semanticweb.org/manufacturingproductionline/data/Machine_6	536	3
http://www.semanticweb.org/manufacturingproductionline/data/Machine_7	538	3
http://www.semanticweb.org/manufacturingproductionline/data/Machine_8	539	3
http://www.semanticweb.org/manufacturingproductionline/data/Machine_9	539	3

Figure 5.9: Listing 4.5 query provides the count of processes performed by machines and the count of tools used by them during a time period.

5.4 DISCUSSION

The current usecase showcases the potential of applying the RGOM to a larger picture. The concepts introduced in RGOM can be utilised for the majority of production lines which will facilitate a wider use of our approach for generating KGs for different use cases. For example, this approach can be extended not only to other similar manufacturing processes such as volleyball and rugby ball production, but also to other more generic production lines that incorporate welding processes. This can help other industries map their customised data into RGOM and construct an industry-specific KG. Additionally, to build KGs for a different manufacturing industry, one should adopt the RGOM framework to add definitions of required classes and relations. Adopting a similar mechanism will help digital transformation for those who have not set up a Linked Data-based production line.

The I4.0 KG dataset can be utilised in predictive maintenance [25]. For example, one useful use case of the I4.0 KG dataset is predicting the temperature of a machine. In manufacturing factories, the temperature of the machine is of high significance and critical. During the process execution, the tools are operating under a set point. The increase in temperature can adversely affect the machine which impacts the product quality. The assessment of the temperature information enables the setting up of condition-based machine tool temperature monitoring and prevents any impact on the quality of the end product. Furthermore, the I4.0 KG can be used by deep learning models to carry out entity matching, node classifications, link prediction, and knowledge graph completion [72, 147, 167].

5.5 SUMMARY

This chapter highlights the adaptability of the Reference Generalized Ontological Model using a football manufacturing production line realistic data. The nine machines deployed in the production line are then explained, including the laser cutting machine, Oval Printing Machine, high-frequency cutting machine, glue spraying machine, heat activating conveyor machine, ball shaping machine, form moulding machine, ball seam glue machine, and heat drying conveyor machine. After this, the data definition and attributing, explain the approach to data integration using RGOM are then highlighted. The integration process is composed of four layers: Unstructured Data Sources, Knowledge Graph Construction, Knowledge Graph, and Users and Application Layer. The upcoming chapters discuss the impact of the SOTA embedding models on the knowledge graph developed from the football dataset using the RGOM.

6 A PERFORMANCE ANALYSIS OF EMBEDDING MODELS FOR LINK PREDICTIONS IN I40KG

This chapter presents a performance analysis of various embedding models on the constructed Knowledge Graph (KG) from the football dataset. It begins by highlighting the challenges associated with link prediction in I40KG and proceeds to provide an analysis of the I40KG dataset. Additionally, different knowledge graph embedding models, including TransE, DistMult, ComplEx, ConvKB, and ConvE, are then discussed. The experiments conducted, the experimental setup, and the obtained results are presented. Finally, the study's findings are discussed, underscoring the importance of selecting appropriate embedding models for effective link prediction in the I40KG.

6.1 THE PROBLEM OF LINK PREDICTION IN I40KG

In recent years, manufacturing industries have been moving towards adopting KG to utilise their data [105]. A growing amount of research is being conducted on building and implementing KGs for use in manufacturing production lines. A manufacturing production line KG represents relationships between various nodes, such as workstations and manufacturing machines, like the one shown in Figure 6.1. Nodes in the KG correspond to machines, manufacturing processes, materials, and their attributes, and edges connecting pairs of nodes represent some facts or labels that capture the relationship, such as "`WorkStation` *hasMachine* `Machine`". On this basis, a KG can be defined as a labelled directed graph G = (V_e, E, T), such that V_e, and E are a set of nodes and labels representing entities and relations, and T represents the triples accordingly [187].

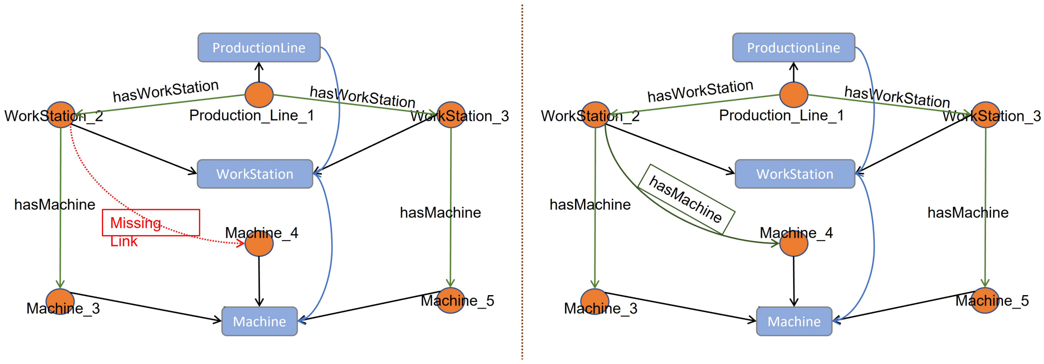


Figure 6.1: The figure depicts a snippet of the football manufacturing production line KG. The green arrow represents the semantics-based connectivity between instances, while the blue arrow represents the connectivity between ontology classes. The black arrows connect the instances to classes.

6.2 I40KG: THE DATA ANALYSIS

This section briefly explores the football production line KG and its dataset. This study investigates the possibility of using a football manufacturing KG to predict new facts. In the process of KG-based data integration, data from various sources is integrated and harmonised into manufacturing settings. The example scenario in Figure 5.1 demonstrates the elements necessary for semantic data integration to create a KG. Although the integration of data offers significant benefits, the KG also provides a means of discovering new relationships or links between data. These links can be created automatically based on the semantics encoded in the KG, allowing missing information to be completed or restored.

The I40KG has a hierarchical structure where the nodes at the top are loosely connected. In state-of-the-art datasets, the nodes are somehow tightly connected due to the node types. The size of the football manufacturing production line KG is comprised of a total of 180701 entities and 35 relations that make 386905 triples. Among these, 9955 nodes are pendants that represent entities that are not highly connected in the KG. The KG has a density [186] of 2.369×10^{-5} which represents the connectivity of the KG and is calculated using Equation 6.1.

$$d = \frac{m}{n^2} \tag{6.1}$$

Where d represents the density, m is the number of edges, and n is the number of nodes. This metric indicates the overall connectivity of the KG. Furthermore, the KG has a mean degree centrality of 2.20×10^{-5}. The degree of centrality of a node is defined as the number of edges it has in the graph, normalised by the maximum possible number of edges. In addition to the KG's degree of centrality, the node *Heat_conveyor_operation* has a maximum degree of centrality, which is

5.53×10^{-6}. This indicates that even the most connected node in the graph has a relatively low number of connections, again in comparison to the maximum possible. Moreover, the KG has a mean network eigenvector centrality of 0.143. This number indicates the average influence of a node in the graph. Unlike degree centrality, eigenvector centrality considers the significance of the nodes to which a node is connected. A mean network eigenvector centrality of 0.143 indicates that, on average, a node in the network holds a certain level of influence.

6.3 KNOWLEDGE GRAPH EMBEDDING MODELS

This section explains the embedding models. To predict the missing links, we choose the five well-known models, that is, ComplEx, DistMult, TransE, ConvKB, and ConvE. About the aforementioned embedding models, we first start with the working process of the TransE model followed by the rest.

6.3.1 TransE

TransE is one of the most popular state-of-the-art embedding models. The training set S is made up of triplets (e_1, r, e_2), where $e_1, e_2 \in E$ (the set of entities) and $r \in L$ (the set of relationships). TransE learns how to embed entities and relationships into these triplets.These embeddings belong to $\mathbb{R}^k$ (k is a model hyperparameter) and are represented by boldface letters. The main concept of TransE is that the functional relation induced by the r-labeled edges corresponds to a translation of the embeddings, that is, it desires $\mathbf{e_1} + \mathbf{r} \approx \mathbf{e_2}$ when $(\mathbf{e_1}, \mathbf{r}, \mathbf{e_2})$ holds ($\mathbf{e_2}$ should be the nearest neighbour of $\mathbf{e_1} + \mathbf{r}$), and $\mathbf{e_1} + \mathbf{r}$ should be distant from $\mathbf{e_2}$, appropriately. Using an energy-based framework, the energy of a triplet equals $d(\mathbf{e_1} + \mathbf{r}, \mathbf{e_2})$ for a dissimilarity measure d, which is chosen by TransE to be either the L1-norm or L2-norm. To obtain embeddings, TransE utilises a margin-based ranking criterion that is minimised over the training set, as given in Equation 6.2.

$$L = \sum_{(e_1,r,e_2)\in S} \sum_{\left(e_1',r,e_2'\right)\in S'(e_1,r,e_2)} \left[\gamma + d\left(\mathbf{e_1} + \mathbf{r}, \mathbf{e_2}\right) - d\left(\mathbf{e_1'} + \mathbf{r}, \mathbf{e_2'}\right)\right]_+ \tag{6.2}$$

where $[x]_+$ represents the positive features of x, $\gamma > 0$ is a margin hyperparameter.

$$S'(h,r,t) = \left\{\left(h',r,t\right) | h' \in E\right\} \cup \left\{\left(h,r,t'\right) | t' \in E\right\} \tag{6.3}$$

To construct a set of corrupted triplets according to Equation 6.3, each training triplet is modified by replacing either the head or tail entity with a random entity, but not both at the same time. This strategy is effective because the loss function Equation 6.2 is designed to assign lower energy values to training triplets compared to corrupted triplets. By doing so, the loss function encourages the model to learn the embeddings that satisfy the intended criterion, and this happens naturally during the training process. It is worth noting that the embedding vector for a given entity is the same whether the entity appears as the head or the tail of a triplet. The optimisation process is conducted using stochastic gradient descent in minibatch mode over the possible e_1, r, and e_2 values.

6.3.2 DistMult

DistMult aims to learn the representations of entities and relations in a KG so that valid triplets receive high scores. Given a KG that is represented as a list of relation triplets (e1, r, e2) denoting a relationship r between entities e1 and e2. In order to learn the embeddings, a two-layer neural network is used. The first layer projects the input entities to low-dimensional vectors, and the second layer combines these vectors using a scoring function with relation-specific parameters to produce a scalar for comparison.

In relation to embedding learning, DistMult associates each input entity with a high-dimensional vector that can be either a "one-hot" index vector or an "n-hot" feature vector. The input vectors for entity e_1 and e_2 are denoted as x_{e_1} and x_{e_2}, respectively. Additionally, the first layer projection matrix is denoted by $\mathbf{W}$.

After passing the input vectors through the neural network, the model learned entity representations y_{e_1} and y_{e_2}. These representations can be expressed through Equation 6.4.

$$y_{e_1} = f(\mathbf{W}x_{e_1}), \quad y_{e_2} = f(\mathbf{W}x_{e_2}) \tag{6.4}$$

where f is a function that can be either linear or non-linear and is applied element-wise to the result of the matrix multiplication between $\mathbf{W}$ and x_{e_1} or x_{e_2}.

Furthermore, DistMult utilises a basic bi-linear scoring function Equation 6.5.

$$g_r^b(y_{e_1}, y_{e_2}) = y_{e1}^T M_r y_{e_2} \tag{6.5}$$

DistMult's scoring function is a modified version of the Neural Tensor Network (NTN) scoring function. The NTN scoring function typically involves a non-linear layer and a linear operator. However, DistMult differs from NTN by removing the aforementioned components and utilizing a 2-dimensional matrix operator $M_r \in \mathbb{R}^{n \times n}$ instead of a tensor operator. Moreover, other matrix factorization models have also utilised the bilinear formulation of DistMult's scoring function, along with various forms of regularisation. To simplify the model and reduce the number of relation parameters, DistMult imposes a constraint on M_r such that it must be a diagonal matrix. This straightforward approach has been shown to be both simple and effective.

6.3.3 ComplEx

Let R and E denote the sets of relations and entities present in a KG. The ComplEx model aims to recover the matrices of scores X_r for all relations $r \in R$. Given two entities e_1 and $e_2 \in E$, the log-odds of the probability that the fact $r(e_1, e_2)$ is true can be expressed in Equation 6.6.

$$P(Y_{r,e_1,e_2} = 1) = \sigma(\varphi(r, e_1, e_2; \Theta)) \tag{6.6}$$

where φ is a scoring function and is based on observed relations factorization, and Θ represents the corresponding model's parameters. Although the entire X matrix is unknown, it is assumed that there exists a set of partially observed adjacency matrices for different relations, denoted as $\{Y_{re_1e_2}\}_{r(e_1,e_2)\in\Omega} \in \{-1, 1\}$. These matrices consist of true and false facts for the observed triples in the KG, where $\Omega \subseteq R \times E \times E$ is the set of observed triples. The objective is to determine the likelihood of whether entries Y_{r',e'_1,e'_2} are true or false, where the triples $r'(e'_1, e'_2)$ are targeted and unobserved, i.e., $r'(e'_1, e'_2) \notin \Omega$.

The scoring function adopted by the ComplEx model is given in Equation 6.7.

$$\sigma(\varphi(r, e_1, e_2; \Theta)) = Re(\langle w_r, e_1, e_2 \rangle) \tag{6.7}$$

where $w_r \in \mathbb{C}^K$ and represents a complex vector. The function $Re(\langle w_r, e_1, e_2 \rangle)$ in Equation 6.7 represents the real part of the complex dot product between the relation r embedding and the embeddings of entities e_1 and e_2.

6.3.4 ConvKB

ConvKB represents the dimensionality of embeddings as k, such that each embedding triple ($\mathbf{ve_1}$, $\mathbf{vr}$, $\mathbf{ve_2}$) is seen as a matrix $\mathbf{A}_i \in \mathbb{R}^{k\times 3}$, with $\mathbf{A}_i \in \mathbb{R}^{1\times 3}$ indicating the i-th row of $\mathbf{A}$. And utilise a filter $\boldsymbol{\omega} \in \mathbb{R}^{1\times 3}$ within the convolution layer. The purpose of $\boldsymbol{\omega}$ is not only to investigate the global relationships between identical dimensional entries of the embedding triple ($\mathbf{ve_1}$, $\mathbf{vr}$, $\mathbf{ve_2}$), but also to capture the transitional features in transition-based models. We repeatedly apply $\boldsymbol{\omega}$ over each row of $\mathbf{A}$ to ultimately produce a feature map $\mathbf{v} = [v_1, v_2, \dots, v_k] \in \mathbb{R}^k$ is given in Equation 6.8.

$$v_i = g(\boldsymbol{\omega} \cdot \mathbf{A}_i + b) \tag{6.8}$$

where $b \in \mathbb{R}$ represents a bias term, and g denotes an activation function, for instance, the Rectified Linear Unit (ReLU).

ConvKB employs different filters $\in \mathbb{R}^{1\times 3}$ to produce distinct feature maps. Denote the collection of filters as Ω and the total number of filters as τ, such that $\tau = |\Omega|$. This leads to the generation of τ feature maps. These τ feature maps are then merged into a single vector in $\mathbb{R}^{\tau k \times 1}$, which is subsequently calculated with a weight vector $\mathbf{w} \in \mathbb{R}^{\tau k \times 1}$ through a dot product, yielding a score for the triple (e_1, r, e_2). Equation 6.9 presents the scoring function that has been adopted by ConvKB.

$$f(e_1, r, e_2) = concat(g([v_{e_1}, v_r, v_{e_2}] * \Omega)) \cdot w \tag{6.9}$$

where Ω and w represent shared parameters that are not dependent on e_1, r, or e_2; the symbol $*$ signifies a convolution operator; and the term 'concat' denotes a concatenation operator. The ConvKB model training loss is minimised via using Adam optimiser with L2 regularization on the weight vector was shown in Equation 6.10.

$$L = \sum_{(e_1,r,e_2)\in\{G\cup G'\}} \log(1 + \exp(l(e_1, r, e_2) \cdot f(e_1, r, e_2))) + \frac{\lambda}{2}\|w\|_2^2 \tag{6.10}$$

where $l(e_1, r, e_2)$ is a function that assigns labels to triples, and G' represents a set of invalid triples created by altering valid triples found in G.

6.3.5 ConvE

ConvE utilises a neural link prediction model that leverages convolutional and fully-connected layers to model the interactions between input entities and the relationships. The key feature of the ConvE model is that the score is established through a convolution performed over embeddings shaped in 2D. ConvE defines the scoring function as follows.

$$\psi_r(e_1, e_2) = f(\mathrm{vec}(f([\overline{e_1}; \overline{r_r}] * \omega)) \cdot W) \cdot e_2 \tag{6.11}$$

where the relation parameter, $r_r \in \mathbb{R}_k$, in the Equation 6.11 depends on r. Additionally, e_1 and r_r are subject to 2D reshaping, denoted as $\overline{e_1}$ and $\overline{r_r}$ respectively.

Specifically, if both e_1 and r_r are elements of $\mathbb{R}^k$, then their reshaped forms $\overline{e_1}$ and $\overline{r_r} \in \mathbb{R}^{k_w \times k_h}$, where k is equal to $k_w \times k_h$.

During the feed-forward pass, the model conducts a row-vector lookup operation on two embedding matrices: one for entities, represented as $\mathbf{E}^{|\mathcal{E}| \times k}$, and another for relations, denoted as $\mathbf{R}^{|\mathcal{R}| \times k'}$. Here, k and k' are the dimensions of entity and relation embeddings respectively, and $|\mathcal{E}|$ and $|\mathcal{R}|$ represent the number of entities and relations respectively. The model concatenates $\overline{\mathbf{e_1}}$ and $\overline{\mathbf{r_r}}$ and uses the resulting vector as input to a 2D convolutional layer with filters ω. This layer produces a feature map tensor $\mathcal{T} \in \mathbb{R}^{c \times m \times n}$, where c is the number of 2D feature maps and m and n are their dimensions. The tensor $\mathcal{T}$ is then reshaped into a vector $\text{vec}(\mathcal{T}) \in \mathbb{R}^{cmn}$, which is subsequently projected into a k-dimensional space via a linear transformation that is parameterised by the matrix $\mathbf{W} \in \mathbb{R}^{cmn \times k}$. Finally, this projection is matched with the object embedding, e_o, through an inner product. It is important to note that the convolutional filter parameters and the matrix W parameters are independent of the parameters used for the entities e_1 and e_2, as well as the relationship r. Equation 6.12 represents the binary cross entropy function that is used to minimise the model loss.

$$\mathcal{L}(p, t) = -\frac{1}{N} \sum_{i=1} (t_i \cdot \log(p_i) + (1 - t_i) \cdot \log(1 - p_i)) \tag{6.12}$$

where t represents the label vector and p_i represents the predicted probability.

6.4 EXPERIMENTS AND RESULTS

6.4.1 Experimental setup

Dataset and Training

This section describes the dataset and the training procedure. 70% of the data is used for training, and 30% is used for testing. Section 4.2 summarises the dataset used in this research. The hypermeters are chosen by trying different values and observing their impact on model performance. Additionally, a learning rate of 0.0001 and latent feature dimensions k of 200 are chosen to train the state-of-the-art model. We set the number of negative triplets to five during training for each positive triplet. With a batch count of 100, the models are trained over 50 epochs. The loss function is minimised using the Adam algorithm.

The missing links are generated by creating corrupted triples, where either the head or tail of a valid triple is replaced with a random entity, but not both at the same time. During evaluation, for each test triple, the model computes and ranks the dissimilarities of these corrupted triples after replacing the head and tail with each entity from the dictionary, in turn, to determine the rank of the correct entity. The performance is then measured using metrics like mean rank and Hits@N, which reflect the proportion of correct entities ranked in the top 10, 3 and 1 predictions. Moreover, the ranking method involves evaluating test triples against all other candidate triples not present in the training, validation, or test sets. This is achieved by substituting either the subject or the object of a test triple with every entity in the knowledge graph, thereby generating candidate triples.

Evaluation Metrics

Here, we discuss the evaluation metrics employed to evaluate the accuracy of the rankings generated by these models. We use two main assessment metrics: Mean Reciprocal Rank (MRR) and Hits@N.

Mean Reciprocal Rank (MRR) calculates the average of the reciprocal ranks of the true (or correct) triplets. The reciprocal rank is the multiplicative inverse of the rank (that is, 1/rank) Equation 6.13.

$$\mathrm{MRR} = \frac{1}{N}\sum_{i=1}^{N}\frac{1}{R_i} \tag{6.13}$$

MRR is sensitive to how well the model ranks the highest-ranked true triplet, and a higher MRR indicates better performance. MRR ranges from 0 to 1, with 1 being the best possible score.

Hits@N Equation 6.14 computes the percentage of true triplets that appear within the top N positions in the ranked list. We have used Hits@1, Hits@3 and Hits@10 for the model evaluation. A higher Hits@N value indicates better performance, as it means a larger proportion of true triplets are ranked within the top N positions.

$$\mathrm{Hits} \; @ \; N = \frac{1}{Q}\sum_{i}^{Q}\delta(\mathrm{rank}_i \leqslant N) \tag{6.14}$$

where Q is the count of positive and negative triples, rank_i is the rank of the positive triples within these triples, and δ is an indicator function that is 1 if $\mathrm{rank}_i \leqslant N$, and 0 otherwise.

By comparing these metrics across different models, we can determine which model performs better in ranking true triplets.

Table 6.1: Comparative evaluation of KG embedding models ComplEx, DistMult, TransE, ConvKB, and ConvE across five test scenarios using Mean Reciprocal Rank (MRR), Hits@10, Hits@3, and Hits@1 as performance metrics.

Models	Test 1				Test 2				Test 3			
	MRR	Hits@10	Hits@3	Hits@1	MRR	Hits@10	Hits@3	Hits@1	MRR	Hits@10	Hits@3	Hits@1
ComplEx	0.271	0.331	0.288	0.236	0.275	0.338	0.301	0.245	0.273	0.335	0.293	0.237
DistMult	0.255	0.311	0.274	0.221	0.252	0.307	0.271	0.218	0.258	0.312	0.276	0.225
TransE	**0.289**	**0.348**	**0.322**	**0.249**	**0.292**	**0.354**	**0.320**	**0.253**	**0.291**	**0.352**	**0.325**	**0.250**
ConvKB	0.240	0.321	0.268	0.193	0.238	0.322	0.266	0.191	0.227	0.320	0.265	0.170
ConvE	0.195	0.281	0.195	0.165	0.191	0.279	0.189	0.162	0.201	0.296	0.197	0.172
	Test 4				Test 5							
	MRR	Hits@10	Hits@3	Hits@1	MRR	Hits@10	Hits@3	Hits@1				
ComplEx	0.272	0.328	0.288	0.238	0.277	0.337	0.296	0.242				
DistMult	0.254	0.311	0.272	0.220	0.243	0.308	0.266	0.203				
TransE	**0.295**	**0.350**	**0.323**	**0.258**	**0.300**	**0.354**	**0.325**	**0.264**				
ConvKB	0.238	0.322	0.271	0.188	0.239	0.323	0.270	0.191				
ConvE	0.193	0.279	0.193	0.165	0.187	0.273	0.185	0.159				

6.4.2 Results

Overall Results

Here, we discuss the overall results of the models on the unseen data. The effectiveness of several KG embedding models, including ComplEx, DistMult, TransE, ConvKB, and ConvE, is thoroughly assessed. We carried out the experiments five times (named as five tests) and evaluated the models using the Mean Reciprocal Rank (MRR), Hits@10, Hits@3, and Hits@1 metrics shown in Table 6.1. It is observed from the overall results that the TransE model outperforms the other models for all test scenarios for football manufacturing production data in terms of MRR, Hits@10, Hits@3, and Hits@1. Additionally, ConvKB showed competitive outcomes, but none of the evaluation metrics saw it outperform TransE.

Model Prediction Results

We now discuss the prediction performance of state-of-the-art embedding models, trained on manufacturing football KG. These models are used to predict the relationships between entities based on the known triples in the KG.

The prediction results (Figure 6.2) on unseen test data from the football manufacturing KG show that the models have achieved varying average accuracy levels between 0 and 1. TransE outperforms the other models with an average accuracy of 0.91, closely followed by ComplEx at 0.87 and DistMult at 0.84. The ConvKB and ConvE models have lower accuracies, with 0.79 and 0.76, respectively. The better performance of the TransE is due to its strategy of modelling the relationships as translations in the entity embedding space. This approach works well for hierarchical data, as entities in a hierarchical structure often have simple and direct relationships. On the other hand, ConvKB and ConvE are based on convolutional neural networks (CNNs), which are better suited for capturing complex and non-

Table 6.2: Performance comparison of KG embedding models on example triples from unseen data.

Triple	ComplEx	DistMult	TransE	ConKB	ConvE
WorkStation_2 hasMachine Machine_4	0.87	0.84	0.92	0.81	0.78
Machine2_Motor_State_177 hasState working	0.89	0.89	0.90	0.83	0.81
Squeegee3_Pressure_Sensor madeObservation Observation_180	0.86	0.80	0.92	0.75	0.73
Machine1_motor1 hasSpeed Machine1_motor_Speed_232	0.85	0.82	0.91	0.79	0.76
Oval_Printning_Process_3 useTool Machine2_Bed3	0.88	0.85	0.90	0.77	0.72

linear patterns in the data. As the manufacturing production line KG has a simple hierarchical structure, the convolutional layers in ConvKB and ConvE could not provide significant results in this case. Table 6.2 shows the accuracy achieved by five trained models, ComplEx, DistMult, TransE, ConvKB, and ConvE, for example, triples of unseen data.

Statistical Results

Significance tests, such as the t-test, are fundamental tools in statistics used to determine whether the differences observed between groups or models in an experiment are statistically significant or merely due to random chance. The p-values indicate the level of significance [145]. We performed pairwise t-tests between the MRR values of all potential model pairs to statistically evaluate the performance of these models. To evaluate the importance of the variations in MRR values between the models, the resulting p-values were computed from Table 6.1. Higher p-values imply that the difference between the compared models is not statistically significant, whereas lower p-values show a statistically significant difference between the compared models. Figure 6.3 represents the mean MRR values of

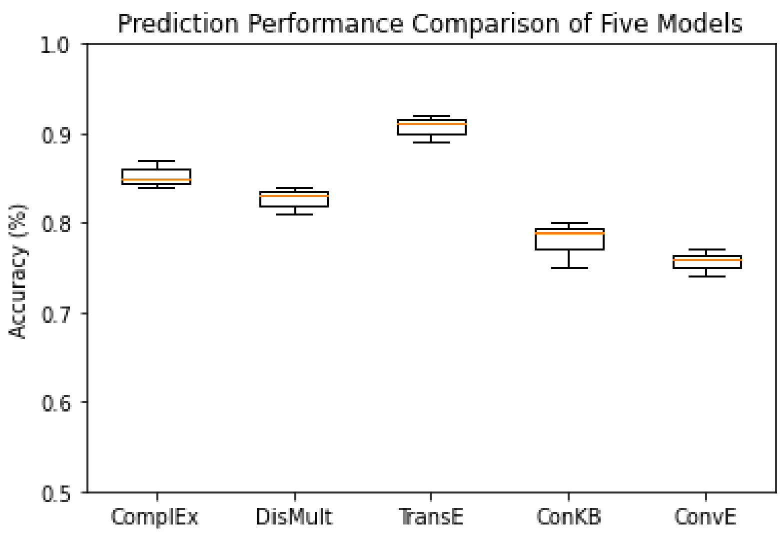


Figure 6.2: Performance comparison of five KG embedding models on unseen test data from the football manufacturing KG.

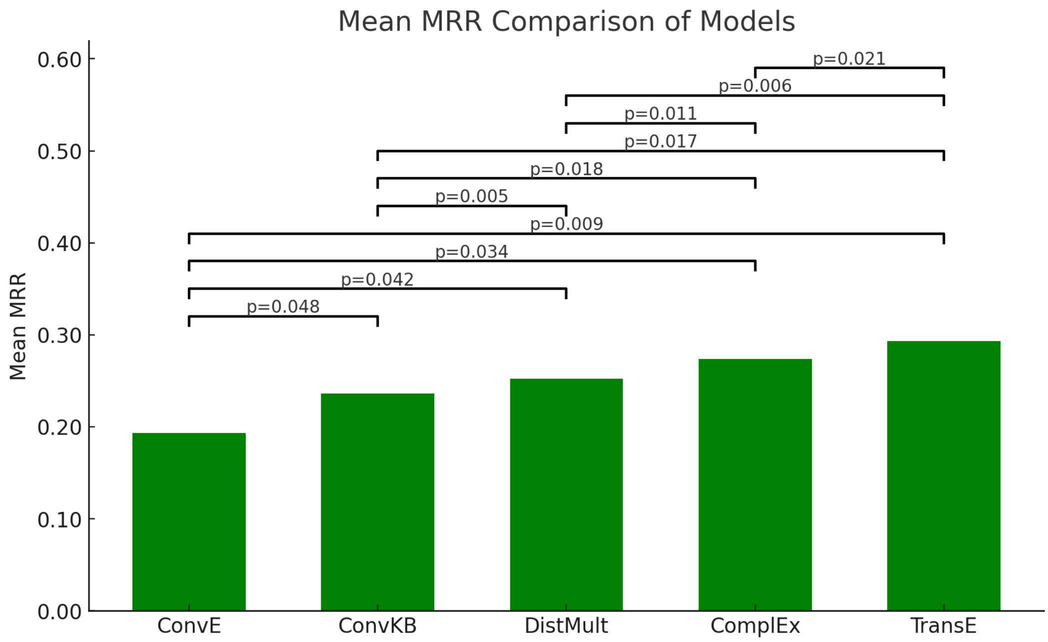
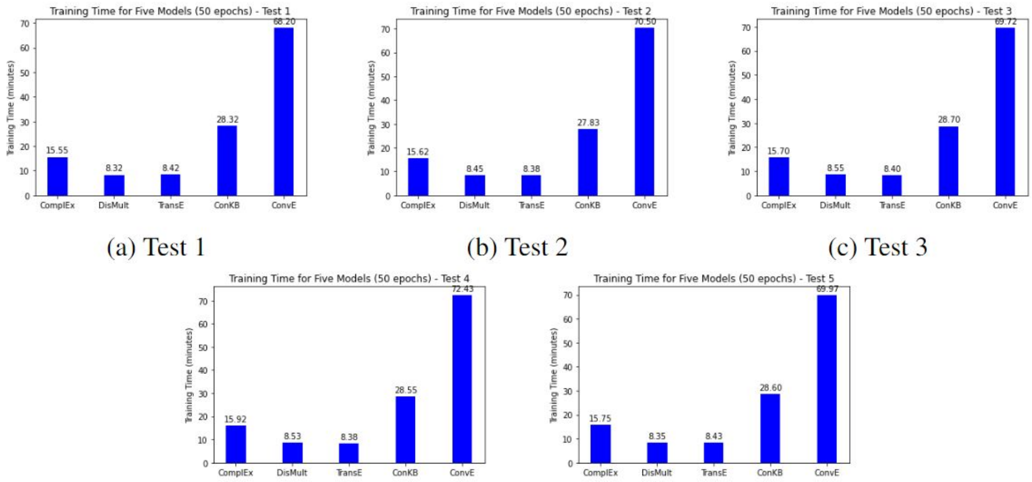


Figure 6.3: Comparison of Mean Reciprocal Rank (MRR) for link prediction models i.e., ComplEx, DistMult, TransE, ConvKB, and ConvE on a football manufacturing dataset. The p-values from pairwise t-tests are annotated above the bars.

the models, along with the pairwise p-values, highlighting the differences in their performance. The chart reveals that TransE outperforms the other models, while ConvE has the lowest mean MRR. Furthermore, the statistical analysis using t-tests shows significant differences between several model pairs, as indicated by the low p-values. Our research shows that the TransE model on the Hierarchical KGs such as the manufacturing football dataset performs best in terms of MRR.

Training Time Analysis of the Models

This section presents the time analysis of training the models on the football KG dataset. Our study has analysed the training times of five state-of-the-art KG embedding models (ComplEx, DistMult, TransE, ConvKB, and ConvE) for 50 epochs. The training times for each model have been recorded across five tests (see 6.4), and the results have been converted to minutes for easier comparison. The hardware used for experiments and implementation involved Nvidia GeForce GTX 1180 (8 GB of RAM) and Ubuntu 18.04.3 LTS (64-bit) operating system. We found that the DistMult and TransE models had the shortest training times, taking an average of 8 minutes and 26 seconds and 8 minutes and 25 seconds, respectively, to complete 50 training epochs. On the other hand, the ConvE model requires the longest training time, with an average of 70 minutes and 12 seconds, indicating



(a) Test 1 (b) Test 2 (c) Test 3

(d) Test 4 (e) Test 5

Figure 6.4: Training Time for Five Models (50 epochs) - Tests 1 to 5.

that it is the slowest model to train for the same number of epochs. The ComplEx and ConvKB models had intermediate training times.

The analysis also revealed some variation in training times between the different tests, particularly for the ConvE model, which has shown significant differences in training times between tests. Despite this variation, the DistMult and TransE models consistently demonstrated the fastest training times throughout all tests. Overall, the findings suggest that the DistMult and TransE models are the most efficient models in terms of training times for KG embedding, whereas the ConvE model is the slowest. These results could inform the choice of KG embedding models in different settings, particularly those where fast training times are crucial.

6.5 FINDINGS OF THE STUDY

We now discuss the findings and analysis in this section. Experts with diverse backgrounds are required to participate in the creation of a knowledge graph (KG), including those in ontology development and data source mapping, which may lead to errors that can directly affect the data's quality. Even with a KG-based integration, there may be disconnections among data silos that can negatively impact the data's completeness and accuracy. Connecting data entities typically necessitates manual intervention and expert knowledge to identify and create necessary connections. However, the approach introduced in this study enables industry experts to assist in identifying potential data links based on their domain expertise. Additionally, the KG's semantics can be utilised to describe and link data silos critical to manufacturing and other industries. Moreover, the method enables the discovery of patterns in the data, enhancing the process of linking different silos.

Table 5.1 presents the results for various KG embedding models, offering critical insights into their interaction within specific KG frameworks, especially in manufacturing production lines. The varied performance among the models, especially the notable efficacy of translation-based models such as TransE compared to neural network approaches like ConvE, highlights the crucial role of choosing the right model based on the KG's unique features, including its straightforward structure and sparse inter-entity connections. These observations emphasise the necessity for customised KG embedding strategies, indicating that models with a more straightforward, direct approach may outperform others in environments with less complex entity relationships. Moreover, these results require further exploration into refining KG embedding techniques, stressing the need to match model strengths with KG attributes to boost performance. This qualitative assessment not only highlights the existing constraints of current models but also paves the way for future research and the practical deployment of KG embeddings tailored to specific domains.

6.6 SUMMARY

In this chapter, the problem of link prediction in Industry 4.0 Knowledge Graphs (I40KG) has been thoroughly examined, and a performance analysis of various embedding models has been conducted. The analysis has involved the evaluation of TransE, DistMult, ComplEx, ConvKB, and ConvE models using the I40KG dataset. The experimental setup, including the selection of evaluation metrics and training parameters, has been explained in detail. The results obtained from the experiments have been presented, highlighting the performance of each model in terms of MRR and Hit@N measurement metrics. The study's findings have emphasised the superior performance of the TransE model, indicating the significance of considering transitional characteristics in manufacturing knowledge graphs. This chapter has provided valuable insights for researchers and practitioners working on link prediction in I40KG, contributing to the advancement of knowledge discovery and decision-making processes in smart manufacturing industries. In the next chapter, we will explain the ontology development process for RSWO in line with RGOM using a Bosch resistance spot welding use case.

7 CONCLUSION AND FUTURE WORK

7.1 CONCLUSION OF THE THESIS

This thesis addresses the challenges of integrating heterogeneous and unstructured data in the context of Industry 4.0 (I4.0) by leveraging semantic web and knowledge graph technologies that can be efficiently used by intelligent smart manufacturing applications. Through a comprehensive investigation, the following key contributions and findings have been achieved:

To answer the first research question, we conducted a thorough literature review analysing the latest developments in production line manufacturing semantic models within the context of Industry 4.0. This analysis resulted in identifying the gaps and areas lacking clarity in the current semantic models, particularly regarding the integration of unstructured data. In response, we proposed employing semantic web technologies as a solution to these identified challenges. A key element of our research is the introduction of the Reference Generalized Ontological Model (RGOM), which stands as a notable contribution. The RGOM uses concepts from existing ontologies while introducing new concepts previously overlooked. It leverages the hierarchical axis knowledge from the RAMI 4.0 reference architecture to illustrate the organisational structure. This incorporation results in a structured ontological framework that accurately represents manufacturing resources, machines, processes, and products. The establishment of RGOM significantly enhances data availability and interoperability within the manufacturing industry. Consequently, this research successfully addresses the **RQ1** by overcoming the **Challenge 1**, fulfilling our first contribution.

Alongside the RGOM, this research introduced a domain-specific ontology customised for Resistance Spot Welding (RSW), called the Resistance Spot Welding Ontology (RSWO). This ontology carefully captures important domain-specific knowledge related to welding processes, machinery, software, and electrode components. The knowledge is gathered from industry experts and ISO documents for RSW. Furthermore, to ensure the ontology's quality and its usability, it went through rigorous evaluation using established metrics, including the FAIR principles and OntoMetrics for ontology assessment. By successfully developing and aligning the RSWO with RGOM and confirming its adherence to quality and usability standards, this work has fulfilled **Contribution 2**. Thus, this thesis successfully addresses **(RQ2)** as **Challenge 2**.

To answer the **RQ3**, we demonstrated the adaptability and practicality of RGOM. A realistic dataset from a football manufacturing production line has been utilised. The first instance of the dataset has been directly acquired from the production line engineers, and the remaining instances have been randomly produced. The production line engineers validate the dataset to ensure its reliability. Through the application of RGOM, unstructured data sources have been integrated into the I4.0 Knowledge Graph (I40KG), enabling a holistic view of the production line. This integration has improved data availability and interoperability, allowing for more insights for users. Hence, this thesis fulfilled **Contribution 3** as **Challenge 3**.

Finally, a comparative analysis of various knowledge graph embedding models, including TransE, DistMult, ComplEx, ConvKB, and ConvE is conducted, focusing on their utility for link prediction within Industry 4.0 Knowledge Graphs (I40KG). This analysis was carried out using a dataset from a football manufacturing production line. Each model is trained and tested on the I40KG and evaluated using MRR and Hit@N metrics which are widely used for link prediction. By assessing the capabilities of each embedding model, the research highlights the important factors that influence their usage in the context of manufacturing KGs. Consequently, this analysis has addressed **RQ4**, leading to **Contribution 4**.

To conclude, this research has contributed to the practical application of semantic web and knowledge graph technologies in the manufacturing domain. The proposed solutions, including RGOM, the RSWO, and the performance analysis of KG embedding models, together have enhanced data integration, modelling, and knowledge extraction in the context of Industry 4.0.

7.2 FUTURE WORK

This section highlights several key research directions for the future work.

7.2.1 Harmonization with Top level ontologies

As a future work, the RGOM can be aligned to top-level ontology such as BFO [9] or DOLCE [19]. By aligning it to top-level ontology it can benefit in standardization that can facilitate communication and interoperability between different domains. The alignment with top-level ontologies can ensure consistency between them that will alternatively reduce ambiguity and confusion.

7.2.2 Extension to Other Production Lines

Furthermore, the approach of the concepts introduced in RGOM can be utilised for the majority of production lines which will facilitate wider use of our approach for generating KGs for different use cases. For example, this approach can be extended not only to other similar manufacturing processes such as volleyball and rugby ball production, but also to other more generic production lines. This can help other industries map their customised data into RGOM and construct an industry-specific KG.

7.2.3 Semantic Mapping of RSWO Terms

Additionally, the creation of semantic mapping of the RSWO terms to other similar welding techniques such as flash welding, and projection welding can be investigated as future work. Moreover, the embedding model features expressively, i.e., capture different relations (transitivity, symmetry, etc.) that can be analysed as future direction.

7.2.4 Exploration of Diverse Research Topics

In general, there is a wide range of topics to be explored: (1) *Commons for Industry* that includes a series of commonly agreed artefacts, methodologies, and best practices, such as a general and standardized ontology for all manufacturing processes, frameworks for sharing data, procedures of cross-domain innovation and strategy negotiation; (2) *Industrial Feedback for Facilitating Research*, the increased research, development and deployment of ontologies and knowledge graphs of industry problems will, in turn, inspire many impactful and challenging research questions that boost research and its interaction with industry, ranging from semantics-based data interoperability, metadata-based or content-based dataset search to neuro-symbolic reasoning for graph data.

BIBLIOGRAPHY

[1] B. Abu-Salih. "Domain-specific knowledge graphs: A survey". In: *Journal of Network and Computer Applications* 185 (2021), p. 103076.

[2] S. Ae Chun and B. MacKellar. "Social health data integration using semantic Web". In: *Proceedings of the 27th annual ACM symposium on applied computing*. 2012, pp. 392–397.

[3] Z. Ahmed and D. Gerhard. "Role of Ontology in Semantic Web Development". In: *First International Workshop on Cultural Heritage on the Semantic Web in conjunction with the 6th International Semantic Web Conference and the 2nd Asian Semantic Web Conference 2007*. 2007, p. 2.

[4] Y. Alfaifi. "Ontology development methodology: a systematic review and case study". In: *2022 2nd International Conference on Computing and Information Technology (ICCIT)*. IEEE. 2022, pp. 446–450.

[5] E. Amdouni, S. Bouazzouni, and C. Jonquet. "O'FAIRe: Ontology FAIRness Evaluator in the AgroPortal Semantic Resource Repository". In: *The Semantic Web: ESWC 2022 Satellite Events*. Ed. by P. Groth, A. Rula, J. Schneider, I. Tiddi, E. Simperl, P. Alexopoulos, R. Hoekstra, M. Alam, A. Dimou, and M. Tamper. Cham: Springer International Publishing, 2022, pp. 89–94. ISBN: 978-3-031-11609-4.

[6] E. Amdouni, S. Bouazzouni, and C. Jonquet. "O'FAIRe: Ontology FAIRness Evaluator in the AgroPortal semantic resource repository". In: *European Semantic Web Conference*. Springer. 2022, pp. 89–94.

[7] R. Anderl. "Industrie 4.0–technological approaches, use cases, and implementation". In: *at-Automatisierungstechnik* 63.10 (2015), pp. 753–765.

[8] M. Arenas, C. Gutierrez, and J. Pérez. "Foundations of RDF Databases". In: *Reasoning Web. Semantic Technologies for Information Systems: 5th International Summer School 2009, Brixen-Bressanone, Italy, August 30 - September 4, 2009, Tutorial Lectures*. Ed. by S. Tessaris, E. Franconi, T. Eiter, C. Gutierrez, S. Handschuh, M.-C. Rousset, and R. A. Schmidt. Berlin, Heidelberg: Springer Berlin Heidelberg, 2009, pp. 158–204.

ISBN: 978-3-642-03754-2. DOI: 10.1007/978-3-642-03754-2_4. URL: https://doi.org/10.1007/978-3-642-03754-2_4.

[9] R. Arp, B. Smith, and A. D. Spear. *Building ontologies with basic formal ontology*. Mit Press, 2015.

[10] V. Babich and G. Hilary. “OM Forum—Distributed ledgers and operations: What operations management researchers should know about blockchain technology”. In: *Manufacturing & Service Operations Management* 22.2 (2020), pp. 223–240.

[11] J. Bakkas and M. Bahaj. “Generating of RDF graph from a relational database using Jena API”. In: *International Journal of Engineering and Technology* 5.2 (2013), pp. 1970–1975.

[12] Z. Balogh, E. Gatial, J. Barbosa, P. Leitão, and T. Matejka. “Reference architecture for a collaborative predictive platform for smart maintenance in manufacturing”. In: *2018 IEEE 22nd International Conference on Intelligent Engineering Systems (ines)*. IEEE. 2018, pp. 000299–000304.

[13] R. Bejarano, B. R. Ferrer, W. M. Mohammed, and J. L. M. Lastra. “Implementing a human-robot collaborative assembly workstation”. In: *2019 IEEE 17th International Conference on Industrial Informatics (INDIN)*. Vol. 1. IEEE. 2019, pp. 557–564.

[14] C. Bezerra, F. Freitas, and F. Santana. “Evaluating ontologies with competency questions”. In: *2013 IEEE/WIC/ACM International Joint Conferences on Web Intelligence (WI) and Intelligent Agent Technologies (IAT)*. Vol. 3. IEEE. 2013, pp. 284–285.

[15] C. Bizer and A. Schultz. “The R2R Framework: Publishing and Discovering Mappings on the Web.” In: *COLD* 665 (2010), pp. 97–108.

[16] A. Bordes, S. Chopra, and J. Weston. “Question answering with subgraph embeddings”. In: *2014 Conference on Empirical Methods in Natural Language Processing, EMNLP 2014*. Association for Computational Linguistics (ACL). 2014, pp. 615–620.

[17] A. Bordes, N. Usunier, A. Garcia-Duran, J. Weston, and O. Yakhnenko. “Translating embeddings for modeling multi-relational data”. In: *Advances in Neural Information Processing Systems* 26 (2013).

[18] S. Borgo and P. Leitão. “Foundations for a core ontology of manufacturing”. In: *Ontologies: a handbook of principles, concepts and applications in information systems* (2007), pp. 751–775.

[19] E. Bottazzi and R. Ferrario. “Preliminaries to a DOLCE ontology of organisations”. In: *International Journal of Business Process Integration and Management* 4.4 (2009), pp. 225–238.

[20] A. Bousdekis, K. Lepenioti, D. Ntalaperas, D. Vergeti, D. Apostolou, and V. Boursinos. “A RAMI 4.0 view of predictive maintenance: software architecture, platform and case study in steel industry”. In: *Advanced Information Systems Engineering Workshops: CAiSE 2019 International Workshops, Rome, Italy, June 3-7, 2019, Proceedings 31*. Springer. 2019, pp. 95–106.

[21] N. Boysen, D. Briskorn, and S. Emde. “Parts-to-picker based order processing in a rack-moving mobile robots environment”. In: *European Journal of Operational Research* 262.2 (2017), pp. 550–562.

[22] J. Brank, M. Grobelnik, and D. Mladenic. “A survey of ontology evaluation techniques”. In: *Proceedings of the Conference on Data Mining and Data Warehouses (SiKDD 2005)*. Citeseer. 2005, pp. 166–170.

[23] P. Bricker. “Ontological commitment”. In: (2014).

[24] G. Buchgeher, D. Gabauer, J. Martinez-Gil, and L. Ehrlinger. “Knowledge graphs in manufacturing and production: A systematic literature review”. In: *IEEE Access* 9 (2021), pp. 55537–55554.

[25] Q. Cao, A. Samet, C. Zanni-Merk, F. d. B. de Beuvron, and C. Reich. “An ontology-based approach for failure classification in predictive maintenance using fuzzy C-means and SWRL rules”. In: *Procedia Computer Science* 159 (2019), pp. 630–639.

[26] B. Chen, J. Wan, L. Shu, P. Li, M. Mukherjee, and B. Yin. “Smart factory of industry 4.0: Key technologies, application case, and challenges”. In: *IEEE Access* 6 (2017), pp. 6505–6519.

[27] I. Y. Chen, M. Agrawal, S. Horng, and D. Sontag. “Robustly extracting medical knowledge from EHRs: a case study of learning a health knowledge graph”. In: *PACIFIC SYMPOSIUM ON BIOCOMPUTING 2020*. World Scientific. 2019, pp. 19–30.

[28] X. Chen, S. Jia, and Y. Xiang. “A review: Knowledge reasoning over knowledge graph”. In: *Expert Systems with Applications* 141 (2020), p. 112948.

[29] H. Cheng, L. Xue, P. Wang, P. Zeng, and H. Yu. "Ontology-based web service integration for flexible manufacturing systems". In: *2017 IEEE 15th International Conference on Industrial Informatics (INDIN)*. IEEE. 2017, pp. 351–356.

[30] H. Cheng, P. Zeng, L. Xue, Z. Shi, P. Wang, and H. Yu. "Manufacturing ontology development based on Industry 4.0 demonstration production line". In: *2016 Third International Conference on Trustworthy Systems and their Applications (TSA)*. IEEE. 2016, pp. 42–47.

[31] T.-M. Choi, S. Kumar, X. Yue, and H.-L. Chan. "Disruptive technologies and operations management in the Industry 4.0 era and beyond". In: *Production and Operations Management* 31.1 (2022), pp. 9–31.

[32] D. Clark, S. Hunt, and P. Malacaria. "Quantitative analysis of the leakage of confidential data". In: *Electronic Notes in Theoretical Computer Science* 59.3 (2002), pp. 238–251.

[33] D. Collarana, M. Galkin, I. Traverso-Ribón, M.-E. Vidal, C. Lange, and S. Auer. "MINTE: semantically integrating RDF graphs". In: *Proceedings of the 7th International Conference on Web Intelligence, Mining and Semantics*. 2017, pp. 1–11.

[34] M. Compton, P. Barnaghi, L. Bermudez, R. Garcia-Castro, O. Corcho, S. Cox, J. Graybeal, M. Hauswirth, C. Henson, A. Herzog, et al. "The SSN ontology of the W3C semantic sensor network incubator group". In: *Journal of Web Semantics* 17 (2012), pp. 25–32.

[35] I. F. Cruz, H. Xiao, et al. "The role of ontologies in data integration". In: *Engineering Intelligent Systems for Electrical Engineering and Communications* 13.4 (2005), p. 245.

[36] C. d'Amato, L. Mahon, P. Monnin, and G. Stamou. "Machine Learning and Knowledge Graphs: Existing Gaps and Future Research Challenges". In: *Transactions on Graph Data and Knowledge* 1.1 (2023), pp. 1–35.

[37] J. Davis, T. Edgar, R. Graybill, P. Korambath, B. Schott, D. Swink, J. Wang, and J. Wetzel. "Smart manufacturing". In: *Annual Review of Chemical and Biomolecular Engineering* 6 (2015), pp. 141–160.

[38] G. De Giacomo, D. Lembo, M. Lenzerini, A. Poggi, and R. Rosati. "Using ontologies for semantic data integration". In: *A Comprehensive Guide Through the Italian Database Research Over the Last 25 Years* (2018), pp. 187–202.

[39] L. De Propris and D. Bailey. "1 Disruptive Industry 4.0". In: *Industry 4.0 and regional transformations* (2020), p. 1.

[40] D. Dessı, F. Osborne, D. R. Recupero, D. Buscaldi, and E. Motta. "Generating knowledge graphs by employing natural language processing and machine learning techniques within the scholarly domain". In: *Future Generation Computer Systems* 116 (2021), pp. 253–264.

[41] T. Dettmers, P. Minervini, P. Stenetorp, and S. Riedel. "Convolutional 2D knowledge graph embeddings". In: *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence and Thirtieth Innovative Applications of Artificial Intelligence Conference and Eighth AAAI Symposium on Educational Advances in Artificial Intelligence*. New Orleans, Louisiana, USA: AAAI Press, 2018. ISBN: 978-1-57735-800-8.

[42] D. Diep, C. Alexakos, and T. Wagner. "An Ontology-based interoperability framework for distributed manufacturing control". In: *2007 IEEE Conference on Emerging Technologies and Factory Automation (EFTA 2007)*. IEEE. 2007, pp. 855–862.

[43] Z. Dong, S. Paul, K. Tassenberg, G. Melton, and H. Dong. "Transformation from human-readable documents and archives in arc welding domain to machine-interpretable data". In: *Computers in Industry* 128 (2021), p. 103439.

[44] T. Ebisu and R. Ichise. "TorusE: Knowledge Graph Embedding on a Lie Group". In: *Proceedings of the AAAI Conference on Artificial Intelligence* 32.1 (Apr. 2018). DOI: 10.1609/aaai.v32i1.11538. URL: https://ojs.aaai.org/index.php/AAAI/article/view/11538.

[45] M. Fernández-López, A. Gómez-Pérez, and N. Juristo. "METHONTOLOGY: From Ontological Art Towards Ontological Engineering". In: *AAAI Conference on Artificial Intelligence*. 1997.

[46] L. Ferreira, N. Lopes, J. Silva, G. D. Putnik, M. M. Cruz-Cunha, and P. S. Ávila. *Technological Developments in Industry 4.0 for Business Applications*. IGI Global, 2018.

[47] B. R. Ferrer, B. Ahmad, A. Lobov, D. A. Vera, J. L. M. Lastra, and R. Harrison. "An approach for knowledge-driven product, process and resource mappings for assembly automation". In: *2015 IEEE International Conference on Automation Science and Engineering (CASE)*. IEEE. 2015, pp. 1104–1109.

[48] B. R. Ferrer, W. M. Mohammed, A. Lobov, A. M. Galera, and J. L. M. Lastra. "Including human tasks as semantic resources in manufacturing ontology models". In: *IECON 2017-43rd Annual Conference of the IEEE Industrial Electronics Society*. IEEE. 2017, pp. 3466–3473.

[49] A. G. Frank, L. S. Dalenogare, and N. F. Ayala. "Industry 4.0 technologies: Implementation patterns in manufacturing companies". In: *International Journal of Production Economics* 210 (2019), pp. 15–26.

[50] N. Freire and S. de Valk. "Automated interpretability of linked data ontologies:: an evaluation within the cultural heritage domain". In: *2019 IEEE International Conference on Big Data (Big Data)*. IEEE. 2019, pp. 3072–3079.

[51] G. Gawriljuk, A. Harth, C. A. Knoblock, and P. Szekely. "A scalable approach to incrementally building knowledge graphs". In: *International Conference on Theory and Practice of Digital Libraries*. Springer. 2016, pp. 188–199.

[52] M. Ghobakhloo. "Determinants of information and digital technology implementation for smart manufacturing". In: *International Journal of Production Research* 58.8 (2020), pp. 2384–2405.

[53] F. Giustozzi, J. Saunier, and C. Zanni-Merk. "Context modeling for industry 4.0: an ontology-based proposal". In: *Procedia Computer Science* 126 (2018), pp. 675–684.

[54] S. A. Gómez and P. R. Fillottrani. "Ontology metrics and evolution in the GF framework for ontology-based data access". In: *Argentine Congress of Computer Science*. Springer. 2021, pp. 237–253.

[55] I. Grangel-González. "A knowledge graph based integration approach for industry 4.0". PhD thesis. Universitäts-und Landesbibliothek Bonn, 2019.

[56] I. Grangel-González, P. Baptista, L. Halilaj, S. Lohmann, M.-E. Vidal, C. Mader, and S. Auer. "The industry 4.0 standards landscape from a semantic integration perspective". In: *2017 22nd IEEE international conference on emerging technologies and factory automation (ETFA)*. IEEE. 2017, pp. 1–8.

[57] I. Grangel-González, L. Halilaj, S. Auer, S. Lohmann, C. Lange, and D. Collarana. "An RDF-based approach for implementing industry 4.0 components with Administration Shells". In: *2016 IEEE 21st International Conference on Emerging Technologies and Factory Automation (ETFA)*. IEEE. 2016, pp. 1–8.

[58] I. Grangel-González, L. Halilaj, G. Coskun, S. Auer, D. Collarana, and M. Hoffmeister. "Towards a semantic administrative shell for industry 4.0 components". In: *2016 IEEE Tenth International Conference on Semantic Computing (ICSC)*. IEEE. 2016, pp. 230–237.

[59] I. Grangel-González, F. Lösch, and 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)*. Vol. 1. IEEE. 2020, pp. 93–100.

[60] M. Grüninger. "Using the PSL ontology". In: *Handbook on Ontologies*. Springer, 2009, pp. 423–443.

[61] J. Gu, Z. Wang, J. Kuen, L. Ma, A. Shahroudy, B. Shuai, T. Liu, X. Wang, G. Wang, J. Cai, and T. Chen. "Recent advances in convolutional neural networks". In: *Pattern Recognition* 77 (2018), pp. 354–377. ISSN: 0031-3203. DOI: https://doi.org/10.1016/j.patcog.2017.10.013. URL: https://www.sciencedirect.com/science/article/pii/S0031320317304120.

[62] N. Guarino, D. Oberle, and S. Staab. "What Is an Ontology?" In: *Handbook on Ontologies*. Ed. by S. Staab and R. Studer. Berlin, Heidelberg: Springer Berlin Heidelberg, 2009, pp. 1–17. ISBN: 978-3-540-92673-3. DOI: 10.1007/978-3-540-92673-3_0. URL: https://doi.org/10.1007/978-3-540-92673-3_0.

[63] Z. Gun and J. Chen. "Novel Knowledge Graph-and Knowledge Reasoning-Based Classification Prototype for OBIA Using High Resolution Remote Sensing Imagery". In: *Remote Sensing* 15.2 (2023), p. 321.

[64] M. Hankel and B. Rexroth. "The reference architectural model industrie 4.0 (rami 4.0)". In: *ZVEI* 2.2 (2015), pp. 4–9.

[65] P. Hitzler, M. Krötzsch, B. Parsia, P. F. Patel-Schneider, S. Rudolph, et al. *OWL 2 web ontology language primer*. Tech. rep. 1. 2009, p. 123.

[66] M. Hodkiewicz, E. Low, and C. Woods. "Towards a Reference Ontology for Maintenance Work Management." In: *I-ESA Workshops*. 2020.

[67] A. Hogan, E. Blomqvist, M. Cochez, C. d'Amato, G. d. Melo, C. Gutierrez, S. Kirrane, J. E. L. Gayo, R. Navigli, S. Neumaier, et al. "Knowledge graphs". In: *ACM Computing Surveys (CSUR)* 54.4 (2021), pp. 1–37.

[68] R. Isele, A. Jentzsch, and C. Bizer. "Silk server-Adding missing links while consuming linked data". In: *CEUR Workshop Proceedings*. Vol. 665. RWTH Aachen. 2010, pp. 85–96.

[69] P. Jain, P. Hitzler, A. P. Sheth, K. Verma, and P. Z. Yeh. "Ontology alignment for linked open data". In: *International Semantic Web Conference*. Springer. 2010, pp. 402–417.

[70] E. Järvenpää, M. Lanz, and N. Siltala. "Formal resource and capability models supporting re-use of manufacturing resources". In: *Procedia Manufacturing* 19 (2018), pp. 87–94.

[71] E. Järvenpää, N. Siltala, O. Hylli, and M. Lanz. "The development of an ontology for describing the capabilities of manufacturing resources". In: *Journal of Intelligent Manufacturing* 30.2 (2019), pp. 959–978.

[72] S. Ji, S. Pan, E. Cambria, P. Marttinen, and S. Y. Philip. "A survey on knowledge graphs: Representation, acquisition, and applications". In: *IEEE Transactions on Neural Networks and Learning Systems* 33.2 (2021), pp. 494–514.

[73] X. Jiang, Q. Wang, and B. Wang. "Adaptive convolution for multi-relational learning". In: *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*. 2019, pp. 978–987.

[74] H. Jung, H. Yoo, and K. Chung. "Associative context mining for ontology-driven hidden knowledge discovery". In: *Cluster Computing* 19 (2016), pp. 2261–2271.

[75] C. Kaar, J. Frysak, C. Stary, U. Kannengiesser, and H. Müller. "Resilient ontology support facilitating multi-perspective process integration in industry 4.0". In: *Proceedings of the 10th International Conference on Subject-Oriented Business Process Management*. 2018, pp. 1–10.

[76] E. G. Kalaycı, I. Grangel González, F. Lösch, G. Xiao, A. ul-Mehdi, E. Kharlamov, and D. Calvanese. "Semantic integration of Bosch manufacturing data using virtual knowledge graphs". In: *The Semantic Web–ISWC 2020: 19th International Semantic Web Conference, Athens, Greece, November 2–6, 2020, Proceedings, Part II 19*. Springer. 2020, pp. 464–481.

[77] H. S. Kang, J. Y. Lee, S. Choi, H. Kim, J. H. Park, J. Y. Son, B. H. Kim, and S. D. Noh. "Smart manufacturing: Past research, present findings, and future directions". In: *International Journal of Precision Engineering and Manufacturing-Green Technology* 3 (2016), pp. 111–128.

[78] M. H. Karray, B. Chebel-Morello, and N. Zerhouni. "A formal ontology for industrial maintenance". In: *Applied Ontology* 7.3 (2012), pp. 269–310. DOI: 10.3233/AO-2012-0112.

[79] A. Keshavarzi and K. J. Kochut. "KGDiff: Tracking the evolution of knowledge graphs". In: *2020 IEEE 21st International Conference on Information Reuse and Integration for Data Science (IRI)*. IEEE. 2020, pp. 279–286.

[80] D. Kibira, K. C. Morris, and S. Kumaraguru. "Methods and tools for performance assurance of smart manufacturing systems". In: *Journal of Research of the National Institute of Standards and Technology* 121 (2016), p. 282.

[81] K.-Y. Kim and F. Ahmed. "Semantic weldability prediction with RSW quality dataset and knowledge construction". In: *Advanced Engineering Informatics* 38 (2018), pp. 41–53.

[82] L. Koh, G. Orzes, and F. J. Jia. "The fourth industrial revolution (Industry 4.0): technologies disruption on operations and supply chain management". In: *International Journal of Operations & Production Management* 39.6/7/8 (2019), pp. 817–828.

[83] E. E. Kosasih, F. Margaroli, S. Gelli, A. Aziz, N. Wildgoose, and A. Brintrup. "Towards knowledge graph reasoning for supply chain risk management using graph neural networks". In: *International Journal of Production Research* (2022), pp. 1–17.

[84] O. Kovalenko, I. Grangel-González, M. Sabou, A. Lüder, S. Biffl, S. Auer, and M.-E. Vidal. "AutomationML ontology: modeling cyber-physical systems for industry 4.0". In: *IOS Press Journal* 1 (2018).

[85] S. Kumar, M. Suhaib, and M. Asjad. "Industry 4.0: Complex, disruptive, but inevitable". In: *Management and Production Engineering Review* (2020).

[86] V. R. S. Kumar, A. Khamis, S. Fiorini, J. L. Carbonera, A. O. Alarcos, M. Habib, P. Goncalves, H. Li, and J. I. Olszewska. "Ontologies for industry 4.0". In: *The Knowledge Engineering Review* 34 (2019), e17.

[87] B. Lantow. "OntoMetrics: Putting Metrics into Use for Ontology Evaluation". In: *International Conference on Knowledge Engineering and Ontology Development*. 2016.

[88] H. Lasi, P. Fettke, H.-G. Kemper, T. Feld, and M. Hoffmann. "Industry 4.0". In: *Business Information Systems Engineering* 6.4 (2014), pp. 239–242. ISSN: 1867-0202. DOI: 10.1007/s12599-014-0334-4. URL: https://doi.org/10.1007/s12599-014-0334-4.

[89] S. Lemaignan, A. Siadat, J.-Y. Dantan, and A. Semenenko. "MASON: A proposal for an ontology of manufacturing domain". In: *IEEE Workshop on Distributed Intelligent Systems: Collective Intelligence and Its Applications (DIS'06)*. IEEE. 2006, pp. 195–200.

[90] G. Li, N. Li, and S. P. Sethi. "Does CSR reduce idiosyncratic risk? Roles of operational efficiency and AI innovation". In: *Production and Operations Management* 30.7 (2021), pp. 2027–2045.

[91] Q. Li, H. Jiang, Q. Tang, Y. Chen, J. Li, and J. Zhou. "Smart manufacturing standardization: reference model and standards framework". In: *On the Move to Meaningful Internet Systems: OTM 2016 Workshops: Confederated International Workshops: EI2N, FBM, ICSP, Meta4eS, and OTMA 2016, Rhodes, Greece, October 24–28, 2016, Revised Selected Papers*. Springer. 2017, pp. 16–25.

[92] Q. Li, Q. Tang, I. Chan, H. Wei, Y. Pu, H. Jiang, J. Li, and J. Zhou. "Smart manufacturing standardization: Architectures, reference models and standards framework". In: *Computers in Industry* 101 (2018), pp. 91–106.

[93] S. Li, L. D. Xu, and S. Zhao. "The internet of things: a survey". In: *Information Systems Frontiers* 17 (2015), pp. 243–259.

[94] S.-W. Lin, B. Miller, J. Durand, R. Joshi, P. Didier, A. Chigani, R. Torenbeek, D. Duggal, R. Martin, G. Bleakley, et al. "Industrial internet reference architecture". In: *Industrial Internet Consortium (IIC), Tech. Rep* (2015).

[95] Y. Lu, K. C. Morris, S. Frechette, et al. "Current standards landscape for smart manufacturing systems". In: *National Institute of Standards and Technology, NISTIR* 8107.3 (2016).

[96] Y. Lu, F. Riddick, and N. Ivezic. "The paradigm shift in smart manufacturing system architecture". In: *Advances in Production Management Systems. Initiatives for a Sustainable World: IFIP WG 5.7 International Conference, APMS 2016, Iguassu Falls, Brazil, September 3-7, 2016, Revised Selected Papers*. Springer. 2016, pp. 767–776.

[97] L. Lü and T. Zhou. "Link prediction in complex networks: A survey". In: *Physica A: statistical mechanics and its applications* 390.6 (2011), pp. 1150–1170.

[98] Y. Ma, P. A. Crook, R. Sarikaya, and E. Fosler-Lussier. "Knowledge graph inference for spoken dialog systems". In: *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE. 2015, pp. 5346–5350.

[99] Z. Ma, M. A. Capretz, and L. Yan. "Storing massive Resource Description Framework (RDF) data: a survey". In: *The Knowledge Engineering Review* 31.4 (2016), pp. 391–413.

[100] A. Maedche, B. Motik, L. Stojanovic, R. Studer, and R. Volz. "An infrastructure for searching, reusing and evolving distributed ontologies". In: *Proceedings of the 12th international conference on World Wide Web*. 2003, pp. 439–448.

[101] A. Magkanaraki, S. Alexaki, V. Christophides, and D. Plexousakis. "Benchmarking rdf schemas for the semantic web". In: *International Semantic Web Conference*. Springer. 2002, pp. 132–146.

[102] G. May and D. Kiritsis. "Zero defect manufacturing strategies and platform for smart factories of industry 4.0". In: *Proceedings of the 4th International Conference on the Industry 4.0 Model for Advanced Manufacturing: AMP 2019 4*. Springer. 2019, pp. 142–152.

[103] L. Mazzola, P. Kapahnke, M. Vujic, and M. Klusch. "CDM-Core: A Manufacturing Domain Ontology in OWL2 for Production and Maintenance". In: *Proceedings of the International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management*. 2016, pp. 136–143.

[104] A. Menon, N. B. Krdzavac, and M. Kraft. "From database to knowledge graph—using data in chemistry". In: *Current Opinion in Chemical Engineering* 26 (2019), pp. 33–37.

[105] L. Nagy, T. Ruppert, and J. Abonyi. "Ontology-Based Analysis of Manufacturing Processes: Lessons Learned from the Case Study of Wire Harness Production". In: *Complexity* 2021 (2021).

[106] D. Q. Nguyen, T. D. Nguyen, D. Q. Nguyen, and D. Phung. "A Novel Embedding Model for Knowledge Base Completion Based on Convolutional Neural Network". In: *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 2 (Short Papers)*. Ed. by M. Walker, H. Ji, and A. Stent. New Orleans, Louisiana: Association for Computational Linguistics, June 2018, pp. 327–333. DOI: 10.18653/v1/N18-2053.

[107] Y. Nishioka. "Industrial Value Chain Initiative for Smart Manufacturing". In: *Tokyo, Japan* 102 (2015).

[108] N. F. Noy. "Semantic integration: a survey of ontology-based approaches". In: *ACM Sigmod Record* 33.4 (2004), pp. 65–70.

[109] J. I. Olszewska, M. Barreto, J. Bermejo-Alonso, J. Carbonera, A. Chibani, S. Fiorini, P. Goncalves, M. Habib, A. Khamis, A. Olivares, et al. "Ontology for autonomous robotics". In: *2017 26th IEEE international symposium on robot and human interactive communication (RO-MAN)*. IEEE. 2017, pp. 189–194.

[110] F. Özcan, C. Lei, A. Quamar, and V. Efthymiou. "Semantic enrichment of data for AI applications". In: *Proceedings of the Fifth Workshop on Data Management for End-To-End Machine Learning*. 2021, pp. 1–7.

[111] A. Pacaci, S. Gonul, A. A. Sinaci, M. Yuksel, and G. B. Laleci Erturkmen. "A semantic transformation methodology for the secondary use of observational healthcare data in postmarketing safety studies". In: *Frontiers in Pharmacology* 9 (2018), p. 435.

[112] F. Pan and J. R. Hobbs. *Time Ontology in OWL*. Tech. rep. 2006.

[113] H. Panetto, M. Dassisti, and A. Tursi. "ONTO-PDM: Product-driven ONTOlogy for Product Data Management interoperability within manufacturing process environment". In: *Advanced Engineering Informatics* 26.2 (2012), pp. 334–348.

[114] J. Park and J. Lee. "Presentation on Korea smart factory program". In: *Proceedings of the international conference on advances in production management systems, Tokyo, Japan*. 2015, pp. 5–9.

[115] P. Patel, M. I. Ali, and A. Sheth. "From raw data to smart manufacturing: AI and semantic web of things for industry 4.0". In: *IEEE Intelligent Systems* 33.4 (2018), pp. 79–86.

[116] N. Petersen, M. Galkin, C. Lange, S. Lohmann, and S. Auer. "Monitoring and automating factories using semantic models". In: *Semantic Technology: 6th Joint International Conference, JIST 2016, Singapore, Singapore, November 2-4, 2016, Revised Selected Papers 6*. Springer. 2016, pp. 315–330.

[117] M. A. Pisching, M. A. Pessoa, F. Junqueira, D. J. dos Santos Filho, and P. E. Miyagi. "An architecture based on RAMI 4.0 to discover equipment to process operations required by products". In: *Computers & Industrial Engineering* 125 (2018), pp. 574–591.

[118] D. G. Pivoto, L. F. de Almeida, R. da Rosa Righi, J. J. Rodrigues, A. B. Lugli, and A. M. Alberti. "Cyber-physical systems architectures for industrial internet of things applications in Industry 4.0: A literature review". In: *Journal of manufacturing systems* 58 (2021), pp. 176–192.

[119] A. Poggi, D. Lembo, D. Calvanese, G. De Giacomo, M. Lenzerini, and R. Rosati. "Linking data to ontologies". In: *Journal on data semantics X*. Springer. 2008, pp. 133–173.

[120] M. Poveda-Villalón, A. Fernández-Izquierdo, M. Fernández-López, and R. García-Castro. "LOT: An industrial oriented ontology engineering framework". In: *Engineering Applications of Artificial Intelligence* 111 (2022), p. 104755.

[121] M. Poveda-Villalón, A. Gómez-Pérez, and M. C. Suárez-Figueroa. "OOPS! (OntOlogy Pitfall Scanner!): An On-line Tool for Ontology Evaluation". In: *International Journal on Semantic Web and Information Systems (IJSWIS)* 10.2 (2014), pp. 7–34.

[122] E. Prestes, J. L. Carbonera, S. R. Fiorini, V. A. Jorge, M. Abel, R. Madhavan, A. Locoro, P. Goncalves, M. E. Barreto, M. Habib, et al. "Towards a core ontology for robotics and automation". In: *Robotics and Autonomous Systems* 61.11 (2013), pp. 1193–1204.

[123] J. Pullmann, N. Petersen, C. Mader, S. Lohmann, and Z. Kemeny. "Ontology-based information modelling in the industrial data space". In: *2017 22nd IEEE International Conference on Emerging Technologies and Factory Automation (ETFA)*. IEEE. 2017, pp. 1–8.

[124] E. Rahm. "The case for holistic data integration". In: *Advances in Databases and Information Systems: 20th East European Conference, ADBIS 2016, Prague, Czech Republic, August 28-31, 2016, Proceedings 20*. Springer. 2016, pp. 11–27.

[125] D. G. Rajpathak. "An ontology based text mining system for knowledge discovery from the diagnosis data in the automotive domain". In: *Computers in Industry* 64.5 (2013), pp. 565–580.

[126] V. J. Ramírez-Durán, I. Berges, and A. Illarramendi. "ExtruOnt: An ontology for describing a type of manufacturing machine for Industry 4.0 systems". In: *Semantic Web* 11.6 (2020), pp. 887–909.

[127] B. Ramis Ferrer, B. Ahmad, D. Vera, A. Lobov, R. Harrison, and J. L. Martínez Lastra. "Product, process and resource model coupling for knowledge-driven assembly automation". In: *at-Automatisierungstechnik* 64.3 (2016), pp. 231–243.

[128] S. Ren, Y. Zhang, Y. Liu, T. Sakao, D. Huisingh, and C. M. Almeida. "A comprehensive review of big data analytics throughout product lifecycle to support sustainable smart manufacturing: A framework, challenges and future research directions". In: *Journal of cleaner production* 210 (2019), pp. 1343–1365.

[129] A. Rhayem, M. B. A. Mhiri, and F. Gargouri. "Semantic web technologies for the internet of things: Systematic literature review". In: *Internet of Things* 11 (2020), p. 100206.

[130] M. Riga, E. Kontopoulos, K. Ioannidis, S. Kintzios, S. Vrochidis, and I. Kompatsiaris. "EUCISE-OWL: An ontology-based representation of the Common Information Sharing Environment (CISE) for the maritime domain". In: *Semantic Web* 12.4 (2021), pp. 603–615.

[131] H. Rijgersberg, M. Van Assem, and J. Top. "Ontology of units of measure and related concepts". In: *Semantic Web* 4.1 (2013), pp. 3–13.

[132] M. de Roode, A. Fernández-Izquierdo, L. Daniele, M. Poveda-Villalón, and R. García-Castro. "SAREF4INMA: A SAREF extension for the industry and manufacturing domain". In: *Semantic Web* 11.6 (2020), pp. 911–926.

[133] A. Rossi, D. Barbosa, D. Firmani, A. Matinata, and P. Merialdo. "Knowledge graph embedding for link prediction: A comparative analysis". In: *ACM Transactions on Knowledge Discovery from Data (TKDD)* 15.2 (2021), pp. 1–49.

[134] S. Saeidlou, M. Saadat, E. Amini Sharifi, and G. D. Jules. "An ontology-based intelligent data query system in manufacturing networks". In: *Production & Manufacturing Research* 5.1 (2017), pp. 250–267.

[135] S. Saha, Z. Usman, W. Li, S. Jones, and N. Shah. "Core domain ontology for joining processes to consolidate welding standards". In: *Robotics and Computer-Integrated Manufacturing* 59 (2019), pp. 417–430.

[136] H. Saitwal, D. Qing, S. Jones, E. V. Bernstam, C. G. Chute, and T. R. Johnson. "Cross-terminology mapping challenges: a demonstration using medication terminological systems". In: *Journal of Biomedical Informatics* 45.4 (2012), pp. 613–625.

[137] S. El-Sappagh, J. M. Alonso, F. Ali, A. Ali, J.-H. Jang, and K.-S. Kwak. "An ontology-based interpretable fuzzy decision support system for diabetes diagnosis". In: *IEEE Access* 6 (2018), pp. 37371–37394.

[138] A. Sarkar, D. Sormaz, and M. H. Karray. "Taxonomy of Manufacturing Joining Operations Based on Process Characterization". In: *International Conference on Flexible Automation and Intelligent Manufacturing*. Springer. 2022, pp. 711–722.

[139] C. Schlenoff, P. Denno, R. Ivester, S. Szykman, and D. Libes. "An analysis of existing ontological systems for applications in manufacturing". In: *International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*. Vol. 19722. American Society of Mechanical Engineers. 1999, pp. 935–944.

[140] B. Schmidt, L. Wang, and D. Galar. "Semantic framework for predictive maintenance in a cloud environment". In: *Procedia CIRP* 62 (2017), pp. 583–588.

[141] G. Schreiber, B. Wielinga, H. Akkermans, W. Van de Velde, and A. Anjewierden. "CML: The commonKADS conceptual modelling language". In: *A Future for Knowledge Acquisition*. Ed. by L. Steels, G. Schreiber, and W. Van de Velde. Berlin, Heidelberg: Springer Berlin Heidelberg, 1994, pp. 1–25. ISBN: 978-3-540-49006-7.

[142] A. Schultz, A. Matteini, R. Isele, C. Bizer, and C. Becker. "LDIF-linked data integration framework". In: *Proceedings of the Second International Conference on Consuming Linked Data-Volume 782*. 2011, pp. 125–130.

[143] A. Seeliger, M. Pfaff, and H. Krcmar. "Semantic web technologies for explainable machine learning models: A literature review." In: *PROFILES/SEMEX@ International Semantic Web Conference (ISWC)* 2465 (2019), pp. 1–16.

[144] A. Seyedamir, B. R. Ferrer, and J. L. M. Lastra. "An ISA-95 based ontology for manufacturing systems knowledge description extended with semantic rules". In: *2018 IEEE 16th International Conference on Industrial Informatics (INDIN)*. IEEE. 2018, pp. 374–380.

[145] M. D. Smucker, J. Allan, and B. Carterette. "A comparison of statistical significance tests for information retrieval evaluation". In: *Proceedings of the Sixteenth ACM Conference on Conference on Information and Knowledge Management*. CIKM '07. Lisbon, Portugal: Association for Computing Machinery, 2007, pp. 623–632. ISBN: 9781595938039. DOI: 10.1145/1321440.1321528. URL: https://doi.org/10.1145/1321440.1321528.

[146] L. Solano. "Ontological modelling of welding processes". In: *IOP Conference Series: Materials Science and Engineering*. Vol. 1193. 1. IOP Publishing. 2021, p. 012019.

[147] B. Steenwinckel, G. Vandewiele, M. Weyns, T. Agozzino, F. D. Turck, and F. Ongenae. "INK: knowledge graph embeddings for node classification". In: *Data Mining and Knowledge Discovery* 36.2 (2022), pp. 620–667.

[148] Y. Svetashova, B. Zhou, T. Pychynski, S. Schmidt, Y. Sure-Vetter, R. Mikut, and E. Kharlamov. "Ontology-enhanced machine learning: a Bosch use case of welding quality monitoring". In: *The Semantic Web–ISWC 2020: 19th International Semantic Web Conference, Athens, Greece, November 2–6, 2020, Proceedings, Part II 19*. Springer. 2020, pp. 531–550.

[149] Y. Svetashova, B. Zhou, S. Schmid, T. Pychynski, and E. Kharlamov. "SemML: Reusable ML for Condition Monitoring in Discrete Manufacturing." In: *International Semantic Web Conference (ISWC) (Demos/Industry)* 2721 (2020), pp. 213–218.

[150] F. Tao, Q. Qi, A. Liu, and A. Kusiak. "Data-driven smart manufacturing". In: *Journal of Manufacturing Systems* 48 (2018), pp. 157–169.

[151] G. Tebes, L. Olsina, D. Peppino, and P. Becker. "TestTDO: A Top-Domain Software Testing Ontology." In: *CIbSE - Iberoamerican Conference on Software Engineering*. 2020, pp. 364–377.

[152] N. Teslya and I. Ryabchikov. "Ontology-driven approach for describing industrial socio-cyberphysical systems' components". In: *MATEC Web of Conferences*. Vol. 161. EDP Sciences. 2018, p. 03027.

[153] K.-D. Thoben, S. Wiesner, and T. Wuest. ""Industrie 4.0" and smart manufacturing-a review of research issues and application examples". In: *International Journal of Automation Technology* 11.1 (2017), pp. 4–16.

[154] I. Tiddi and S. Schlobach. "Knowledge graphs as tools for explainable machine learning: A survey". In: *Artificial Intelligence* 302 (2022), p. 103627.

[155] N.-H. Tran, H.-S. Park, Q.-V. Nguyen, and T.-D. Hoang. "Development of a smart cyber-physical manufacturing system in the industry 4.0 context". In: *Applied Sciences* 9.16 (2019), p. 3325.

[156] D. Tranfield, D. Denyer, and P. Smart. "Towards a methodology for developing evidence-informed management knowledge by means of systematic review". In: *British Journal of mManagement* 14.3 (2003), pp. 207–222.

[157] T. Trouillon, J. Welbl, S. Riedel, É. Gaussier, and G. Bouchard. "Complex embeddings for simple link prediction". In: *International conference on machine learning*. PMLR. 2016, pp. 2071–2080.

[158] A. Ustundag, E. Cevikcan, C. Salkin, M. Oner, A. Ustundag, and E. Cevikcan. "A conceptual framework for Industry 4.0". In: *Industry 4.0: Managing the Digital Transformation* (2018), pp. 3–23.

[159] M. Vidoni and A. Vecchietti. "Towards a Reference Architecture for Advanced Planning Systems". In: *Proceedings of the 18th International Conference on Enterprise Information Systems (ICEIS 2016-Volume 1)*. SciTePress. 2016, pp. 434–440.

[160] T. Vu, T. D. Nguyen, D. Q. Nguyen, D. Phung, et al. "A capsule network-based embedding model for knowledge graph completion and search personalization". In: *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*. 2019, pp. 2180–2189.

[161] J. Wan, B. Chen, M. Imran, F. Tao, D. Li, C. Liu, and S. Ahmad. "Toward dynamic resources management for IoT-based manufacturing". In: *IEEE Communications Magazine* 56.2 (2018), pp. 52–59.

[162] J. Wan, S. Tang, D. Li, M. Imran, C. Zhang, C. Liu, and Z. Pang. "Reconfigurable smart factory for drug packing in healthcare industry 4.0". In: *IEEE transactions on industrial informatics* 15.1 (2018), pp. 507–516.

[163] J. Wan, B. Yin, D. Li, A. Celesti, F. Tao, and Q. Hua. "An ontology-based resource reconfiguration method for manufacturing cyber-physical systems". In: *IEEE/ASME Transactions on Mechatronics* 23.6 (2018), pp. 2537–2546.

[164] H. Wang, Z. Cui, R. Liu, L. Fang, and Y. Sha. "A Multi-type Transferable Method for Missing Link Prediction in Heterogeneous Social Networks". In: *IEEE Transactions on Knowledge and Data Engineering* (2023).

[165] H. Wang and Z. Le. "Seven-layer model in complex networks link prediction: A survey". In: *Sensors* 20.22 (2020), p. 6560.

[166] L. Wang, M. Törngren, and M. Onori. "Current status and advancement of cyber-physical systems in manufacturing". In: *Journal of Manufacturing Systems* 37 (2015), pp. 517–527.

[167] M. Wang, L. Qiu, and X. Wang. "A survey on knowledge graph embeddings for link prediction". In: *Symmetry* 13.3 (2021), p. 485.

[168] S. Wang, J. Wan, D. Zhang, D. Li, and C. Zhang. "Towards smart factory for industry 4.0: a self-organized multi-agent system with big data based feedback and coordination". In: *Computer Networks* 101 (2016). Industrial Technologies and Applications for the Internet of Things, pp. 158–168. ISSN: 1389-1286.

[169] M. D. Wilkinson, M. Dumontier, I. J. Aalbersberg, G. Appleton, M. Axton, A. Baak, N. Blomberg, J.-W. Boiten, L. B. da Silva Santos, P. E. Bourne, J. Bouwman, A. J. Brookes, et al. "The FAIR Guiding Principles for scientific data management and stewardship". In: *Scientific Data* 3.1 (2016), p. 160018. ISSN: 2052-4463.

[170] S. Xiao, Y. Hu, J. Han, R. Zhou, and J. Wen. "Bayesian networks-based association rules and knowledge reuse in maintenance decision-making of industrial product-service systems". In: *Procedia Cirp* 47 (2016), pp. 198–203.

[171] L. Xiaoxue, B. Xuesong, W. Longhe, R. Bingyuan, L. Shuhan, and L. Lin. "Review and Trend Analysis of Knowledge Graphs for Crop Pest and Diseases". In: *IEEE Access* 7 (2019), pp. 62251–62264.

[172] C. Xu, B. Min, X. Sun, J. Hu, B. Li, and Y. Duan. "MULAPI: A tool for API method and usage location recommendation". In: *2019 IEEE/ACM 41st international conference on software engineering: companion proceedings (ICSE-Companion)*. IEEE. 2019, pp. 119–122.

[173] M. Yahya, A. Ali, Q. Mehmood, L. Yang, J. G. Breslin, and M. I. Ali. "A benchmark dataset with Knowledge Graph generation for Industry 4.0 production lines". In: *Semantic Web* Preprint (), pp. 1–19.

[174] M. Yahya, A. Breathnach, F. Khan, I. Abaspur, and R. Ranganathan. "Towards Semantic Modeling of Camera from Image Quality Testing Perspective: Valeo Vision Systems Case". In: *Proceedings of the 2nd International Workshop on Semantic Industrial Information Modelling (SemIIM 2023) co-located with 22nd International Semantic Web Conference (ISWC 2023)*. 2023.

[175] M. Yahya, J. G. Breslin, and M. I. Ali. "Semantic web and knowledge graphs for industry 4.0". In: *Applied Sciences* 11.11 (2021), p. 5110.

[176] H. Yan, J. Yang, and J. Wan. "KnowIME: a system to construct a knowledge graph for intelligent manufacturing equipment". In: *IEEE Access* 8 (2020), pp. 41805–41813.

[177] B. Yang, S. W.-t. Yih, X. He, J. Gao, and L. Deng. "Embedding Entities and Relations for Learning and Inference in Knowledge Bases". In: *Proceedings of the International Conference on Learning Representations (ICLR) 2015*. 2015.

[178] C. Yong, W. Lee, and C. Lazarus. "Industry 4.0 reference architectural models: Critical review and opportunities". In: *International Journal of Advanced Research in Engineering Innovation* 3.4 (2021), pp. 40–49.

[179] X. Zeng, X. Tu, Y. Liu, X. Fu, and Y. Su. "Toward better drug discovery with knowledge graph". In: *Current opinion in structural biology* 72 (2022), pp. 114–126.

[180] F. Zhang, N. J. Yuan, D. Lian, X. Xie, and W.-Y. Ma. "Collaborative Knowledge Base Embedding for Recommender Systems". In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. KDD '16. San Francisco, California, USA: Association for Computing Machinery, 2016, pp. 353–362. ISBN: 9781450342322.

[181] W. Zhang, B. Paudel, W. Zhang, A. Bernstein, and H. Chen. "Interaction embeddings for prediction and explanation in knowledge graphs". In: *Proceedings of the Twelfth ACM International Conference on Web Search and Data Mining*. 2019, pp. 96–104.

[182] B. Zhou, Y. Svetashova, A. Gusmao, A. Soylu, G. Cheng, R. Mikut, A. Waaler, and E. Kharlamov. "SemML: Facilitating development of ML models for condition monitoring with semantics". In: *Journal of Web Semantics* 71 (2021), p. 100664.

[183] B. Zhou, D. Zhou, J. Chen, Y. Svetashova, G. Cheng, and E. Kharlamov. "Scaling usability of ML analytics with knowledge graphs: Exemplified with a Bosch welding case". In: *The 10th International Joint Conference on Knowledge Graphs*. 2021, pp. 54–63.

[184] K. Zhou and P. Yao. "Overview of recent advances of process analysis and quality control in resistance spot welding". In: *Mechanical Systems and Signal Processing* 124 (2019), pp. 170–198.

[185] K. Zhou, T. Liu, and L. Zhou. “Industry 4.0: Towards future industrial opportunities and challenges”. In: *2015 12th International Conference on Fuzzy Systems and Knowledge Discovery (FSKD)*. 2015, pp. 2147–2152.

[186] M. Zloch, M. Acosta, D. Hienert, S. Conrad, and S. Dietze. “Charaterizing RDF graphs through graph-based measures–framework and assessment”. In: *Semantic Web* 12.5 (2021), pp. 789–812.

[187] X. Zou. “A survey on application of knowledge graph”. In: *Journal of Physics: Conference Series*. Vol. 1487. IOP Publishing. 2020, p. 012016.