CRANFIELD UNIVERSITY

IÑIGO FERNÁNDEZ DEL AMO BLANCO

ONTOLOGY-BASED AUGMENTED REALITY CONTENT-RELATED TECHNIQUES AND THEIR IMPACT IN KNOWLEDGE CAPTURE AND RE-USE WITHIN MAINTENANCE DIAGNOSIS

SCHOOL OF AEROSPACE, TRANSPORT AND MANUFACTURING
Phd In Manufacturing

PhD
Academic Year: 2017 - 2020

Supervisor: Dr. John Ahmet Erkoyuncu
Associate Supervisor: Dr. Maryam Farsi
May 2020

CRANFIELD UNIVERSITY

SCHOOL OF AEROSPACE, TRANSPORT AND MANUFACTURING
Phd In Manufacturing

PhD

Academic Year 2017 - 2020

IÑIGO FERNÁNDEZ DEL AMO BLANCO

Ontology-based Augmented Reality content-related techniques and their impact in knowledge capture and re-use within maintenance diagnosis

Supervisor: Dr. John Ahmet Erkoyuncu
Associate Supervisor: Dr. Maryam Farsi
May 2020

Abstract

This PhD thesis aims to study ontology-based AR content-related methods and their impact in knowledge transfer, capture and re-use for cost-effective human knowledge integration in digital diagnostic systems. Industry 4.0 has revealed the importance of maintainers' knowledge capture and re-use in diagnostics systems for providing satisfactory solutions in cases where those systems cannot (e.g. no-fault-found). Augmented Reality (AR) utilises content-related techniques to transfer knowledge to maintainers for improving efficiency and effectiveness of diagnosis tasks. Academic literature has shown that AR can also be utilised for knowledge capture and re-use, but this has only been demonstrated in simple, step-by-step repair operations. In diagnosis research, ontology-based methods are applied to capture and re-use knowledge from unstructured and heterogenous sources like humans. Nevertheless, these methods have not made use of AR potential to contextualise knowledge and so, improve efficiency and effectiveness of knowledge capture and re-use diagnosis operations.

This PhD thesis aims to demonstrate that ontology-based AR content-related methods can enable knowledge transfer, capture and re-use knowledge in different diagnosis operations. Thus, improving diagnosis efficiency and effectiveness while reducing implementation costs to enhance human knowledge integration in digital diagnostic systems. For that purpose, this thesis proposes three contributions:

1. Ontology-based reporting and monitoring methods to enable knowledge capture and re-use for improving efficiency of diverse diagnosis tasks.

2. An ontology-based, automatic AR content-creation method to reduce AR implementation costs while gaining sufficient knowledge transfer effectiveness for improving operational efficiency and effectiveness.

3. An ontology-based AR knowledge recommender method to enable knowledge re-use for improving efficiency and effectiveness of diagnosis reporting tasks.

Through diverse validation methods including stopwatch experiments, usability surveys and expert interviews applied to several case studies in complex assets of variable nature (e.g. mechanical or electronic), these contributions have proven that:

1. Ontology-based reporting and monitoring methods can increase structure and accuracy of captured knowledge to enhance knowledge integration and to re-use it for improving complex fault-finding tasks.
2. The ontology-based, automatic AR content-creation method standardises content-creation processes for diverse maintenance operations and reduces related costs to facilitate integration of AR in information systems.
3. The ontology-based AR recommender method contextualises and standardises expert knowledge capture and re-use to enhance digitalisation of diagnosis reporting tasks.

Collectively, these contributions enable to create and adapt augmented content automatically as well as to capture and re-use diagnosis knowledge through web and AR applications. Therefore, they can be considered to conform a framework for human knowledge integration in digital diagnostic systems. Such framework has been implemented as a prototype system for validation purposes not only of this research, but also of consequent research works regarding human knowledge integration in Digital Twins and degradation assessment. Besides, it has been tested in industrial environments with satisfactory first impressions. Hence, future research should investigate its applicability to other maintenance operations and assets' lifecycle phases to demonstrate its potential for human knowledge integration in Industry 4.0 contexts.

Acknowledgements

I am indebted to all the great people who have guided and assisted me during my doctoral studies and those whose support made this thesis possible.

First and foremost, I would like to express my deepest gratitude to my supervisor, Dr. John Ahmet Erkoyuncu, for his outstanding guidance, encouragement and professionalism. His vision and expertise have inspired me to surpass myself academically, professionally and personally. The provided opportunities to undertake this PhD and contribute beyond it have been highly rewarding. I would like to thank my associate supervisor, Dr. Maryam Farsi, whose attentive and constant support have improved the quality of this thesis. I would also like to acknowledge Prof. Rajkumar Roy, whose direction and creativity helped me set this research in the right path.

I would like to express my sincere appreciation to Babcock International and its associates for funding and supporting this research. In particular, I would like to thank Stephen Wilding, whose continuous support and motivation have inspired me to be as innovative as one can be. I would also like to acknowledge Jim Sibson for his encouraging recognition and the highly rewarding opportunities to contribute beyond this research.

My gratitude to the Through-Life Engineering Service Centre and its members for supporting me throughout the hardest stages of this research. Special thanks to Dr. Pavan Addepalli, Samuel Court, Dr. Riccardo Palmarini and Teresa Bandee for their advice without which this thesis would have not been possible.

Last but not least, I would like to express my most sincere gratitude to my family and friends to whom I am indebted for being there in good and bad times. Special thanks to my mother Ana, who always supported me through thick and thin, and to my father Gregorio, may his soul keep guiding me in death. I am really grateful to Marta, whose love and patience these years made my journey worthwhile.

List of publications

Journal papers

1. Fernández Del Amo, I., Erkoyuncu, J. A. and Farsi, M. (2020) Hybrid recommendations and dynamic authoring for AR knowledge capture and re-use in maintenance diagnosis applications, Knowledge-Based Systems. *Submitted, June 2020.*

2. Fernández Del Amo, I., Erkoyuncu, J. A. and Farsi, M. (2020b) Programmable content and a real-time ontology-based pattern-matching algorithm for automatic adaptive authoring in AR for maintenance, Expert Systems with Applications. *Submitted, June 2020*

3. Fernández Del Amo, I., Erkoyuncu, J., Bulka, D., Farsi, M. and Wilding, S. (2020) Ontology-based diagnosis reporting and monitoring to improve fault finding in Industry 4.0, Knowledge-Based Systems. *Submitted, June 2020.*

4. Erkoyuncu, J. A., Fernández del Amo, I., Ariansyaha, D., Bulka, D., Vrabič, R. and Roy, R. (2020) 'A Design Framework for Adaptive Digital Twins', CIRP Annals - Manufacturing Technology, 00, pp. 8–11. doi: 10.1016/j.cirp.2020.04.086.

5. Fernández del Amo, I., Erkoyuncu, J., Vrabič, R., Frayssinet, R., Vazquez Reynel, C. and Roy, R. (2019) 'Structured authoring for AR-based communication to enhance efficiency in remote diagnosis for complex equipment', Advanced Engineering Informatics. Elsevier, 45(March), p. 15. doi: 10.1016/j.aei.2020.101096.

6. Fernández Del Amo, I., Erkoyuncu, J. A., Roy, R., Palmarini, R. and Onoufriou, D. (2018) 'A systematic review of Augmented Reality content-related techniques for knowledge transfer in maintenance applications', Computers in Industry, 103, pp. 47–71. doi: 10.1016/j.compind.2018.08.007.

7. Erkoyuncu, J. A., Fernández del Amo, I., Dalle Mura, M., Roy, R. and Dini, G. (2017) 'Improving efficiency of industrial maintenance with context aware adaptive authoring in augmented reality', CIRP Annals - Manufacturing Technology. doi: 10.1016/j.cirp.2017.04.006.

Conference papers

1. Ariansyah, D., Fernández del Amo, I., Erkoyuncu, J. A., Agha, M., Bulka, D., De la Puante, J., Langlois, M., M'Khinini, Y., Sibson, J. and Penver, S. (2020) 'Digital Twin Development: A Step by Step Guideline', in 9th International Conference on Through-life Engineering Services (TESConf 2020). Cranfield, United Kingdom. *Submitted, June 2020.*

2. Fernández Del Amo, I., Galeotti, E., Palmarini, R., Dini, G., Erkoyuncu, J. and Roy, R. (2018) 'An innovative user-centred support tool for Augmented Reality maintenance systems design: A preliminary study', in Procedia CIRP, pp. 362–367. doi: 10.1016/j.procir.2018.02.020.

3. Fernández Del Amo, I., Erkoyuncu, J. A., Roy, R. and Wilding, S. (2018) 'Augmented Reality in Maintenance: An information-centred design framework', in Procedia Manufacturing, pp. 148–155. doi: 10.1016/j.promfg.2018.01.021.

4. Palmarini, R., Del Amo, I. F., Bertolino, G., Dini, G., Erkoyuncu, J. A., Roy, R. and Farnsworth, M. (2018) 'Designing an AR interface to improve trust in Human-Robots collaboration', in Procedia CIRP, pp. 350–355. doi: 10.1016/j.procir.2018.01.009.

Book chapters

Fernández Del Amo, I., Erkoyuncu, J. and Roy, R. (2019) An Augmented Reality System Model for Integration with Maintenance Information Systems, Augmented, Virtual, and Mixed Reality Applications in Advanced Manufacturing. World Scientific. *Accepted for publication, December 2018.*

Table of contents

List of figures

List of tables

List of abbreviations

AR	Augmented Reality
A	Authoring
CA	Context-Awareness
IA	Interaction-Analyses
DT	Digital Twins
AI	Artificial Intelligence
AR	Attribute Richness
RR	Relationship Richness
IR	Inheritance Richness
SLR	Systematic Literature Review

Chapter 1
Introduction

Augmented Reality (AR) can help to integrate human knowledge in Industry 4.0's digital revolution (Gattullo *et al.*, 2019). AR is a set of human-computer interaction technologies that enrich users' real-world perception by embedding virtual information in spatial and temporal coexistence with real-world objects (Azuma, 2016; Bottani and Vignali, 2019). AR technologies have therefore the ability to transfer knowledge from humans to information systems and vice versa (Longo, Nicoletti and Padovano, 2019). Thus, allowing to improve efficiency and effectiveness of human tasks (e.g. repair) and to enrich more digital processes (e.g. monitoring) through knowledge capture and re-use (van Lopik *et al.*, 2020).

Knowledge capture and re-use refer to the processes of retrieving explicit or implicit knowledge (capture) from humans or artefacts and re-applying it in other actions or contexts (re-use) (Becerra-Fernandez and Sabherwal, 2010; Pérez-Salazar *et al.*, 2019). The role of humans in Industry 4.0 is fostering research attention because of their abilities to ubiquitously capture and re-use knowledge for managing its increasing complexities (Li, Fast-Berglund and Paulin, 2019). A relevant area where humans are of great use is maintenance (Roy *et al.*, 2016). Unlike robots or intelligent systems, humans can adapt to conduct complex repairs or diagnose unpredicted failures using their experience (Pistofidis *et al.*, 2016). Through knowledge capture and re-use, AR technologies can provide new valuable information streams to enable more fault-tolerant and adaptive assets and production systems (Kong *et al.*, 2019).

1.1 Augmented Reality and content-related techniques

AR benefits in maintenance of complex assets include efficiency increase (57%), safety growth (55%), productivity rise (52%) and complexity reduction (Bottani and Vignali, 2019). These benefits have made AR maintenance applications to be among the first closing the gap towards commercial deployment (Capgemini Research Institute, 2018). Despite their maturity, there are still significant research challenges to overcome:

1. Added value quantification: there is little academic evidence on standard frameworks for AR system's (Li, Nee and Ong, 2017) and cost-benefit models and analysis of AR applications (Suárez-Warden *et al.*, 2015).

2. Hardware enhancement: improved ergonomics and acceptance require better capabilities of optical see-through devices (e.g. wider field-of-view) and more complex input/output methods (e.g. eye tracking) (Azuma, 2016).

3. Software commercialisation: improved robustness is essential to achieve sufficient commercial reliability and cost-effectiveness (Palmarini *et al.*, 2018). Relevant research areas include:

a. Improved image processing algorithms for real-object tracking in varying environments with accuracy (X. Wang, Ong and Nee, 2016).

b. Automation of content-related techniques to reduce deployment costs (Egger and Masood, 2020).

4. Advanced applications: enhanced content-related techniques for improving interactivity of maintainers with AR content (X. Wang, Ong and Nee, 2016) can increase AR added-value for long-term maintenance challenges like human knowledge integration (Bottani and Vignali, 2019).

AR content-related techniques (Table 1-1) are software methods that enable to create (authoring), adapt (context-awareness) and enrich (interaction-analysis) augmented content for effective knowledge transfer (Gattullo *et al.*, 2019). Recent research has also demonstrated their potential for knowledge capture and re-use applications, but only in tangible operations like assembly tasks (Ramirez-Amaro, Beetz and Cheng, 2017; Suarez-Warden and González Mendívil, 2017). For more conceptual operations such as diagnosis, there are still research challenges to solve (Fernández Del Amo *et al.*, 2018). These include automatic authoring of pre-existing maintenance information (Egger and Masood, 2020) or context-aware data filtering for accurate knowledge capture (Wang *et al.*, 2020) among others.

Table 1-1. Definitions of main software methods related to AR applications.

Method	Definition
Authoring	Software techniques that create and maintain augmented content overtime for AR applications regarding specific knowledge to transfer to specific users in specific environments (Zhu, Ong and Nee, 2015; Palmarini *et al.*, 2018).
Context-Awareness	Software techniques that determine the context of the augmented scene in terms of user, environment and application for modifying augmented content accordingly (Manuri and Sanna, 2016; Akçayir and Akçayir, 2017)
Interaction-Analysis	Software techniques that analyse augmented content regarding support provided for the task at hand and return feedback automatically (X. Wang, Ong and Nee, 2016).
Ontologies	Sets of statements and rules that can capture and specify vocabulary within software applications, making it understandable by humans and processable by computers (Breitman, Casanova and Truszkowski, 2007; Flotyński and Walczak, 2017)
Recommender systems	Software techniques that predict the rating of information items for recommending them to users according to variable contextual parameters (Adomavicius and Tuzhilin, 2015; Torres-Ruiz *et al.*, 2020)

1.2 Digital maintenance and human knowledge integration

Industry 4.0 refers to the ongoing digital transformation and automation of conventional manufacturing and industry-related value chains (Oztemel and Gursev, 2020). Industry 4.0 has revealed the need to digitalise maintenance for sustaining competitive advantages (Rødseth, Schjølberg and Marhaug, 2017). Digital maintenance refers to the use of information and communication technologies for improving asset availability, safety and sustainability and dealing with the ever increasing complexity of engineering assets (Candell, Karim and Söderholm, 2009; Roy *et al.*, 2016). Technologies like Artificial Intelligence (AI) and Digital Twins (DT) allow for real-time monitoring and prognosis for improving operational efficiency and decision-making (Cimino, Negri and Fumagalli, 2019). These improvements are enabled through automatic data exchange between multiple devices such as sensors, 3D scanners, robots or Radio Frequency Identification Devices (RFID) (Khan and Yairi, 2018). These require integrated data management that is still challenging due to heterogeneity (e.g. audio or images) and lack of structure (e.g. manuals or reports) of numerous data sources (Angelopoulos *et al.*, 2020).

Integrated data management is particularly significant in maintenance diagnosis (Yazdi, 2019). Diagnostic systems are becoming more automated, but still require human experts when those do not provide satisfactory solutions (e.g. no-fault-found) (Vogl, Weiss and Helu, 2019). Explicitly capturing their knowledge can enhance maintenance efficiency and effectiveness (Nuñez and Borsato, 2017) due to its potential to be reused by others (e.g. novices) or in different activities (e.g. repair or monitoring) (Pistofidis *et al.*, 2016). Maintenance diagnosis research has utilised semantic computing methods such as ontologies (Table 1-1) to capture knowledge due to their abilities to correlate information from unstructured sources (Zhong *et al.*, 2018). Nevertheless, knowledge re-use is still challenging due to the complexities associated with variable contexts such as configuration management (Liu *et al.*, 2019). Recommendation techniques (Table 1-1) like case-based reasoning have been used in diagnosis knowledge re-use, although context-aware approaches are limited due to the difficulty of capturing contextual data (Wan *et al.*, 2019). AR technologies use context-aware and interaction-analysis methods to analyse real environments for enhancing virtual information that AR embeds in real-world scenarios. These could be applied to analyse diagnosis scenarios for improving knowledge re-use and thus, achieving the full potential of human knowledge integration in maintenance diagnosis.

1.3 Research outline

1.3.1 Research motivation

This PhD thesis intends to contribute to some of the research challenges described above regarding AR technologies and human knowledge integration in digital maintenance. In particular, those focusing on effective knowledge capture and re-use in diagnosis and AR content-related techniques to demonstrate their added value for human knowledge integration in digital maintenance. This PhD thesis contributions expect to fulfil relevant research gaps in these areas by developing and validating ontology-based AR content-related techniques for integrating human knowledge in maintenance information systems.

Besides research motives, there are also industrial interests to study the impact of AR technologies through knowledge integration in digital maintenance. This PhD thesis has been sponsored by a multinational organisation that provides support services for complex engineering assets and infrastructures in critical and safety environments around the globe. As part of their operations, they conduct complex diagnosis operations that heavily rely on human knowledge to produce satisfactory results. This organisation counts with a very knowledgeable but ageing workforce to conduct these processes that aim to maintain. Therefore, they have a particular interest on studying AR technologies to enable knowledge transfer, capture and re-use for improving their competitive advantage in a cost-effective manner. As part of their sponsorship, this organisation has provided human and material resources to conduct expert interviews and industrial tests to validate this thesis contributions from an industrial perspective.

The basis to demonstrate AR added value for human knowledge integration lays down on the assumption that automating content-related methods can make AR technologies cost-effective. Therefore, automatic and adaptive content-related methods are needed for effective knowledge transfer in maintenance applications. Besides, according to its definition knowledge re-use needs to be validated in two scenarios for different contexts, and for different applications.

1.3.2 Aim and objectives

This PhD thesis hypothesises that:

"By automatically creating, adapting and recommending augmented content, AR applications can transfer, capture and re-use knowledge in different tasks and contexts to improve efficiency and effectiveness of diagnosis operations while reducing implementation costs for enhancing human knowledge integration in digital maintenance".

Therefore, the aim of this research is:

"To develop automatic and ontology-based AR content-related techniques for transferring, capturing and re-using knowledge to improve efficiency and effectiveness of failure diagnosis operations".

In order to achieve this aim, specific research activities are distributed across the following objectives:

1. To identify the relations among AR content-related techniques and knowledge transfer, capture and re-use for integrating human knowledge in digital maintenance.

2. To develop ontology-based reporting and monitoring methods and validate their ability to capture and re-use knowledge in failure diagnosis operations.

3. To develop an ontology-based method for automatic and adaptive authoring and validate its ability for effective knowledge transfer in diverse maintenance operations.

4. To develop an ontology-based, context-aware, interaction-analysis AR recommender method and validate its ability to capture and re-use knowledge in diagnosis reporting.

1.3.3 Methodology

This PhD thesis research methodology is based on Design Science Research (DSR) *"a well-established methodology for research in information systems"* (Nuñez and Borsato, 2017) as illustrated in Figure 1-1.

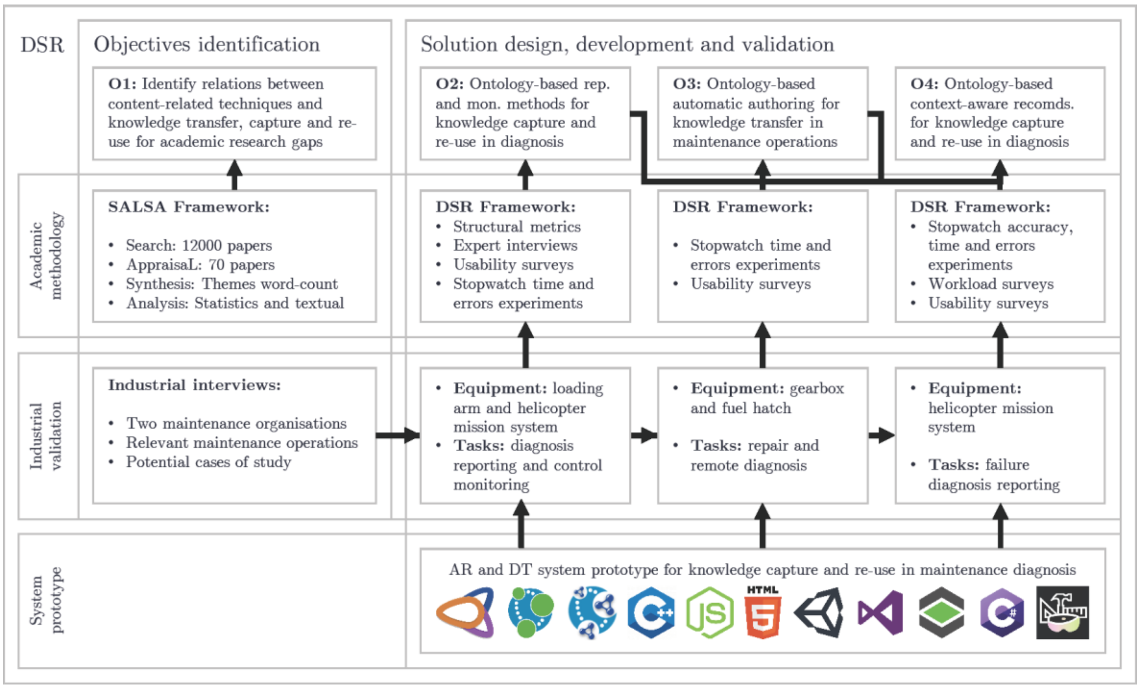


Figure 1-1. Overview of thesis' methodology and independent research objectives including validation methods, cases of study and related software prototyping.

The doctoral research followed such deductive approach to ensure that research objectives contributed to this PhD thesis aim as well as academic knowledge and industrial innovation. DSR provided a framework to conduct research consistently with prior literature and coherently with the academic contributions to validate (Peffers *et al.*, 2008). Each research objective utilised its own specific methodology to ensure that their contributions were validated and consistent with academic literature. The first objective (Figure 1-1 – O1) employed the SALSA framework (Booth, Papaioannou and Sutton, 2012) to conduct a systematic review for finding relations among AR content-related techniques and knowledge transfer, capture and re-use and so, identify this thesis subsequent objectives. Second, third and fourth objectives (Figure 1-1 – O1, O2, O3) utilised their own implementations of the DSR framework to corroborate the proposed research gaps and develop solutions that will contribute to academic literature. These methodologies also included quantitative and qualitative research methods (e.g. stopwatch experiments and usability surveys) and industrially corroborated case studies to demonstrate research contributions' validity from both, academic and industrial, perspectives. These contributions were also implemented as a system prototype to further corroborate the overall thesis aim. Further details on each objective's methodology can be found at their corresponding chapters.

1.3.4 Thesis layout

The thesis comprises six further chapters. These chapters are described below according to their outcomes and implications regarding the PhD thesis aim and objectives. Figure 1-2 is utilised as reference to explain the connections between thesis chapters and their contributions regarding thesis objectives.

Chapter 2 presents a systematic literature review on AR content-related methods, their relations with knowledge transfer, capture and re-use and relevant research gaps (Objective 1). Its results helped to determine this PhD thesis subsequent objectives and its focus on diagnosis for demonstrating AR added value for human knowledge integration in digital maintenance. These included the need to automate AR authoring for cost-effective knowledge transfer and context-aware and interaction-analysis methods to enable knowledge capture and re-use.

Chapter 3 describes ontology-based reporting and monitoring methods to improve failure diagnosis operations (Objective 2). Chapter 2 identifies ontologies describing human-related diagnosis tasks as a relevant research gap for advancing AR knowledge capture and re-use capabilities. Chapter 3 proposes it along with ontology-based diagnosis reporting and monitoring methods (Figure 1-2 – Objective 2) to demonstrate the added value of capturing and re-using human knowledge in different maintenance operations. For validation purposes, this chapter analyses theoretical, experimental and survey results to evaluate diagnosis knowledge capture accuracy and re-use impact in monitoring operations.

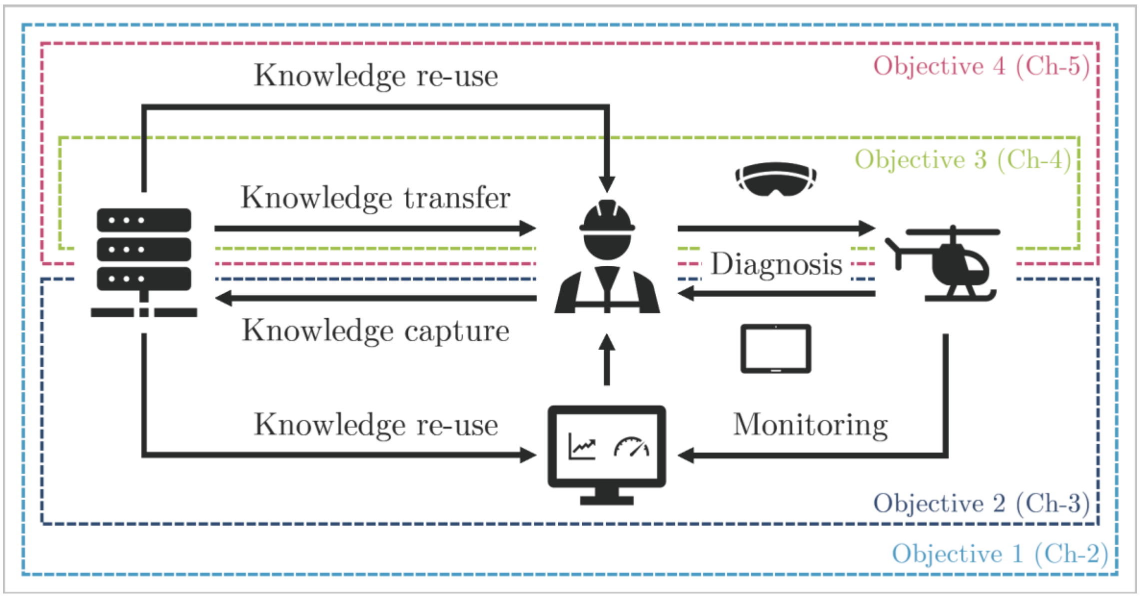


Figure 1-2. Overview of this thesis' objectives, chapters and relations among them.

Chapter 4 introduces an ontology-based, pattern-matching method for automatic adaptive authoring in AR maintenance applications (Objective 3). Chapter 2 devises automatic authoring as not only necessary to enable knowledge transfer but also to reduce AR implementation costs. Therefore, the method that Chapter 4 proposes for automatic adaptive authoring aims to lay the foundations of more advanced techniques for AR knowledge capture and re-use (Figure 1-2 – Objective 3). This method proposes semantic analysis of ontology-based maintenance information to pair it with programmable content formats for enabling automatic adaptive authoring. Research validation includes efficiency experiments and usability surveys to evaluate the impact of resultant augmented content compared to that of conventional authoring methods in two case studies: repair and remote diagnosis. These aim to demonstrate that automatic authoring can achieve similar levels of knowledge transfer effectiveness than manual authoring solutions while reducing AR implementation costs.

Chapter 5 presents an ontology-based, context-aware and interaction-analysis recommendation method for AR knowledge capture and re-use in failure diagnosis (Objective 4). Chapter 2 indicates that enabling AR knowledge capture and re-use requires accurate knowledge domains (Chapter 3) and effective knowledge transfer (Chapter 4). Chapter 5 utilises the contributions described in those two to propose an AR-based hybrid recommender method (Figure 1-2 – Objective 4) that analyses maintenance and user contextual features to suggest most-probable faults for more efficient and effective failure diagnosis reporting. Research validation comprises stopwatch experiments and usability and workload surveys to evaluate the impact of the proposed method in terms of recommendations accuracy and reporting efficiency and effectiveness. These aim to demonstrate that AR not only be useful for capturing knowledge, but also for re-using it in different diagnosis contexts.

Chapter 3, Chapter 4 and Chapter 5 aim to demonstrate that human knowledge integration can be achieved cost-effectively through AR to enable effective knowledge capture and re-use for different diagnosis operations (reporting and monitoring) and contexts (failures). Besides validation results and academic contributions being discussed at each chapter, Chapter 6 presents a global discussion of this thesis results. It examines researches results validity in accordance with this thesis aim as well as the limitations regarding the case studies utilised. Besides, it describes in detail the relationships among chapters contributions, their resultant system prototype and their industrial impact.

Chapter 7 closes this thesis by presenting its conclusions and suggesting future research works. It summarises the implication of its results and its contributions to academic knowledge. It also outlines future research works towards integration of human knowledge in digital maintenance and other fields of AR application.

1.3.5 Related academic publications

This PhD thesis has been written in a paper format style. Each contributory chapter (2-5) has been written as an independent research work as if they were journal publications. Some of them have already been published while others are in the process of submission. Table 1-2 presents the relation between academic publications and this PhD thesis chapters. Additional publications made within the PhD registration period can be found at List of publications (Page v).

Table 1-2. List of academic publications related to this thesis chapters and objectives.

Chapter	Object.	Title	Authors	Journal	Status
2	1	A systematic review of Augmented Reality content-related techniques for knowledge transfer in maintenance applications	I. Fernández del Amo, J. Erkoyuncu, R. Roy, R. Palmarini, D. Onoufriou	Computers in Industry 103 (2018) 47–71	P
3	2	Ontology-based diagnosis reporting and monitoring to improve fault finding in Industry 4.0	I. Fernández del Amo, J. Erkoyuncu, D. Bulka, M. Farsi, S. Wilding	Knowledge-Based Systems	S
4	3	Programmable content and a real-time ontology-based pattern-matching algorithm for automatic adaptive authoring in AR for maintenance	I. Fernández del Amo, J. Erkoyuncu, M. Farsi	Expert Systems with Applications	S
5	4	Hybrid recommendations and dynamic authoring for AR knowledge capture and re-use in maintenance diagnosis applications	I. Fernández del Amo, J. Erkoyuncu, M. Farsi	Knowledge-Based Systems	S

Legend: P = Published; S = To be submitted by June 2020

Chapter 2
A systematic literature review of Augmented Reality content-related techniques for knowledge transfer in maintenance applications

2.1 Introduction

The Augmented Reality (AR) definition has evolved over the years alongside techniques and applications. According to its extended capabilities, AR can be defined as a set of human-computer interaction techniques (X. Wang, Ong and Nee, 2016) that enriches user's real-world experience (Bottani and Vignali, 2019) by embedding contextualised information (Gattullo *et al.*, 2019) into user's space in coexistence with real-world objects (Azuma, 2016). Moreover, Nonaka (1994) defines knowledge as information in context. Knowledge transfer is also defined as "the conveyance of knowledge from one place, person, system or ownership to another" (Liyanage et al., 2009). Therefore, if AR is able to transfer information and put it into context, then it should be able to transfer knowledge to the users (Longo, Nicoletti and Padovano, 2019).

The idea of AR being a knowledge transfer technology is also confirmed by latest research in the area. Literature reviews in different application fields such as design and manufacturing (Bottani and Vignali, 2019), maintenance (Palmarini et al., 2018), surgery (Bernhardt et al., 2017), or education (Akçayir and Akçayir, 2017) have discussed research gaps regarding AR knowledge transfer abilities.

Besides, these gaps were always related with at least one content-related method: creation (Authoring), adaptation (Context-Awareness) or improvement (Interaction-Analysis) of augmented content. These methods are emerging AR research areas in their own:

- Authoring (A): software techniques that aim to create augmented content and properly display it in the real world (Zhu, Ong and Nee, 2015).
- Context-Awareness (CA): software techniques that aim to use contextual information to characterise augmented content (Zhu, Ong and Nee, 2014).
- Interaction-Analysis (IA): software techniques that analyse the status of the interaction between user and augmented content to provide relevant feedback and/or improve the interaction (Webel et al., 2013; Westerfield, Mitrovic and Billinghurst, 2015).

To the best of author's knowledge, no research has been found to review the state-of-the-art of these techniques altogether. Moreover, there is no research focused on clarifying their relation with AR knowledge transfer capabilities. Such research can involve an immense amount of work if all AR fields of application were to be considered. An intelligent strategy would be to narrow it down to an application where AR knowledge transfer capabilities can have a great impact.

Maintenance has a critical role improving organisations' competitiveness and contributing to their sustainable development (Rødseth, Schjølberg and Marhaug, 2017). The global-market size of high-value products maintenance-industries has been estimated in £490 billion by 2015 and £710 billion by 2025 (Mehta, 2015). High-value products are increasingly complex, technology intensive, expensive and critically reliable, requiring from continuous maintenance throughout their lifecycle (Roy et al., 2016). This leads to two of the main challenges that drive maintenance research (Rødseth, Schjølberg and Marhaug, 2017; Longo, Nicoletti and Padovano, 2019; Angelopoulos *et al.*, 2020):

1. Extend life of high-value products with optimum cost.
2. Improve efficiency and effectiveness of maintenance processes.

Due to high-value products' features, maintenance processes are knowledge intensive for maintainers (Candell, Karim and Söderholm, 2009; Dini and Dalle Mura, 2015). Some of their features are (Masood *et al.*, 2015; Roy *et al.*, 2016):

1. Number of equipment, subsystems and components implicates a large number of operations.
2. Complexity involves a large variety of different tasks from diagnosis to repair.
3. Long life causes varying levels of quality, standards and depth in documentation.

Therefore, the provision of the right information to the right user in the right quality and time is critical to increase efficiency of these maintenance processes (Parida and Kumar, 2004; Lee et al., 2008). As a visualisation technology, AR can provide support to maintainers with these knowledge-intensive challenges described above (Ong, Yuan and Nee, 2008). Authoring, Context-Awareness and Interaction-Analysis techniques are identified as important AR areas to enhance maintenance efficiency and effectiveness (Palmarini *et al.*, 2018; Bottani and Vignali, 2019):

- Authoring (A) to provide proper maintenance processes visualisation and so enhance their efficiency.
- Context-Awareness (CA) to adapt visualisation to the maintainer and so enhance their effectiveness.
- Interaction-Analysis (IA) to capture maintainers feedback and analyse their performance for enhancing visualisation and so improve their efficiency.

Therefore, A, CA and IA can help to enhance maintenance processes efficiency and effectiveness by providing an adaptive, increasing effective visualisation of maintenance processes to maintainers (Roy et al., 2016).

This thesis chapter aims to review the state-of-the-art of A, CA and IA research areas and establish a relation with AR knowledge transfer capabilities in the context of maintenance applications. A Systematic Literature Review (SLR) is conducted to achieve the following research objectives:

1. Identify the state-of-the-art of A, CA and IA techniques in AR maintenance applications.
2. Determine types and modes of knowledge transferred in AR maintenance applications.
3. Determine A, CA and IA research relations with knowledge transfer.
4. Identify A, CA and IA current challenges and potential future developments.
5. Identify new potential AR applications in knowledge transfer.

The rest of the chapter is organised in four sections. Section 2.2 describes the methodological approach for the SLR. Section 2.3 presents the results of the thematic analysis from the SLR. These results are used in Section 2.4 to provide an answer for the research questions. Then, Section 2.5 discusses the fulfilment of the research objectives. Finally, Section 2.6 reports this chapter's conclusions and proposes future research works.

2.2 Methodology

The research objectives identified in the previous section indicate the need to review existing literature. Besides, the specific research method to conduct this review was inspired by similar works in the field (Corrêa dos Santos, Delamaro and Nunes, 2013; Bacca et al., 2014; Palmarini et al., 2018). The comparison among these, suggested to conduct a SLR defined by Booth et al. (Booth, Papaioannou and Sutton, 2012) as a *"systematic, explicit, and reproducible method for identifying, evaluating, and synthesising the existing body of completed and recorded work made by researchers, scholars, and practitioners"*. In their book they also present the SALSA Framework (Grant and Booth, 2009), a methodology to determine research protocols for SLR's. The description of this methodology's steps is presented in Figure 2-1. It also includes the steps' outcomes and the research methods identified in this SLR to achieve them.

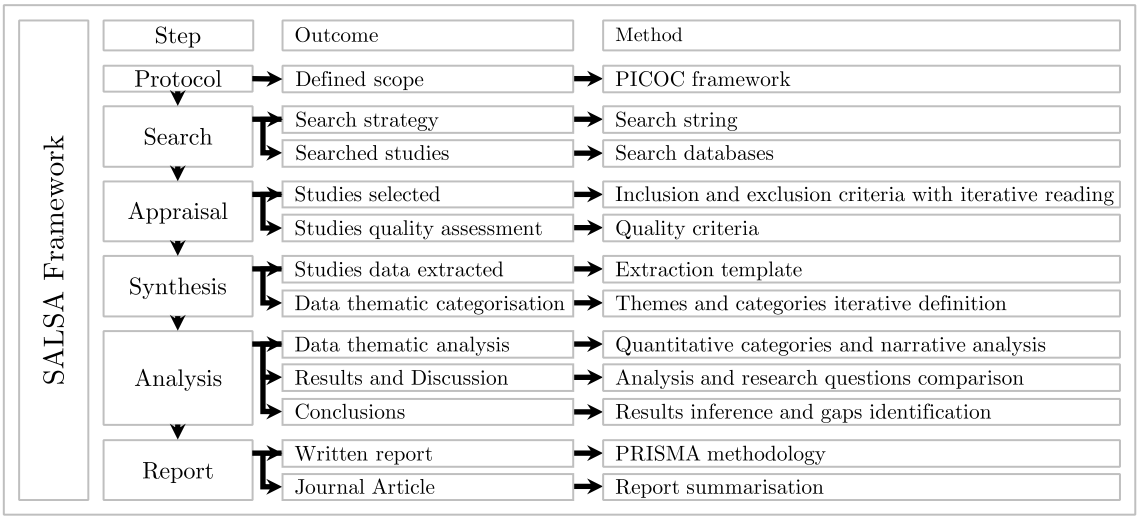


Figure 2-1. SLR methodology description: Includes steps (dark blue), outcomes (green) and the methods (purple) identified to achieve the outcomes inspired by Booth, Papaioannou and Sutton (2012).

Each step and the research methods to obtain their outcomes are explained in detail in the following subsections.

2.2.1 Protocol – SLR methodology step 1

The need for a research protocol for SLR's is identified by Booth Papaioannou and Sutton (2012) in the consideration of transparency, transferability and replicability, which are the characteristics that make a literature review systematic. The most critical stage in the protocol definition phase is determining the research scope. The scope helps to formulate answerable research questions and establish research boundaries to identify research methods for the SLR steps (see Figure 2-1). In order to determine the scope of this research, the PICOC framework has been used (Booth, Papaioannou and Sutton, 2012). PICOC (Population, Intervention, Comparison, Outcome and Context) is a formal structure to decompose research questions by their component concepts and therefore specifying the research's scope. Table 2-1 presents the application of the PICOC framework to this research objectives, along with the definitions for each concept.

Table 2-1. SLR research scope: obtained through the application of the PICOC framework to the SLR research objectives (Booth, Papaioannou and Sutton, 2012).

Concept	Definition	SLR application
Population	The problem or situation the research is dealing with.	Augmented reality for maintenance applications research: training, monitoring, inspection, diagnosis, repair, assembly, reporting, maintenance.
Intervention	Existing techniques utilised to address the problem identified.	Methods, tools and techniques for augmented content creation, adaptation, analysis and re-use: Authoring, Context-Awareness and interaction analysis.
Comparison	Techniques to contrast the intervention against.	Contrast between intervention techniques.
Outcome(s)	The measure to assess the effect of the techniques in the population.	Maintenance key performance indicators: completion time and errors, etc. Augmented content effectiveness: data quality, overlay accuracy, etc. Learning effects: task normalisation, equipment use, shared experiences, etc.
Context	The particular settings or areas of the population.	Maintenance of medium-long life complex assets.

The research scope identified in Table 2-1 will be used to determine the approach to consequent SLR steps. Besides, it also helps to refine this SLR's objectives. These are presented in the form of research questions in the list below:

1. What is the state-of-the-art in A, CA and IA techniques for AR maintenance applications?

2. What are the research gaps in A, CA and IA for AR maintenance applications?

3. What are the relations between A, CA and IA techniques and knowledge transfer?

4. What knowledge types are transferable by AR?

5. What potential applications of AR knowledge transfer are in maintenance contexts?

These are the research questions that will be answered by this SLR through the conduction of the steps described in the following subsections.

2.2.2 Search – SLR methodology step 2

The search phase consisted of identifying sources of information that could be relevant for this research. In order to do so, it was required to first identify where to find those sources and then retrieve them. Those two were the steps in which this phase was separated into: search strategy and delivery.

The search strategy step aimed to identify search databases and define the search string that can obtain relevant documentation on this systematic review's scope. The search string definition was defined using the terminology identified for the population in this scope (see Table 2-1). This scope was also used to identify relevant databases within the research area. First, the author created a list of search databases and engines considering those presented in similar works (Corrêa dos Santos, Delamaro and Nunes, 2013; Palmarini *et al.*, 2018). Then, those databases which did not allow to download references from the search results for further data processing were discarded. The resulted search databases/engines and search string are:

- Search databases: ACM, IEEE Xplore, ScienceDirect, Scopus, Web of Science.
- Search string: *("Augmented Reality") AND (Maintenance OR Assembly OR Repair OR Diagnosis OR Training OR Reporting OR Monitoring OR Inspection).*

The search delivery step involved the application of the search string to the databases in order to retrieve related academic papers as search results. These results are classified by database as summarised in Table 2-2.

Table 2-2. SLR search: search delivery results classified by database.

Search String[1]	Database	Date	Papers
"augmented reality" AND (maintenance OR assembly OR repair OR diagnosis OR training OR reporting OR monitoring OR inspection)	ACM	30/05/17	1712
(“augmented reality”) AND (maintenance OR assembly OR repair OR diagnosis OR training OR reporting OR monitoring OR inspection)	IEEE Xplore	31/05/17	1687
("augmented reality" AND (maintenance OR assembly OR repair OR diagnosis OR training OR reporting OR monitoring OR inspection))	ScienceDirect	13/06/17	3835
TITLE-ABS-KEY ("augmented reality" AND (maintenance OR assembly OR repair OR diagnosis OR training OR reporting OR monitoring OR inspection))	Scopus	03/06/17	2485
TS = ("augmented reality" AND (maintenance OR assembly OR repair OR diagnosis OR training OR reporting OR monitoring OR inspection))	WOS	12/06/17	2174
		Total	11893

Several observations according to the SLR search phase are made:

- The number of results was quite high, due to the length of the search string.
- As the searches in each database were independent, the results included duplicates.
- The results also included papers from different applicable contexts.

Hence, strict inclusion and exclusion criteria was needed to narrow down the results in the study selection process.

2.2.3 Appraisal – SLR methodology step 3

The appraisal phase comprised the evaluation of search results in order to select those papers that are relevant according to this research scope and describe their validity. These two objectives were achieved in two different steps: study selection and quality assessment.

[1] To be used within advance search queries of the database (link)

The study selection consisted of the screening of search results for selecting those papers that were relevant according to this review. In order to make this process as systematic and repeatable as possible, the author used a set of predetermined inclusion and exclusion criteria. The author defined these criteria to be relevant according to the review's scope (see Table 2-3) and inspired by the structure of the selection criteria presented in similar reviews (Corrêa dos Santos, Delamaro and Nunes, 2013; Akçayir and Akçayir, 2017; Palmarini *et al.*, 2018).

Table 2-3. SLR appraisal: study selection inclusion and exclusion criteria.

Type	ID	Statement
Inclusion	I1	Application-research studies that present evidences of Authoring
Inclusion	I2	Application-research studies that present evidences of Context-Awareness
Inclusion	I3	Application-research studies that present evidences of interaction analysis
Exclusion	E1	Papers not written in English
Exclusion	E2	Papers that were published before 2012
Exclusion	E3	Papers that are duplicated within the search documents
Exclusion	E4	Papers that do not meet any of the inclusion criteria
Exclusion	E5	Papers whose evidence is not applicable to maintenance of medium-long life complex assets
Exclusion	E6	Papers that are not primary research
Exclusion	E7	Papers that are not journal papers
Exclusion	E8	Papers that are not accessible

Table 2-3 presents the inclusion and exclusion criteria utilised. While some criterions are directly related to the interactions and context of this research scope (I1, I2, I3, E4, E5), others consider research-community focus in these matters (E2), and others focus on research quality and validity (E6, E7) and data accessibility (E1, E3, E8). Further discussions on the criteria definition can be found in Subsection 2.5.1.

These criteria were applied for screening a total 11893 papers. In order to increase screening efficiency, experts suggest to apply the criteria iteratively at different reading phases (Moher et al., 2010; Booth, Papaioannou and Sutton, 2012). The screening process and the resultant included and excluded papers are presented in Figure 2-2. After removing old (E2) and duplicated (E3) papers, exclusion criteria (E1-E8) was applied for reviewing papers' title (5127), abstracts (1959), and introduction and conclusion sections (681). Then, inclusion and exclusion criteria were applied when fully reading the papers (92) to classify them by the evidence (I1, I2 or I3) shown. The process resulted on the selection of a sample of 74 relevant papers that complies with the selection criteria.

Reading step	*Criterions applied*	*Total*	*Included*	*Excluded*
Database search results	Search string	11893	11893	0
Remove papers before 2012	E2	11893	6874	5019
Remove duplicated papers	E3	6874	5127	1747
Title reading	E1 to E8	5127	1959	3168
Abstract reading	E1 to E8	1959	681	1278
Intro.-conclusions reading	E1 to E8	681	92	589
Full-text reading	I1 to I3 and E1 to E8	92	74	18

Figure 2-2. SLR appraisal: total, included and excluded number of papers presented by reading steps according to the criteria applied on each step.

The quality assessment step comprised the evaluation of the internal (research methods) and external (results) validity of selected-relevant papers (Booth, Papaioannou and Sutton, 2012). Frameworks for systematic reviews, such as SALSA (Booth, Papaioannou and Sutton, 2012) or PRISMA (Moher et al., 2010), include this step to "localise how weaknesses or flaws of included studies may impact upon the review's findings" (Booth, Papaioannou and Sutton, 2012). In case any biases or contradictory ideas appear when analysing selected-relevant papers, quality assessment's results could be used to provide more transparent, and repeatable findings. The assessment has been conducted following the validity criteria described by similar works (Corrêa dos Santos, Delamaro and Nunes, 2013; Palmarini et al., 2018) that is presented in Table 2-4. The total 74 selected-relevant papers were evaluated against each criterion with a score from 0 (no compliance), 0.5 (partial compliance) to 1 (full compliance).

Table 2-4. SLR appraisal: quality assessment criteria, scores and statistical results classified by validity aspect.

id	Criterion	Score	Mean	S.D.
	Internal validity	0-6	3.82	1.06
1	Appropriateness and clarity of research objectives	0-1	0.86	0.25
2	Appropriateness and clarity of research methodology	0-1	0.55	0.32
3	Appropriateness and clarity of research process	0-1	0.64	0.33
4	Appropriateness and clarity of data support of analysis	0-1	0.63	0.36
5	Appropriateness and clarity of analysis methods	0-1	0.48	0.38
6	Appropriateness and clarity of conclusions	0-1	0.67	0.31
	External validity	0-6	3.71	0.93
a	Evidence for A, C-A and I-A	0-1	0.76	0.26
b	Appropriateness and clarity of system architecture	0-1	0.68	0.38
c	Appropriateness and clarity of knowledge involved	0-1	0.81	0.26
d	Case studies and applications not obsolete	0-1	0.82	0.32
e	Results applicable to maintenance of complex assets	0-1	0.64	0.39

Table 2-4 also presents the average statistical results (mean and standard deviation) for each quality criterion of all 74 papers assessed. It is important to note that the author did not have the need to use these results further in the review. This was because no biases or contradictions were found in the analysis of the 74 selected-relevant papers. Still, the results are presented to provide the reader with a tool to assess the quality of this systematic review's findings. The application of validity criteria for quality assessment is a process subjected to biases itself. Therefore, while the average mean (see Table 2-4) represents the quality of the papers assessed, the standard deviation for each criterion provides a numerical valuation on the potential author's bias in the assessment process.

2.2.4 Synthesis – SLR methodology step 4

The synthesis phase consisted of the extraction and classification of relevant data from selected papers in order to map the evidence base for its analysis. Among others (e.g. logic models, Bayesian meta-synthesis), thematic synthesis was selected as research method for this phase as suggested by Booth et. al (Booth, Papaioannou and Sutton, 2012). The reasons to make this choice were two. First, it is a mature method for synthesising qualitative data. And second, themes could be directly identified according to the research scope (see Table 2-1). This method consisted of two steps: data extraction and thematic categorisation.

The data extraction step involved identification and extraction of relevant data in the 74 selected papers according to the themes identified in the research scope (Booth, Papaioannou and Sutton, 2012). Table 2-6 presents their definitions. The first four themes (Asset, Operation, Task and Knowledge) provide a description of the maintenance operation and the AR application. The last three themes (i.e. A, CA and IA) detail the techniques used to provide knowledge transfer. The data related to each theme of each paper was extracted into an Excel sheet for data processing as shown in the example in Table 2-5.

Table 2-5. SLR synthesis: thematic data extraction example from two papers.

Theme	Paper 1	…	Paper 3
Asset	thermal power plant	…	construction
Operation	periodical and repairing maintenance	…	progress monitoring and documentation
Task	occupational safety and work instructions	…	inspection, survey and annotation of construction progress
Knowledge	procedural	…	procedural
A	knowledge-based rules, manual-content	…	knowledge-based algorithms, automatic-content
C-A	user-experience	…	aerial images, georeferenced positions
I-A	none	…	none

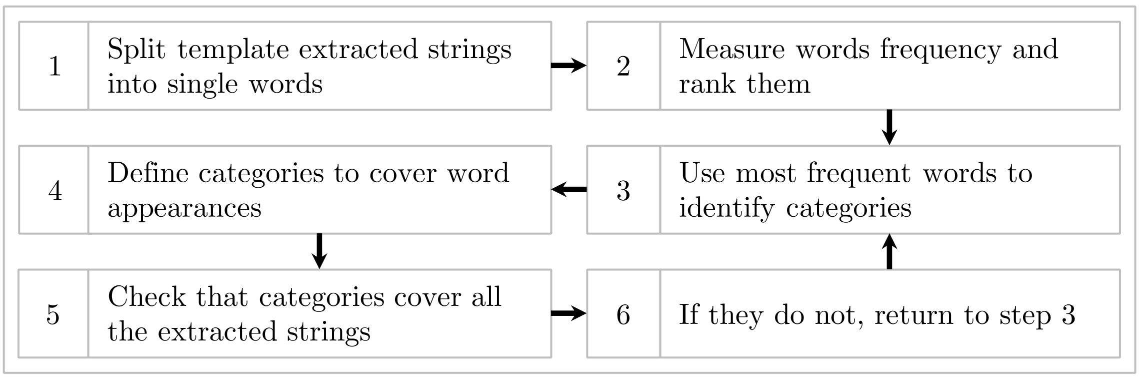


Figure 2-3. SLR synthesis: definition of steps included in the word-counting, iterative method used for categories identification in data synthesis.

The thematic categorisation step included the classification and processing of the data extracted to prepare it for further analysis. In order to provide quantitative evidence about qualitative data, several categories for each theme are determined according to the data extracted. These categories comprise the different groups in which the ideas presented by the author about the themes can be classified. A word-counting, iterative method (see Figure 2-3) was proposed to identify categories based on the qualitative data extracted. The resultant categories are presented in Table 2-6 along with their themes.

Table 2-6. SLR synthesis: definition of themes used for data extraction and identified categories in data synthesis.

Theme	Definition	Category	Definition
Asset	The size of the augmented objects the AR application is targeting. The augmented object size is directly related with the type of assets considered within different industries (e.g. civil construction, manufacturing, etc.).	Small	The size of the object is similar to that of a device, product, equipment, etc.
		Medium	The size of the object is similar to that of a plant, construction, infrastructure, etc.
		Large	The size of the object is similar to that of an open space, environment, city, etc.
Operation	The maintenance processes the AR application described by the authors is supporting, e.g. diagnosis, repair, design, training.	Design	Any process related to the design of a maintenance operation.
		Assembly	Any process related to objects' assembly or disassembly within maintenance operations.
		Diagnosis	Any process related to the diagnosis of malfunctioning objects within maintenance operations.
		Repair	Any process related to objects' repair or replacement within maintenance operations.
		Training	Any process related to the training of a maintenance operation.
		Management	Any process related to the management of maintenance operations.
Task	The function that the augmented content described in the paper is conducting, e.g. monitoring, guidance, etc. The task directly relates the AR support given with the maintenance operation assisted.	Monitoring	Any AR function related to the observation and control of objects.
		Guidance	Any AR function related to the guidance of the user.
		Collaboration	Any AR function related to the collaboration among different users.

		Simulation	Any AR function related to the simulation of a real processes of objects.
Knowledge	The type of knowledge the AR application described in the paper is transferring and/or capturing (e.g. procedural, tacit, explicit).	Procedural	The knowledge related to the performance or conduction of a specific procedure.
		Declarative	The knowledge related to a specific fact or situation and the relations between them.
Authoring	The set of software methods, tools and techniques to create augmented contents for the AR applications described in the papers.	Users	Consumers of Authoring tools, who can be either AR developers or subject-matter experts.
		Rules	The logic under which augmented content is created. These can be either specific-application algorithms or domain knowledge-structures (when the knowledge about a specific domain can be articulated, and so the information related directly identified).
		Automation	When the Authoring tools, can create the content automatically, semi-automatically (users-approved) or manually (users-made).
Context-Awareness	The set of software methods, tools and techniques to characterise the augmented content with relevant data regarding the task of the AR application described in the paper.	Contexts	The set of contexts the methods consider, it can be one or more of the followings: user, augmented object, activity, environment.
		Rules	The set of rules that determine how the context(s) are considered. These can be either specific-application algorithms or domain knowledge-structures.
Interaction analysis	The set of software methods, tools and techniques to analyse the interaction between the user and the augmented content, regarding the task status, in order to provide feedback for improving task's achievement.	Data	Interaction analysis tools can acquire data either automatically (sensors) or manually (users feedback).
		Automation	Whether the analysis made by these tools is automatic or manual (made by users).

The thematic categorisation method was proposed because most frequent words are more probable to be identified as categories (Booth, Papaioannou and Sutton, 2012), as they cover the meaning of similar approaches to a specific theme. Only relevant words (excluding those not related to the themes) from the most frequents were used to find specific definitions for the categories (steps 4, 5 and 6 in Figure 2-3). This process required few iterations for obtaining consistent and coherent definitions for the categories.

2.2.5 Analysis – SLR methodology step 5

The analysis phase comprised the evaluation of synthesised data and the extraction of conclusions for providing enough findings to answer the SLR research questions (Subsection 2.2.1). It consists of three steps: (i) independent analysis of themes (i.e. thematic analysis), (ii) discussion of analysis results to answer research questions (i.e. results discussion) and (iii) extraction of conclusions (i.e. conclusions drawn).

The thematic analysis involved the assessment of quantitative results and qualitative data extracted during the synthesis phase. Its aim was to map how each theme was covered among the selected papers and find the relations between themes. Each subsection in Section 2.3 presents the narrative analysis of the data extracted for each theme, along with the quantitative figures about the demographic description of categories between selected-relevant papers.

The results discussion step included the evaluation of the evidence-base given by the analyses in the previous step. It allowed to discuss and answer the research questions through the relations between themes. In order to simplify the narrative of this chapter, Section 2.4 briefly presents the answer to the research as results. Then, the discussion about those results is detailed in Section 2.5.

Finally, the conclusions drawn step incorporated inferring insights from the previous discussions. As a result, research conclusions and future works are presented in Section 2.6.

2.2.6 Report – SLR methodology step 6

The report step involved the description and presentation of the methods and results on the research conducted. Three tasks were determined for this step: (i) description of the main elements in a SLR under a standard form (PRISMA report), (ii) research summary for public presentation (Journal article) and (iii) extended description from doctoral thesis presentation. The PRISMA report step comprised the presentation of the methods and results of this SLR as well as the description of any step in the process using a standardised method called PRISMA methodology (Moher et al., 2010). This step resulted in the creation of a more detailed report utilised to write a journal article. It comprised the presentation of the research methods and results in a more comprehensive and detailed manner. So, that this research can be publicly available for scientific purposes within the scope of the doctoral research.

2.3 Analysis

The thematic analysis consisted of the assessment of the research included in the 74 selected-relevant papers according to the themes, categories (Table 2-6) and data extracted in the synthesis phase. In order to provide the context on the papers' research, demographics results about them are presented. Table 2-7 shows the top 10 publications with more selected-relevant papers. The maximum number of selected-relevant papers from one publication is five. This represents a 7% of the total, which is not enough percentage to consider it as a reference publication in the research this SLR involved.

Table 2-7. SLR analysis – Demographic description: top 10 journals with most publications within the 74 selected-relevant papers.

Publication	Papers
Automation in Construction	5
CIRP Annals - Manufacturing Technology	5
Computers in Industry	5
Augmented Reality in Architecture, Engineering, and Construction	3
International Journal of Advanced Manufacturing Technology	3
International Journal on Interactive Design and Manufacturing	3
Robotics and Computer-Integrated Manufacturing	3
Assembly Automation	2
Personal Ubiquitous Computing	2
Proceedings of the IEEE	2
Others	41
Total	74

Further categorisation and analysis are required to map the research in the selected-relevant papers. According to the research scope, the synthesis themes and categories will be used to drive the analysis and describe the research. So, this analysis can provide enough evidence base to answer the research questions. The following subsections provide an analysis (Subsection 2.2.5) of each theme and its categories (see Table 2-6). First, selected-relevant papers are examined according to the 'assets' (Subsection 2.3.1) the AR applications provide support to. Second, they are analysed according to the maintenance 'operations' supported and how this support is provided (Subsection 2.3.2). Then, the focus is set on how the support 'tasks' relate with AR knowledge transfer capabilities (Subsection 2.3.3) and the 'knowledge' types involved (Subsection 2.3.4). Finally, the Authoring (Subsection 2.3.5), Context-Awareness (Subsection 2.3.6) and Interaction-Analysis (Subsection 2.3.7) techniques are independently analysed, classified and their relation with AR knowledge transfer capabilities assessed.

2.3.1 Asset

As defined in the thematic categorisation (Table 2-6), by ‘asset’ the author considers the size of the objects targeted by the AR applications described in the selected relevant papers. These assets can be ‘small’ (e.g. equipment, devices, etc.), ‘medium’ (e.g. infrastructure, buildings, etc.) and ‘large’ (e.g. open spaces, cities, etc.). Figure 2-4 presents the percentages of the 74 selected-relevant papers that target the three different sizes of ‘Assets’.

According to Figure 2-4, most of the selected-relevant papers (68%) have proposed AR-maintenance applications for ‘small’ assets (e.g. machine tools, motherboards, etc.). Only a few of these papers (5%) have focused in ‘large’ assets such as rivers or cities. The ‘Asset’ size affects the kind of maintenance operations that are considered by the AR applications in the papers. That is the reason why the 74 selected-relevant are described in the following subsection by their ‘Operations’ according to the size of their ‘Asset’.

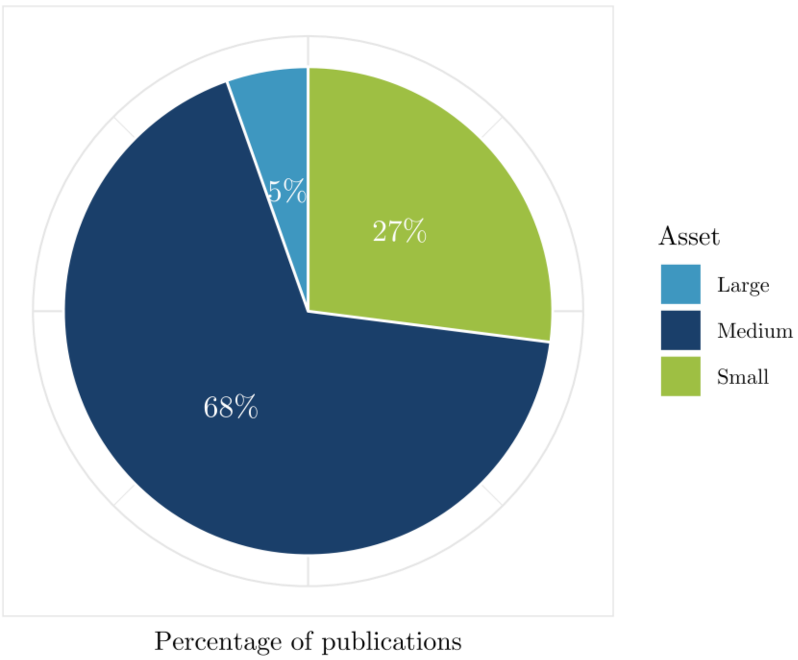


Figure 2-4. SLR analysis – ‘Asset’ analysis: classification of 74 selected-relevant papers by type of ‘Asset’ targeted by their AR applications.

2.3.2 Operation

By 'Operation', the author refers to the maintenance process that the AR application is supporting (Table 2-6). Figure 2-5 presents the papers classified by maintenance 'operation': 'design', 'assembly', 'diagnosis', 'repair', 'training' and 'management'. Compared to other reviews with similar categories (Palmarini et al., 2018), few operations such as 'design' and 'management' were added. This was due to the method (Subsection 2.2.4) that used paper data to create categories, whose validity was endorsed by results' similarities with other reviews.

The analysis showed that AR applications were similar for each 'operation', but differed from 'asset' to 'asset'. So, it was worthy to analyse 'operations' based on their 'asset'. Also, not all 'operations' have AR applications for every 'asset' and not all 'operations' within each 'asset' have a similar amount of relevant papers for all the years included. This fact can offer valuable insights about the maturity of AR applications in specific 'operations' and 'assets'. Besides, it can also point to differences in the A, CA and IA techniques and their requirements by 'asset' and/or 'operation'. Following subsections describe AR applications according to the maintenance 'operations' they support for 'large', 'medium' and 'small' assets.

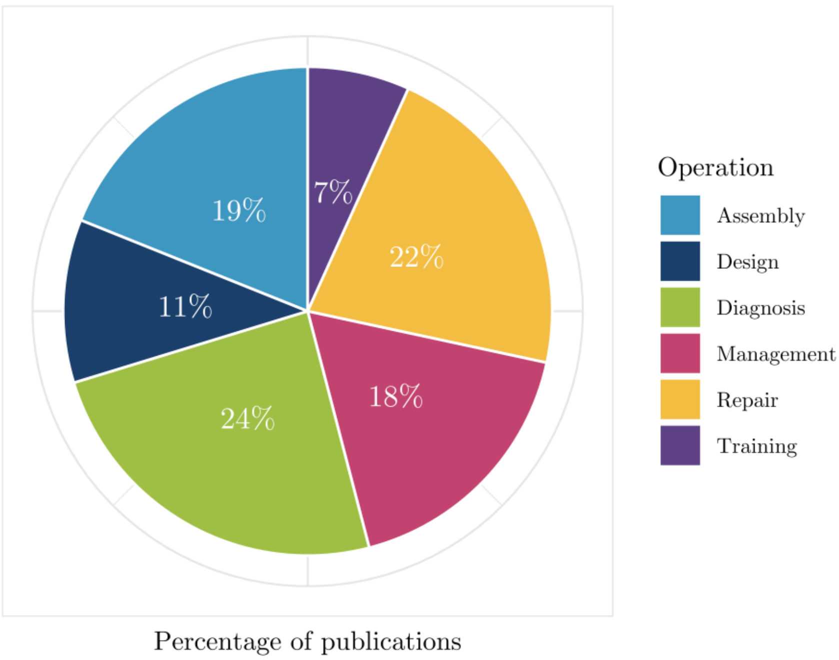


Figure 2-5. SLR analysis – 'Operation' analysis: classification of 74 selected-relevant papers by type of 'Operation' supported by their AR applications.

2.3.2.1 Operations in 'Large' Assets

The selected-relevant papers classified by 'operation' within the 'large' asset category are presented in Figure 2-6. Only 4 papers were found relevant under this category at the study selection process. These were the only papers to consider the AR research areas of Authoring, Context-Awareness and Interaction Analysis. Due to the AR technological challenges in outdoor applications, it appears that research in AR for 'Large' Assets is not mature enough (Veas et al., 2013) to consider yet the broad application of these techniques.

Sebillo et. al (2016) presented a training application for emergency responders that aimed to enhance their effectiveness with emergency technologies and procedures. Apart from an Authoring tool that was able to display location-context information regarding the user's tasks during the training, this application also had an IA tool that enabled trainers to automatically collect data from trainees and adapt their training tasks on real-time.

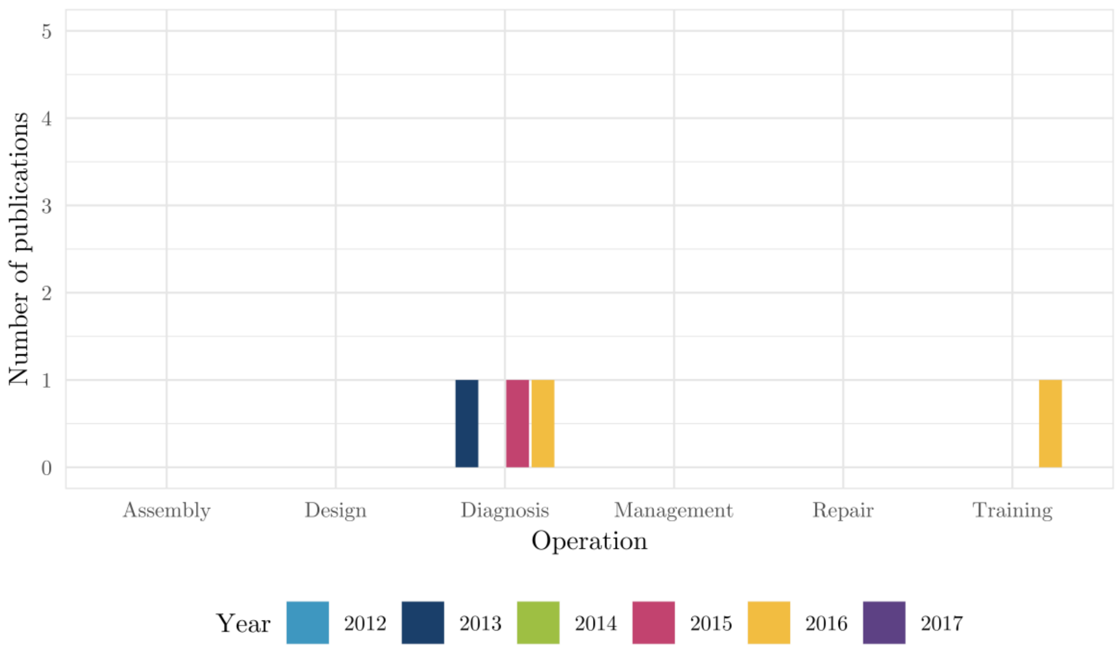


Figure 2-6. SLR analysis – 'Operation' analysis: classification of 4 selected-relevant papers by type of 'operation' in 'large' asset.

On the other side, the three papers describing ‘diagnosis’ operations were focused on the monitoring of different environments. First, Veas et. al (2013) presented an automatic Authoring tool to generate environmental sensor measures content on augmented environments to enhance data visualisation and interaction of domain experts. Pokric et. al (2015) described an Authoring tool for creating pollution monitoring applications in smart cities for citizen participation. This tool allows experts to create content using serious gaming concepts based on the data acquired from the monitoring IoT (Internet of Things) devices. Pierdicca et. al (2016) presented an AR application for riverbanks maintenance. It included an Authoring tool based on a standard data layer that considered all the tasks to be done by the environmental inspectors. According to the task selected by the inspector, different content automatically generated was displayed over the riverbanks to enhance inspection efficiency.

2.3.2.2 Operations in ‘Medium’ Assets

The selected-relevant papers targeting ‘medium’ assets classified by ‘operations’ are presented in Figure 2-7. There is a total of 20 papers, mainly distributed in the earlier years considered within this review. Only one is from 2017 while 16 of them are from 2012, 2013 and 2014. Only two papers – one in ‘assembly’ (tele-operated cranes support) (Chi *et al.*, 2012) and another in ‘training’ operations (escape guidance for radioactive accidents) (Tsai and Yau, 2013) – are not from ‘diagnosis’ and ‘management’ operations.

In ‘diagnosis’ operations, the reviewed papers between 2012 and 2017 are focused on three topics. The first one involves monitoring of different defects like segment displacement in tunnels (Zhou, Luo and Yang, 2017), underground manholes (Yang *et al.*, 2015) and building damage reconnaissance (Dong, Feng and Kamat, 2013). The second one relates to monitoring of building power consumption data for energy performance (Ham and Golparvar-Fard, 2013; Golparvar-Fard and Ham, 2014; Chou *et al.*, 2016). Finally, the third includes construction site progress monitoring (Zollmann, Hoppe and Kluckner, 2014). All these AR applications present automatic Authoring tools based on physic models and BIM data. Besides, some present additional Interaction-Analysis techniques to support users in their evaluation and inspection tasks.

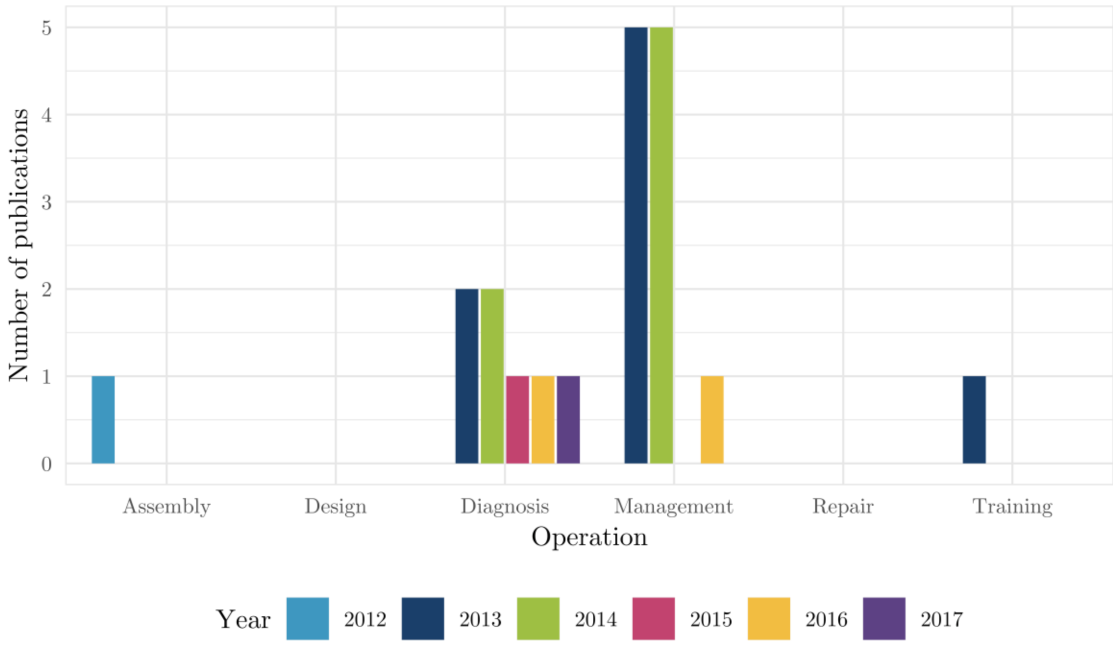


Figure 2-7. SLR analysis – 'Operation' analysis: classification of 20 selected-relevant papers by type of 'operation' in 'medium' asset.

In 'management' operations, the focus of papers is different. There are some case-specific applications about construction-site defect management (Park *et al.*, 2013; Kwon, Park and Lim, 2014), library management (Shatte, Holdsworth and Lee, 2014), or air-traffic management (Rohacs, Rohacs and Jankovics, 2016). Nevertheless, most of them are focused on project management for the construction site. The Architecture, Engineering and Construction (AEC) industry is advanced in the representation of knowledge for project management (Chi, Kang and Wang, 2013). And Building Information Modelling (BIM) is a mature technology in this area. Therefore, most of these AR applications are focused on Authoring tools for BIM data visualisation. Different approaches have been taken in different contexts. For example, the usage of GIS (Geographic Information System) data for underground infrastructure (Schall, Zollmann and Reitmayr, 2013) or Photogrammetry in the facility management of oil refineries (Hou *et al.*, 2014). BIM data visualisation using AR can be considered a mature application. This is because there already exist different Authoring tools for automatically creating and augmenting BIM data (Irizarry *et al.*, 2013; Jiao *et al.*, 2013; Kim *et al.*, 2013; Javier *et al.*, 2014; Meža, Turk and Dolenc, 2014). Nevertheless, there is less reported evidence on Context-Awareness and Interaction-Analysis techniques. Only the case-specific applications above-mentioned considered them. Besides, the BIM data considered in the Authoring application papers do not mention neither contextual nor task- and user-status data that could trigger the development of these techniques.

2.3.2.3 Operations in 'Small' Assets

The 'small' assets category is the most relevant in number of selected-relevant papers (50). Figure 2-8 presents these papers classified by different 'operation' categories. 'Small' asset is the only category with representation in all 'operation' categories. The papers within each 'operation' category are described below in ascending order of number of papers.

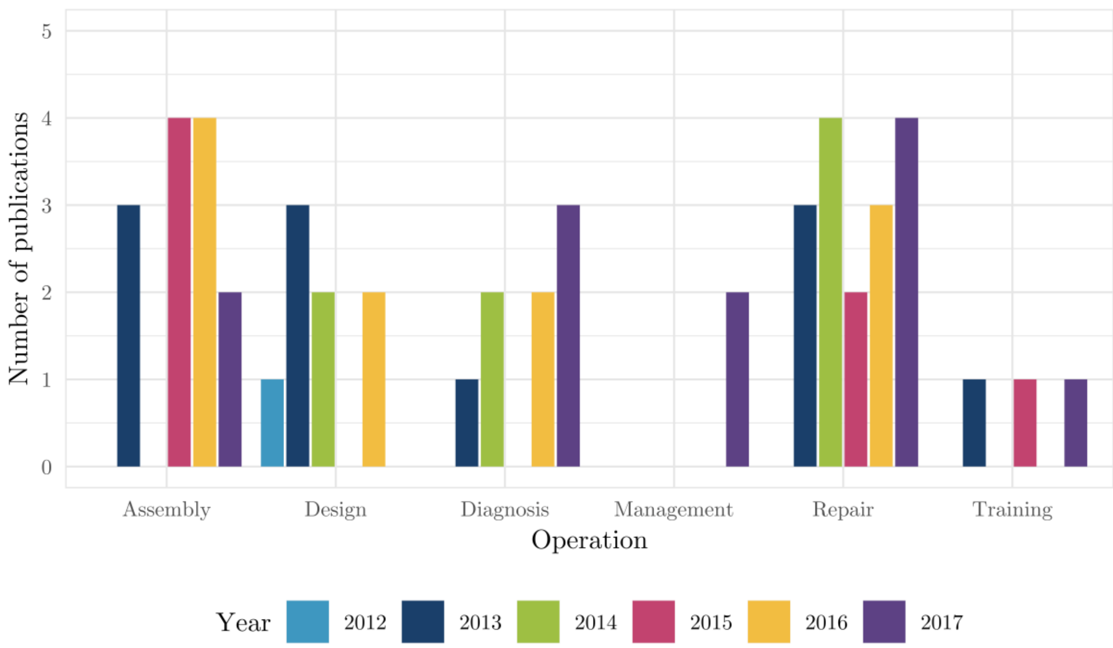


Figure 2-8. SLR analysis – 'Operation' analysis: classification of 50 selected-relevant papers by type of 'operation' in 'small' asset.

In the case of 'management' operations there are only 2 selected-relevant papers. Suarez-Warden and González Mendívil (2017) proposed an AR-based method to transfer procedural knowledge in the aerospace service industry. Differently, Liu et. al (2017) proposed an AR interface for cyber-physical machine tools that allowed the user to control machining processes in real time. This paper presented evidence on all A, CA and IA techniques. An Authoring tool that automatically generates the simulation of the machining process, a Context-Aware technique to adapt the augmented simulation based on monitoring data, and an Interaction-Analysis approach to evaluate the feasibility of the process.

'Training' operations are similar to 'management' regarding their use of A, CA and IA techniques. The use of expert systems to support traditional AR training is proposed by Westerfield, Mitrovic and Billinghurst (2013) to enhance learning processes rather than just training procedural skills. In contrast, Webel et. al (2013) proposed the use of haptic feedback based on training status to enhance the practice on the undergoing skills of procedural works. On a different application level, Okazaki and Takaseki (2017) proposed an AR-based training simulator for maritime navigation, which only included evidence of Authoring methods. Meanwhile, the other researches on training of procedural works also included research on Context-Awareness and Interaction-Analysis.

Within 'design' operations, one of the most relevant fields is the design, planning and simulation of assembly procedures. Different papers have shown different Authoring tools for automatically creating assembly procedures based on real and virtual components (Wang et al., 2013; Liu, Li and Wang, 2014). Based on visualising assembly procedures, user-virtual components-interaction capabilities have been included in following works, in order to enhance design and planning effectiveness. That is the case of Wang, Ong and Nee (2013); they proposed a bare-hand interface to contextualise the manipulation of virtual components. Moreover, they extended their own work with Interaction-Analysis methods to evaluate the status of assembly situations based on user interaction and provide more accurate assembly simulations (Ng et al., 2013; X Wang, Ong and Nee, 2016b). Additionally, other design applications have been proposed in the context of ‘small’ assets. One was an Authoring tool for robot trajectory planning and simulation (Fang, Ong and Nee, 2012). Another was a different Authoring tool for collaborative design through tele-presence systems (Wang *et al.*, 2014). The latest allowed users to create and interact with their own-created and other users-created content in the same virtual space. Also, a context-based approach was proposed by Laroche et al. (2016) to access enterprise knowledge in design processes through AR. Nevertheless, there was not that much research and so, only independent Authoring and Context-Awareness techniques have been considered within selected relevant papers in ‘design’ operations in ‘small’ assets.

Figure 2-8 shows an increasing interest in ‘diagnosis’ operations due to the increase on published papers over the last years in this area. Nevertheless, the specificity of inspection and monitoring processes in most of the cases has narrowed the research. Authors have been able to investigate together either Authoring and Context-Awareness or Authoring and Interaction-Analysis to provide further insights within AR applications about the actual processes. Some examples are: post-impact inspection of thin structures (De Marchi et al., 2013), crack growth monitoring in bonded single-lap joints (Bernasconi et al., 2014), ‘Situated Analytics’ for health product evaluation (ElSayed et al., 2016), strain and stress visualisation for mechanical systems (Naets, Cosco and Desmet, 2017), finite element analysis visualisation and interaction (Huang, Ong and Nee, 2017) and real-time monitoring by comparison of 3D printing processes (Ceruti, Liverani and Bombardi, 2017). Conversely, inspection applications on less specific domains have only considered Authoring tools for automatic content creation. That is the case of AR applications for support of maintainers inspection procedural tasks (Wójcicki, 2014) or visualisation support of acceptance sampling procedures (Franceschini et al., 2016).

‘Assembly’ is one of the most famous fields of AR application. Thirteen (13) relevant papers have been identified and they mostly cover visualisation support of assembly procedures. The fame of assembly as AR application has led to an extensive representation of assembly knowledge. Moreover, it has also revealed the latest advancements in Context-Awareness and Interaction-Analysis techniques. The AR applications studied are able to achieve knowledge capture and discovery of procedural assembly expertise. Initially, research was focused on the automatic content generation of assembly procedures based on existing virtual data for human users (Makris et al., 2013; Radkowski, Herrema and Oliver, 2015) or for programming robots without programming skills (Lakshantha and Egerton, 2016; Reina et al., 2017). Then, Context-Awareness methods have been introduced to provide more effective support based on user's experience and assembly status (Ong and Zhu, 2013; X Wang, Ong and Nee, 2016a; Yew, Ong and Nee, 2016). Based on determining assembly status, Interaction-Analysis methods have been used to capture assembly expertise, evaluate it and transfer it to robots (Lambrecht et al., 2013; Makris et al., 2016; Ramirez-Amaro, Beetz and Cheng, 2017) or other human users (Liu et al., 2014; Bleser et al., 2015; Wang, Ong and Nee, 2015). Nevertheless, the introduction of different areas of research (CA and IA) has been progressive and based on the advancements of the previous (Authoring). Contextualisation could not be provided until there was automatic augmented content. Then, based on the insights in task status and user experience given by Context-Awareness methods, interaction analysis techniques for assembly evaluation and optimisation could be introduced.

'Repair' is another of the most famous and oldest fields of AR application. It comprises repair and replacement procedures in corrective or preventive maintenance. These have the need to enhance their efficiency by better visualising existing procedural documentation. Based on the sixteen relevant papers selected within this operation, the same pattern as in 'small' assets 'assembly' or 'large' assets 'management' operations was identified. First, Authoring tools are developed for different applications such as machine tools (Gimeno *et al.*, 2013), aircrafts (Golański, Perz-Osowska and Szczekala, 2014), consumer products (Shaaban, Che Mat and Mahayudin, 2015), hazardous environments (Martínez et al., 2014), or industrial-like environments (Fiorentino et al., 2014, 2016; Re, Oliver and Bordegoni, 2016). Then, Authoring automation and Context-Awareness techniques are developed once the areas of application are better investigated. That is the case of several systems such as CARMMI (Espíndola et al., 2013) (integrating existing information from different sources) or ACARS (Zhu, Ong and Nee, 2013), COARS (Zhu, Ong and Nee, 2014) and ACAAR (Zhu, Ong and Nee, 2015). In the latest three, the same authors first developed a context-aware method, and then included Authoring capabilities, first for programming experts and then for non-programmers. Also ARAUM (Erkoyuncu *et al.*, 2017) considered methods to adapt content and interactions to users' expertise. Context-Awareness techniques have also been used to enhance safety in such difficult environments (Alam et al., 2017; Tatic and Tešic, 2017). Following the same pattern, Interaction-Analysis techniques are introduced after to track the performance of users and status of repairing processes. That was the case of Nakai and Suzuki (2016) for faster procedures in chemical plants. Also from Mourtzis, Vlachou and Zogopoulos (2017b), who were the latest authors to present a new approach for machine tools servitisation. This was based on remote AR-supported maintenance that enable knowledge transfer from remote experts to on-site technicians.

2.3.2.4 Findings summary

From the narrative analysis that supports the numerical results presented in Subsections 2.3.2.1, 2.3.2.2 and 2.3.2.3, their key findings are summarised below:

- There appears to be a relation between knowledge representations in an 'operation' domain and the relevancy of Authoring, Context-Awareness and Interaction-Analysis research in those 'operations'.
- There appears to be certain 'operations' within certain 'industries' where more generic knowledge-domain representations have been studied (e.g. equipment repair or construction management). Instead, there are others in which these representations are more limited.
- Those 'operations' with more limited knowledge-domain representations do not present as much evidence in Authoring, Context-Awareness and Interaction-Analysis as those with more generic ones.
- Generality of knowledge-domain representations can be seen as an indicator of knowledge transfer effectiveness. The easier to represent knowledge, the easier to transfer it effectively.

Therefore, it seems reasonable to identify a relation between knowledge transfer and Authoring, Context-Awareness and Interaction-Analysis techniques.

2.3.3 Task

‘Task’ has been defined (Table 2-6) as the support function that augmented content provides to AR users. While ‘operations’ refer to the labours associated to the ‘asset’ (e.g. repair, assemble, design, etc.), ‘tasks’ relate to the AR information delivery that support those. Therefore, ‘tasks’ can be considered as the methods AR enables to transfer knowledge (“information in context” (Nonaka, 1994)). Four non-mutually exclusive categories have been declared for ‘tasks’ (Table 2-6): ‘monitoring’, ‘guidance’, ‘simulation’ and ‘collaborative’. Good explanations from each category can be found in (Wang *et al.*, 2014; Zollmann, Hoppe and Kluckner, 2014; Naets, Cosco and Desmet, 2017; Tatic and Tešic, 2017) respectively. Besides, it seemed interesting to analyse the relation between ‘operations’ and ‘tasks’ (Figure 2-9) and ‘tasks’ and A, CA and IA techniques (Figure 2-11) to further understand AR knowledge transfer capabilities.

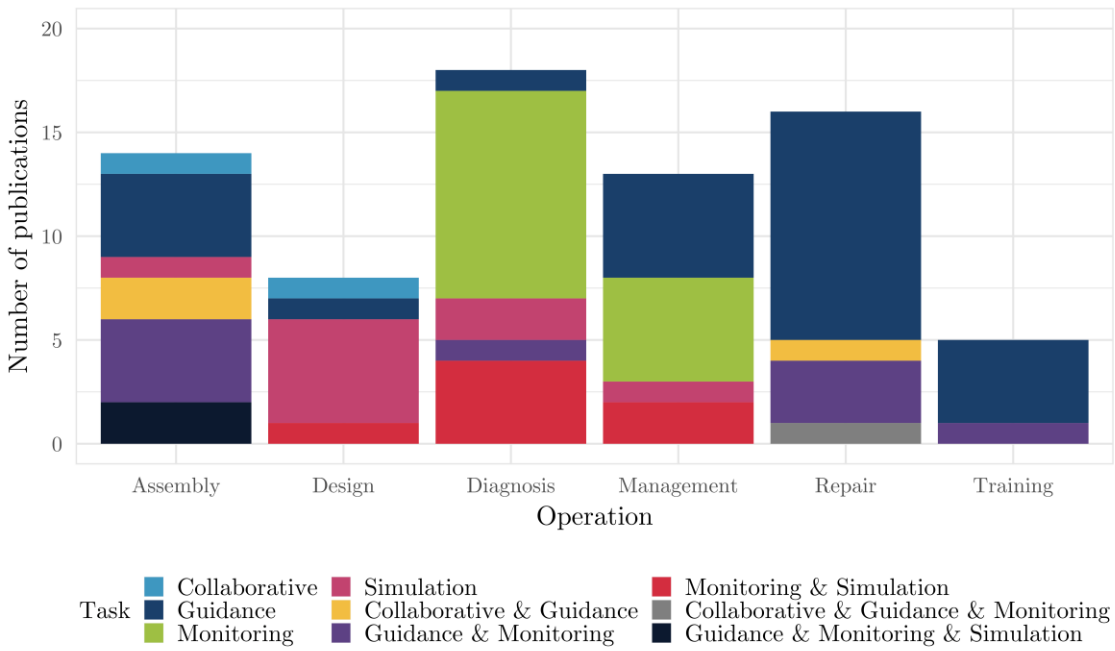


Figure 2-9. SLR analysis – ‘Task’ analysis: classification of 74 selected-relevant papers by type of ‘operation’ and type of ‘task’.

If ‘task’ is considered as the support offered by AR applications to maintenance ‘operations’, then it can be said that the type of support varies from one ‘operation’ to another. Based on results presented in Figure 2-9, all ‘operations’ have a component of ‘guidance’ tasks. Meanwhile, ‘collaborative’ has a minimum effect in any ‘operation’. Therefore, it seems that ‘monitoring’ and ‘simulation’ are the tasks that differentiate between ‘operations’. At this point in the analysis, the author noted that the more mature AR applications for those ‘operations’, the more combination in ‘tasks’ (e.g. ‘guidance’ and ‘monitoring’, ‘monitoring’ and ‘simulation’, etc.) they provided. Figure 2-10 is used to discuss whether the previous statement can be considered correct. As it can be seen, AR applications from later years have more ‘tasks’ combined (e.g. ‘simulation’ and ‘guidance’) than those for previous years. From a 0% of combinative support in 2012 and 20% in 2013 to a 40% in 2017, it seems there is an increasing trend in research for AR maintenance applications that offered support to multiple ‘tasks’.

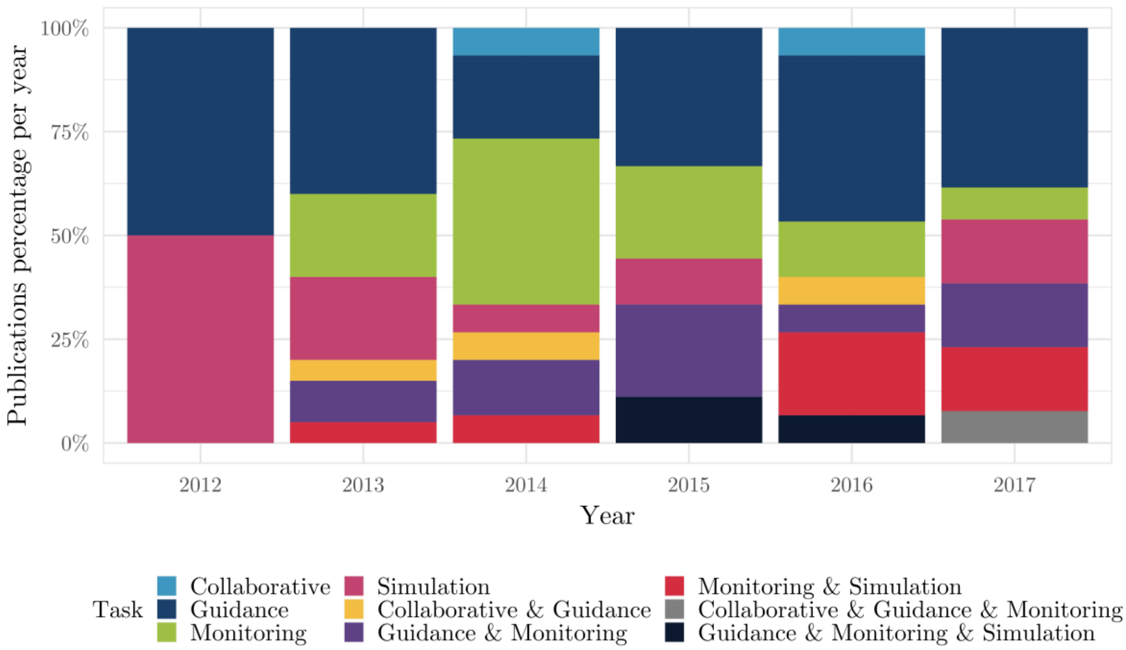


Figure 2-10. SLR analysis – ‘Task’ analysis: classification of 74 selected-relevant papers by year and type of ‘task’.

The previous paragraph supports the idea that for every ‘operation’, the more mature AR applications, the more ‘tasks’ they can offer to support users in their maintenance processes. Considering the relations between 'tasks' and knowledge transfer, it can be said that the more 'tasks' an AR application provides, the more knowledge it can transfer. Due to the objectives of this SLR, it seemed important to analyse whether more knowledge transferred is accompanied with more A, CA and IA techniques. For this matter Figure 2-11 is presented. It shows that the papers with more combined 'tasks' present more evidence of Authoring, Context-Awareness and Interaction-Analysis techniques. Therefore, it seems reasonable to establish a relation between A, CA and IA techniques and knowledge transfer. Nevertheless, how each technique relates to knowledge transfer has not been identified yet. This will be discussed in each technique Subsection (2.3.5, 2.3.6 and 2.3.7).

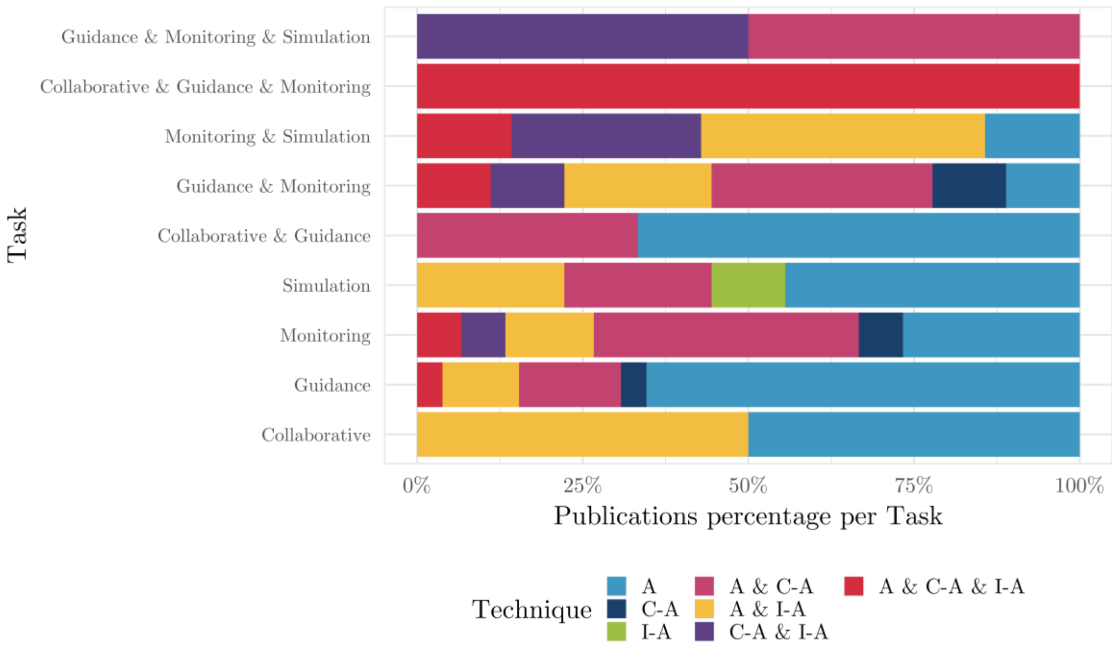


Figure 2-11. SLR analysis – 'Task' analysis: classification of 74 selected-relevant papers by type of 'task' and Authoring (A), Context-Awareness (C-A) and Interaction-Analysis (I-A) techniques included.

2.3.3.1 Findings summary

From the analysis presented in this subsection, several findings can be summarised:

- The more mature AR applications are, the more 'tasks' they provide.
- The more 'tasks' AR applications provide, the more knowledge these applications can transfer.
- The more 'tasks' AR applications provide, the more evidence shown of A, CA and IA techniques.
- Therefore, it appears to be a relation between A, CA and IA techniques and AR knowledge transfer capabilities enabled.

The inference above is only an appreciation. So, further analysis is required to identify how each technique relates to AR knowledge transfer capabilities.

2.3.4 Knowledge

Nonaka (Nonaka, 1994) defines knowledge as information in context. Moreover, he also defines as 'explicit' the knowledge that can be easily transferred to others as information (Nonaka, 1994). In other words, knowledge that can be easily represented or codified by data with a specified format (Becerra-Fernandez and Sabherwal, 2010). Because AR transfers knowledge by putting information in an explicit context, it is limited to transfer explicit knowledge. Table 2-6 identifies the categories of explicit knowledge: (i) procedural (sequences of steps, actions) and (ii) declarative (relationships among variables, facts) (Becerra-Fernandez and Sabherwal, 2010). Regarding this SLR objectives, it seemed important to analyse the relation between knowledge types and AR knowledge transfer capabilities.

Figure 2-12 presents 'knowledge' types transferred by AR applications considering the maintenance 'operations' they support. Compared to others, 'diagnosis' and 'management' operations transfer more 'declarative' than 'procedural' knowledge. A similar relation between 'operations' and 'tasks' was found in Subsection 2.3.2. So, it would be interesting to see the relations between 'task' and 'knowledge'.

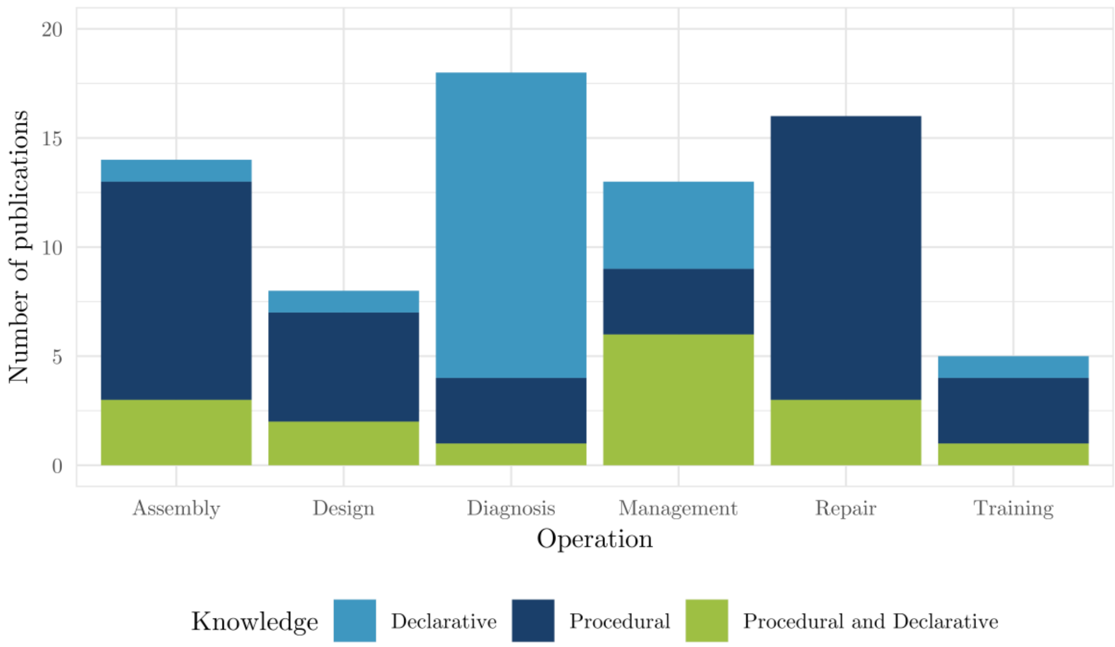


Figure 2-12. SLR analysis – 'Knowledge' analysis: classification of 74 selected-relevant papers by type of 'operation' and type of 'knowledge'.

Figure 2-13 presents the comparison between 'task' and 'knowledge' types. The results show that 'simulation' and 'monitoring' tasks provide more support in the form of 'declarative' rather than 'procedural' knowledge. Different examples can be found in (Zollmann, Hoppe and Kluckner, 2014; Ceruti, Liverani and Bombardi, 2017; Naets, Cosco and Desmet, 2017; Zhou, Luo and Yang, 2017). These papers emphasise the idea of using AR to provide the necessary knowledge for users to make certain decisions on their maintenance 'operations'. Nevertheless, less evidence has been found on how AR can be used to provide instructions ('procedural' knowledge) for making those decisions in order to increase maintenance efficiency.

Another relation to analyse is between 'knowledge' types and evidence on A, CA and IA techniques. That is presented in Figure 2-14. Apart from papers where only IA techniques are presented, the rest present similar numbers on the 'knowledge' types to which of A, CA and IA techniques are applied. Therefore, it seems reasonable to say that A, CA and IA techniques do not affect the 'knowledge' types being transferred.

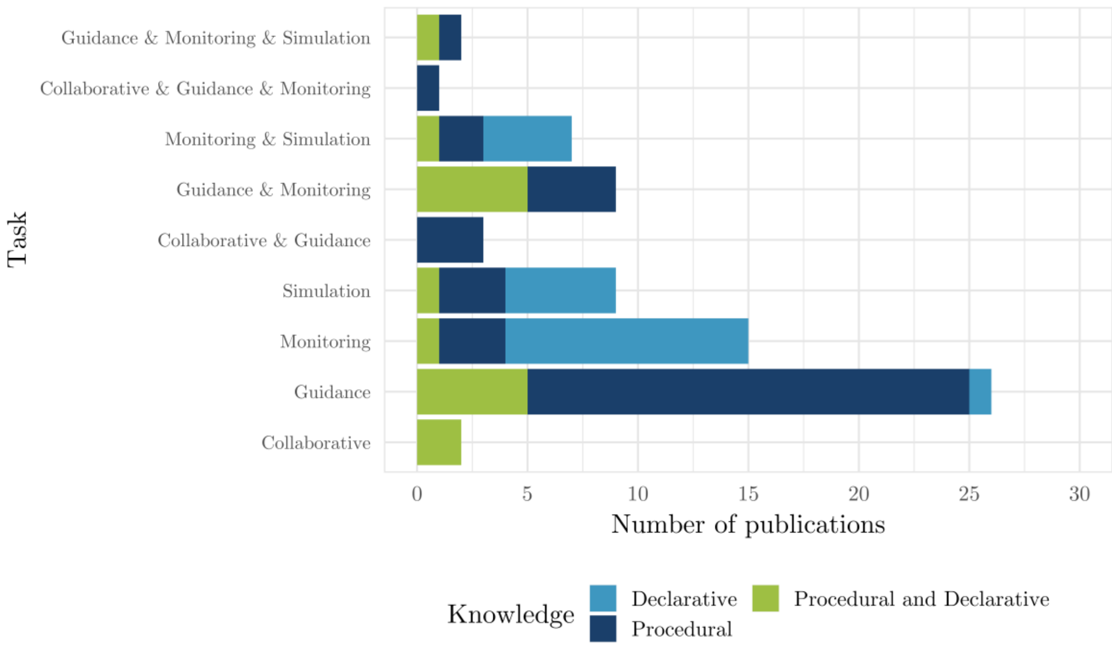


Figure 2-13. SLR analysis – 'Knowledge' analysis: classification of 74 selected-relevant papers by type of 'task' and type of 'knowledge'.

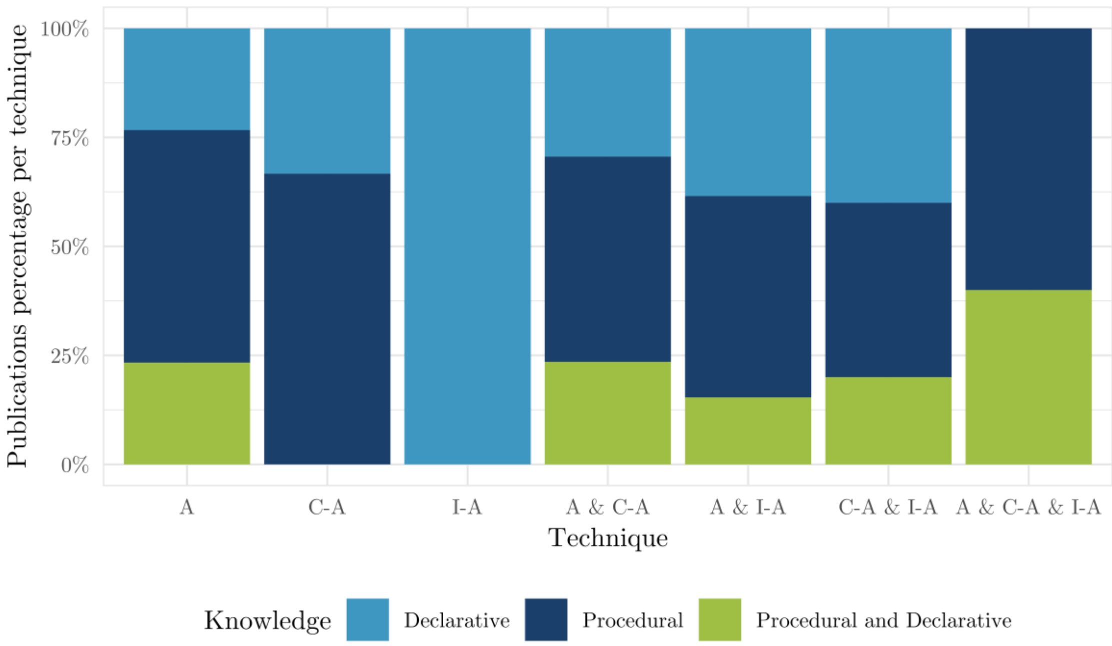


Figure 2-14. SLR analysis – 'Knowledge' analysis: classification of 74 selected-relevant papers by Authoring (A), Context-Awareness (C-A) and Interaction-Analysis (I-A) techniques and type of 'knowledge'.

2.3.4.1 Findings summary

From the analysis presented in this subsection, several findings can be summarised:

- There appears to be lack of 'procedural' knowledge for 'diagnosis' and 'management' operations.
- There appears to be a relation between 'tasks' and 'knowledge' types provided. 'Declarative' knowledge is more common in 'simulation' and 'monitoring' tasks.
- There is no reason to believe certain 'tasks' (e.g. guidance) cannot be applied to certain 'operations' (for 'diagnosis'). It seems that lack of 'tasks' to support certain 'operations' is due to the inability to explicit they 'knowledge' type being provided.
- Therefore, further research is required to specify explicit 'procedural' knowledge in 'diagnosis' and 'management' operations.
- There appears to be no relation between 'knowledge' types and the use Authoring, Context-Awareness and Interaction-Analysis techniques.

Therefore, it seems reasonable to believe that A, CA and IA techniques affect the way knowledge is being transferred but not its type. The following subsections discuss existing A, CA and IA techniques and how these affect to AR knowledge transfer capabilities.

2.3.5 Authoring

'Authoring' techniques are the set of software methods, tools and techniques to create augmented content for AR applications (Ong and Zhu, 2013). Figure 2-15 presents the existing categorised types and their distribution among those selected-relevant papers (66) that present evidence of 'Authoring' techniques.

The categories identified (Table 2-6) for 'Authoring' techniques can be classified according to: (i) potential users, (ii) automation, and (iii) content-creation rules.

Techniques that are automatic (i) do not require from users (ii) and vice versa. But, all can be classified according to their content-creation rules (iii). Therefore, six types are identified within the selected-relevant papers (Figure 2-15), two automatic, two for software developers (AR expertise), and two for application experts (maintenance expertise). Each pair is further categorised according to content creation rules. These rules can be programmed ad-hoc for each application (algorithmic) or taken from a structured knowledge-domain (or knowledge-based). These types are presented in the following paragraphs, starting from the top and reading anti-clockwise on Figure 2-15.

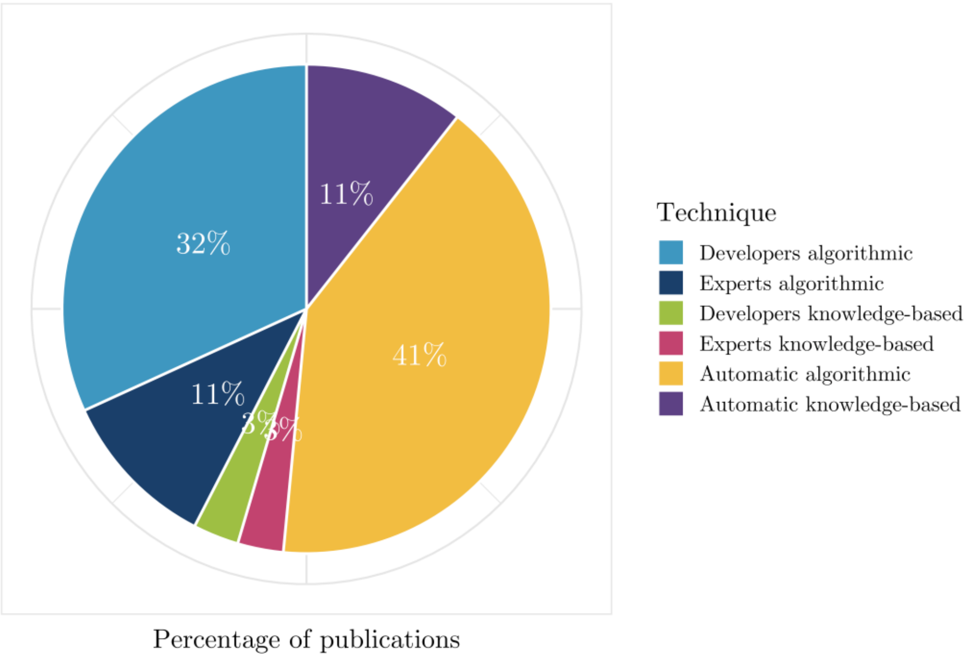


Figure 2-15. SLR analysis – 'Authoring' analysis: classification of 66 selected-relevant papers by type of 'Authoring' techniques.

'Developers algorithmic' techniques are those that manually create and display the content. This means that AR programmers (developers) "hard-code" the content and its interactions ad-hoc for the application being considered. A clear example is presented by Tatić and Tešić (2017), they proposed a step-by-step logic which displays content based on markers recognition. They also described the storage and the format given to the content created. 'Developers algorithmic' is the simplest form of Authoring and is often used in applications which focus is solving more fundamental AR challenges (Gimeno *et al.*, 2013). That is why these "manual" techniques are still being used in research of new fields of application. The author has noted evidence in papers presenting new AR applications in: safety monitoring (Alam et al., 2017), human-robot collaborative assembly (Makris et al., 2016), remote maintenance in radioactive environments (Martínez et al., 2014) and procedural guidance in chemical plants (Nakai and Suzuki, 2016). Moreover, the papers presenting evidence of 'Developers algorithmic' techniques mention the need to research more effective Authoring methods. These should enhance AR industrial implementation and decrease development and support costs.

'Experts algorithmic' techniques are an advancement towards AR industrial implementation and the previous step to introduce knowledge-structures for creating augmented content. These techniques are based on pre-programmed algorithms that allow non-programmers (i.e. application experts) to create content and determine its overlay. Application experts use interfaces, desktop or AR, to access virtual data, format it and generate the content. A good example is presented by Wang et. al in (2014) for collaborative design. In this case, application experts are given a desktop tool to import virtual data and allocate it as content in an augmented desktop where the design is taking place. Therefore, 'experts algorithmic' techniques are easier to implement because no AR experts are required to create content. Still, the algorithms for content generation are ad-hoc; and so, any changes to the rationale for content display have to be "hard-coded". That is why 'Experts algorithmic' share challenges (development cost, generality, etc.) and research focus (new AR applications (Gimeno *et al.*, 2013; Wang *et al.*, 2014; Yew, Ong and Nee, 2016; Ceruti, Liverani and Bombardi, 2017)) with 'Developers algorithmic' techniques.

Compared to 'algorithmic' techniques, 'developers' and 'experts knowledge-base' techniques differ in the way their algorithms are programmed. Instead of being "hard-coded", 'knowledge-based' algorithms are based on knowledge-domain structures that represent the 'operations' and/or 'tasks' being supported by the AR application. These structures for knowledge-representations (often ontologies

(Zhu, Ong and Nee, 2015) or taxonomies (Pokric et al., 2015)) provide a generic rationale that describe the knowledge and information associated to 'operations' and 'tasks'. So, there is no need to reprogram algorithms to augment new kinds of information. Although, they are the next logic step towards Authoring automation, fewer evidence has been found regarding these techniques compared to automatic ones. This is because once the knowledge representation for an 'operation' or 'task' is obtained, the automation of content creation is straightforward. Only those cases where the virtual data to be augmented needs from reformatting, these techniques have been considered. The author would like to note two relevant examples. ARAMS (Ong and Zhu, 2013) presents a bi-directional Authoring tool in which developers create content using a repair ontology and experts cross-validate the content created. The other example is called CARAGS (X Wang, Ong and Nee, 2016a). It presents an ontology for creating assembly content based on existing data (e.g. CAD models, assembly paths, etc.). So, experts only have to decide about the order of the content rather than its format. Overall, 'developers' and 'experts knowledge-base' techniques are better than the previous two, as they reduce development costs while maintaining accuracy and validity of content created, and so the knowledge being transferred.

'Automatic algorithmic' are the most present Authoring techniques within the SLR relevant papers. These techniques are programmed to create augmented content automatically from existing data without the need of experts' or developers' input. A clear example to describe these techniques is presented by Erkoyuncu et al. (2017). The Authoring software retrieves text and 3D models from an existing database and creates instructions and animations to overlay regarding the repair operations about the asset being tracked. These techniques are a step forward towards automatic content creation. But, they are still quite limited when considering adaptability of the AR applications to different scenarios (e.g. diagnosis instructions in repair operations) and acceptable data formats (often pre-determined by algorithms). Therefore, although these techniques reduce even further development costs, they don't perform better in terms of adaptability (data, scenarios) for industrial implementation. Relevant evidence of these techniques can be found in (Ham and Golparvar-Fard, 2013; Ng *et al.*, 2013;

Veas *et al.*, 2013; Golparvar-Fard and Ham, 2014; ElSayed *et al.*, 2016; Mourtzis, Vlachou and Zogopoulos, 2017b).

'Automatic knowledge-based' techniques are the ultimate approach to augmented content creation. These create content automatically based on existing data, as 'automatic algorithmic', but they also contextualise it according to knowledge-domain structures, as 'developers' and 'experts knowledge-base' techniques. Therefore, 'automatic knowledge-based' techniques hold the advantages (cost reduction, knowledge transfer accuracy) of those while reducing their disadvantages (validity and adaptability). Nevertheless, although these techniques overperform others, they are as difficult to achieve as the knowledge-domain representations they require to work. That is why only few 'operations' in certain 'industries' have achieved them. Relevant 'automatic knowledge-based' techniques have been found in 'equipment' 'assembly' (X Wang, Ong and Nee, 2016b), 'equipment' 'repair' (Espíndola et al., 2013) and 'environmental' 'inspection' (Pierdicca et al., 2016).

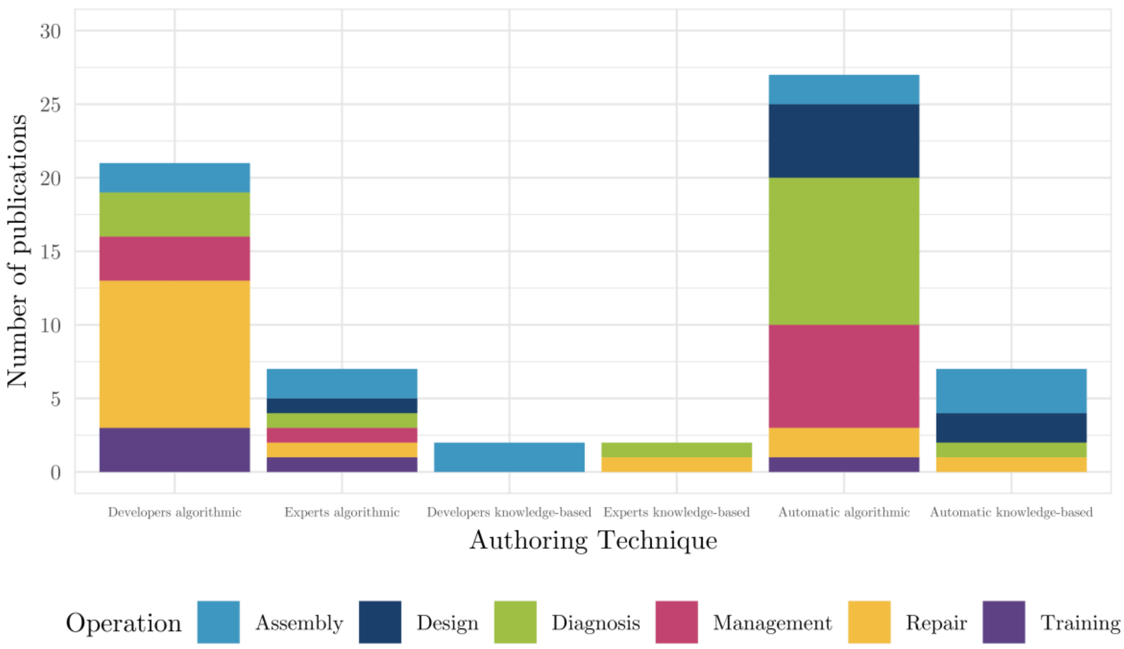


Figure 2-16. SLR analysis – 'Authoring' analysis: classification of 66 selected-relevant papers by type of 'Authoring' techniques and type of 'Operations'

2.3.5.1 Findings summary

Figure 2-16 presents Authoring techniques classified by supported 'operations.' Such classification helps to support the findings discussed above, which are summarised below:

- There appears to be a relation between the advancements in Authoring techniques and solutions for more fundamental AR challenges that is dependent on the 'operation' supported. Application fields where fundamental AR challenges have been solved have more advancements in Authoring techniques.
- AR knowledge transfer is affected by Authoring through its ability to create content. If the augmented content is correct, then the knowledge transfer obtained is valid.
- Development costs, industrial implementation, data and 'operation' adaptability, and knowledge transfer validity appear to be the research challenges in which Authoring techniques are focused on.
- 'Automatic' techniques perform better regarding development costs and industrial implementation.
- 'Knowledge-based' techniques outperform in data and 'operation' adaptability as they represent better the maintenance 'operations' and data formats to be supported by AR applications.
- Research gaps in Authoring depend on the techniques their selves and the maintenance 'operations' where are being applied.
- There appears to be lack of research in Authoring automation for less-matured applications.
- Further research is required to automatise Authoring generically independently from the application.
- There appears to be lack of research in knowledge-domain structures for automatised Authoring.

- Further research is required in knowledge-domain representations for maintenance 'operations'.
- There appears to be lack of research in the effect of Authoring in knowledge transfer validity.
- Further research is required to improve Authoring adaptability to changing scenarios to increase AR knowledge transfer validity.

2.3.6 Context-Awareness

Context-Awareness techniques are defined as software methods, tools and techniques that use contextual information to characterise augmented content (Erkoyuncu *et al.*, 2017). Where context is understood as *"any information that can be utilized to describe the situation of an entity. Where the entity can be a place, a person, or an object that is relevant to the interaction between a user and an application, such as time, location, activities, etc."* (Zhu, Ong and Nee, 2014).

Contextualising augmented content in AR applications has the target of enhancing the user in his/her consecution of a 'task' (Alam et al., 2017). Context-Awareness techniques achieve this by modifying already created content according to data obtained about the relevant context. This explanation includes the categories identified (Table 2-6) to classify Context-Awareness techniques: (i) contexts and (ii) rules. Rules refer to the logic/rational used to modify the content. These rules can be made ad-hoc for the AR application ('algorithmic') or based on knowledge-domain representations of the 'task' supported ('knowledge base'). Contexts refer to the relevant data according to which the content is being modified. These can be 'single', when rules consider only one piece of data, or 'multiple', when multivariable data is considered. Figure 2-17 presents the distribution of techniques categorised in the selected-relevant papers (29) that mention them.

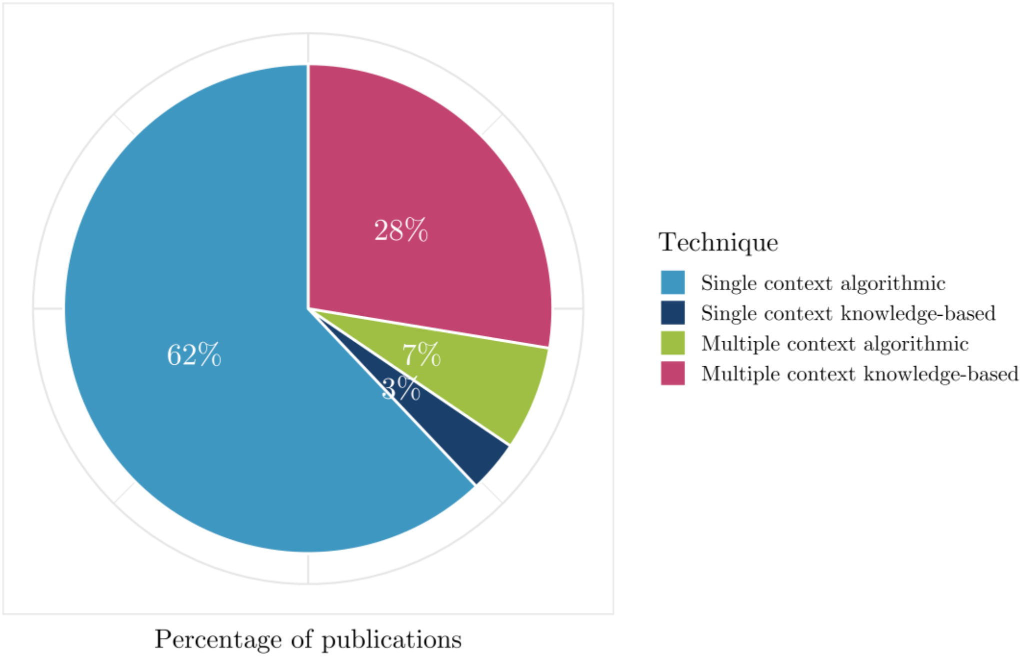


Figure 2-17. SLR analysis – 'Context-Awareness' analysis: classification of 29 selected-relevant papers by type of 'Context-Awareness' techniques.

Starting from the top and reading anti-clockwise the first technique is 'Single context algorithmic'. It is worthy to note that 62% of selected-relevant papers that present evidence of Context-Awareness are related to this type. 'Single context algorithmic' is the simplest method as it modifies augmented content according to one single variable through an algorithm "hard-coded" for the application. So, the adaptability of this technique to different situations is limited. A good example is presented by Ceruti, Liverani and Bombardi (2017). They proposed an algorithm to adapt the percentage of 3D model shown to the technician according to the percentage of 3D model printed. So, the technician can check for potential errors in the 3D printing process. These techniques help to modify the content accurately to the context. So, they are really good to increase the effectiveness of knowledge transfer. Relevant evidence can be found in (Veas *et al.*, 2013; Pokric *et al.*, 2015; Liu *et al.*, 2017; Mourtzis, Vlachou and Zogopoulos, 2017b; Zhou, Luo and Yang, 2017).

A similar approach is taken by 'Multiple context algorithmic' techniques. A clear example can be found in (Radkowski, Herrema and Oliver, 2015). In this case, the "hard-coded" algorithm considers the complexity of the visual interface and the difficulty of an ‘assembly’ operation to adapt the format of the content being shown. However, the content shown is not modified but instead, new content is created for each context scenario. Although, the benefits of these techniques are similar to the previous, they arise some drawbacks: increased development cost and difficulty to analyse the context. Content development costs are increased as more formats, and so contents, have to be created in order to keep up with contextualisation. Besides, the more complex the context is, the more difficult is to analyse it and identify the relevant variables.

The reason to use ‘knowledge-based’ techniques is to overcome the drawbacks mentioned above. They use knowledge-domain representations to identify context variables and determine the rules for modifying the content based on those. These techniques have direct advantages compared to the previous. First, there is no need to analyse relevant variables independently, as knowledge representations already consider them. Second, rules for adapting content can adapt to data formats; and so, content can be contextualised automatically. Therefore, there is no need to duplicate content and the associated costs can be reduced. The difference between ‘single context’ and ‘multiple context knowledge-based’ techniques is also related to the number of variables considered. Although, consideration of variables now depends on the ability to access or capture related data. A good example is presented by Zhu, Ong and Nee (2013), they proposed a Context Ontology for Maintenance Services (COMS) in which they consider variables (e.g. equipment model, expertise level, etc.) from different contexts (equipment, technician, etc.). They also use this ontology to add features to the content (e.g. text with associated 3D model), so it does not have to be duplicate it when contextualising it. More relevant evidence about these techniques can be found in (De Marchi et al., 2013; Zhu, Ong and Nee, 2014, 2015; Bleser et al., 2015; Radkowski, Herrema and Oliver, 2015; Laroche et al., 2016; Rohacs, Rohacs and Jankovics, 2016; X Wang, Ong and Nee, 2016a; Ceruti, Liverani and Bombardi, 2017; Ramirez-Amaro, Beetz and Cheng, 2017).

2.3.6.1 Findings summary

Figure 2-18 presents Context-Awareness techniques classified by ‘operations’.

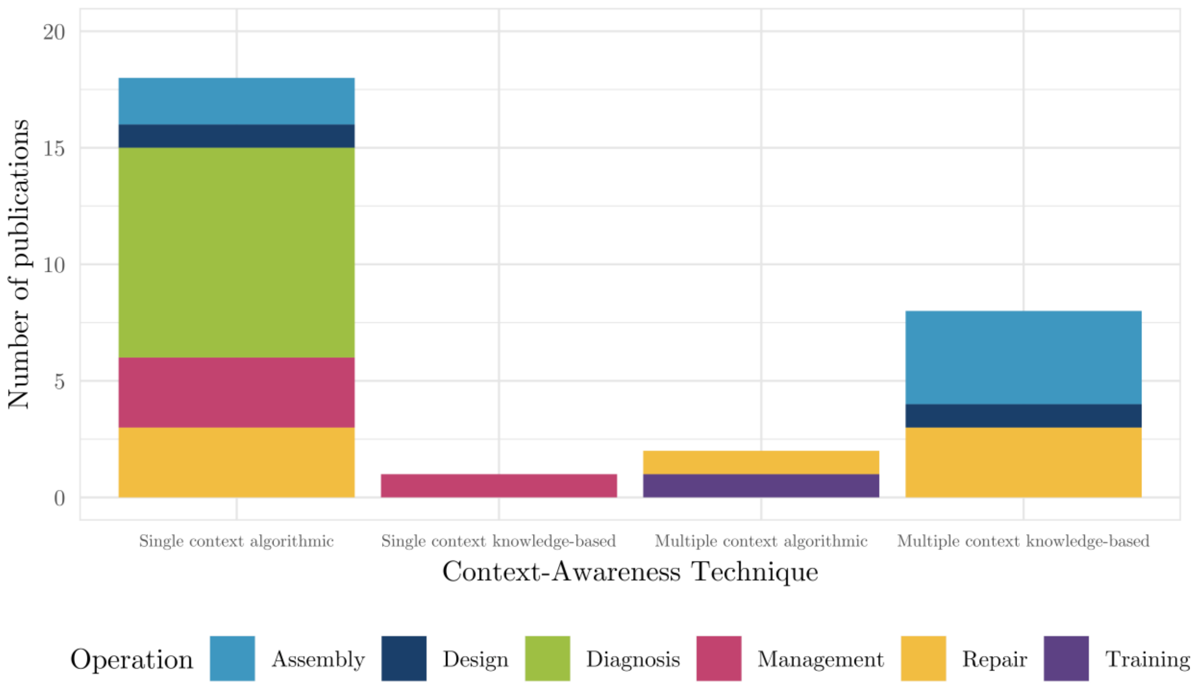


Figure 2-18. SLR analysis – 'Context-Awareness' analysis: classification of 29 selected-relevant papers by type of 'Context-Awareness' techniques and type of 'Operations'

Such classification helps to support the findings discussed above, which are summarised below:

- There appears to be a relation between research in 'Context-Awareness' and advancements in 'Authoring' techniques. The more content is automatically created, the more content is contextualised.
- AR knowledge transfer is affected by Context-Awareness through its ability to adapt content. If the content is more accurate according to its context, then the knowledge transfer is more effective.
- Development costs, content accuracy and knowledge transfer effectiveness appear to be the research challenges in which Context-Awareness techniques are focused on.
- 'Algorithmic' techniques achieve a more accurate contextualisation at the expense of being limited and costly (replicated content).

- ‘Knowledge base’ techniques provide a wider contextualisation at the expense of being less accurate.
- Research gaps in Context-Awareness depend on accuracy, adaptability trade-off and the maintenance ‘operations’ in which is being applied.
- There appears to be lack of automatic contextualisation in ‘algorithmic’ techniques.
- Further research is required to automate contextualisation and avoid content replication.
- There appears to be lack of data acquisition for context accuracy in ‘knowledge base’ techniques.
- Further research is required in automatic data acquisition to increase contextualisation accuracy.
- There appears to be lack of knowledge representations for context of certain maintenance ‘operations'.
- Further research is required in knowledge-domain representations in ‘diagnosis’, ‘management’ and ‘training’ operations.

2.3.7 Interaction-Analysis

‘Interaction-Analysis’ techniques are defined as software tools, methods or techniques that analyse the status of the interaction between the user and the augmented content to provide relevant feedback and/or improve the interaction itself (Webel et al., 2013; Westerfield, Mitrovic and Billinghurst, 2015). These techniques can be classified (Table 2-6) according to the level of automation regarding data (i) acquisition and (ii) analysis. These categories identified the level of user input required to conduct the analysis of interactions. Figure 2-19 presents the distribution of categorised types among the selected-relevant papers (25) that present evidence of ‘Interaction-Analysis’ techniques.

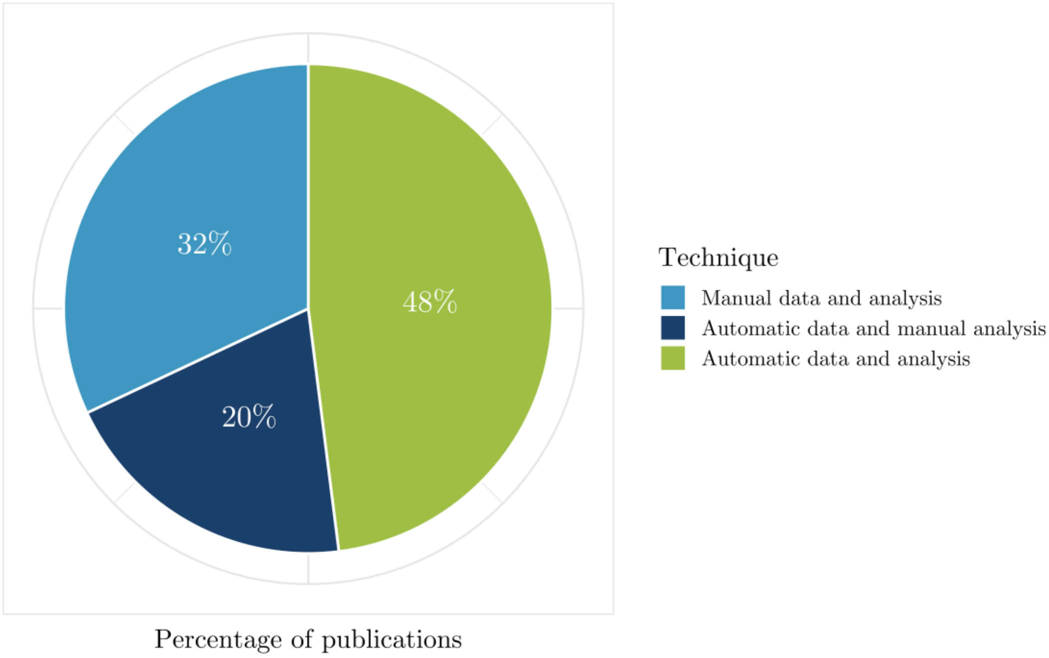


Figure 2-19. SLR analysis – 'Interaction-Analysis' analysis: classification of 25 selected-relevant papers by type of 'Interaction-Analysis' techniques.

'Manual data and analysis' are the second most extended 'Interaction-Analysis' techniques within the selected-relevant papers. These techniques provide manually acquired data (user feedback) for content creators to analyse it and modify contents and/or interactions. A good example is presented by Mourtzis, Vlachou and Zogopoulos (2017b). They utilised users' feedback about an asset for maintenance experts (content creators) to provide efficient augmented guidelines about its repair in a product-service environment. These techniques help to increase knowledge transfer efficiency through a more accurate content at the expense of real-time content creation. They are mainly focused in human collaboration (Webel *et al.*, 2013; Zhu, Ong and Nee, 2013; Mourtzis, Vlachou and Zogopoulos, 2017b), human-robot collaboration 'management' operations (Kim et al., 2013; Schall, Zollmann and Reitmayr, 2013; Kwon, Park and Lim, 2014) and procedural guidance (Nakai and Suzuki, 2016). Nevertheless, no real analysis is made by these techniques. They only give support to users, who really made the analysis and/or the decisions.

'Automatic data and manual analysis' techniques are the least represented within selected-relevant papers. They automatically acquire data for users to achieve a more efficient analysis. Compared to 'manual data and analysis' techniques, the structure of data captured speed up the analysis increasing the efficiency of knowledge transfer. A clear example is described by Sebillo et al. (2016). Their AR system allows trainers to make decisions on next training steps based on real-time trainees' status. The same approach has been described in different contexts: assembly design (Ng et al., 2013), tunnelling construction inspection (Zhou, Luo and Yang, 2017) and building energy performance evaluation (Ham and Golparvar-Fard, 2013). Nevertheless, all these papers mention the need to provide automatic analysis in order to increase the efficiency of these techniques.

'Automatic data and analysis' techniques are the most represented (48%) within selected-relevant papers (see Figure 2-19). These techniques acquire data and analyse it automatically, whose results are then used to modify content. Most techniques (Park et al., 2013; Golparvar-Fard and Ham, 2014; Liu et al., 2014, 2017; Westerfield, Mitrovic and Billinghurst, 2015; Chou et al., 2016; ElSayed et al., 2016; Rohacs, Rohacs and Jankovics, 2016) still provide these results to experts for them to update augmented content. Instead, few latest papers have achieved to automatically connect these techniques with 'automatic' Authoring. So, they are able to create content or modify existing automatically according to the interaction between users and augmented content. Apart from increasing knowledge transfer efficiency and reducing developments costs, these mixed techniques (Bleser et al., 2015; Wang, Ong and Nee, 2015; X Wang, Ong and Nee, 2016a; Ramirez-Amaro, Beetz and Cheng, 2017) have achieved knowledge capture capabilities. A good example is presented by Ramírez-Amaro, Beetz and Cheng (2017). They track the interaction between users and 3D models to analyse their trajectories and infer movement tasks associated with the real objects the 3D models represent. These tasks are then transferred to humanoid robots (programmed) or other users (animations). Apart from merging Authoring and Interaction-Analysis, these papers also present evidence of Context-Awareness for automatic interaction-data acquisition. Therefore, it appears reasonable to believe that joining Authoring, Context-Awareness and Interaction-Analysis techniques it is possible to also enable knowledge capture capabilities in AR technologies.

2.3.7.1 Findings summary

Figure 2-20 presents Interaction-Analysis techniques classified by 'operation' categories.

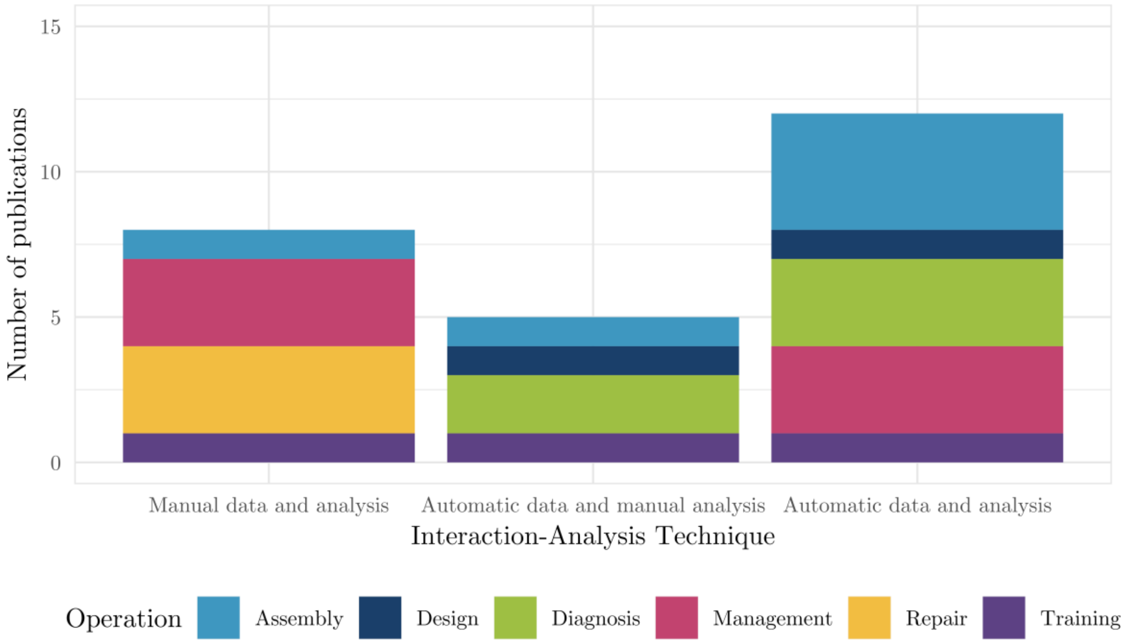


Figure 2-20. SLR analysis – 'Interaction-Analysis' analysis: classification of 25 selected-relevant papers by type of 'Interaction-Analysis' techniques and type of 'Operations'

Such classification helps to support the findings discussed above, which are summarised below:

- There appears to be a relation between advancements in Interaction-Analysis and Context-Awareness and Authoring advancements classified by maintenance 'operations'.
- AR knowledge transfer is affected by Interaction-Analysis through its ability to improve content. If content's accuracy and correctness are increased, then knowledge transfer is more efficient.
- AR knowledge capture capabilities are obtained when Authoring, Context-Awareness and Interaction-Analysis techniques are enabled jointly.

- AR knowledge capture capabilities have only been achieve in 'assembly' operations.
- Knowledge transfer efficiency and development costs appear to be the research challenges in which Interaction-Analysis techniques are focused on.
- The more automation in data acquisition and analysis, the more efficient knowledge transfer becomes. Although, the ability to adapt content to different situations (content validity) decreases.
- Research gaps in Interaction-Analysis depend on their relation with maintenance 'operations' and Authoring and Context-Awareness research gaps.
- There appears to be lack of manual data acquisition research in 'design' and 'diagnosis' operations.
- Further research is required in how to capture user feedback in maintenance 'design' and 'diagnosis'.
- There appears to be lack of automatic analysis research in 'management' and 'repair' operations.
- Further research is required to understand how AR interactions affect maintenance 'management' and 'repair' efficiency.
- There appears to be lack of research in AR knowledge capture capabilities in maintenance 'operations'.
- Further joint research in Authoring, Context-Awareness and Interaction-Analysis is required in maintenance 'operations' but 'assembly'.

2.4 Results

The thematic analysis provides enough evidence base to answer adequately the research questions. These answers, which are the results of the research conducted, are presented in the following subsections.

2.4.1 What is the state-of-the-art in Authoring, Context-Awareness and Interaction-Analysis techniques for AR maintenance applications?

A detailed description of existing techniques in Authoring (Subsection 2.3.5), Context-Awareness (Subsection 2.3.6) and Interaction-Analysis (Subsection 2.3.7) was provided earlier in this chapter. Besides, here the author summarises the latest techniques and their advantages, disadvantages and application areas:

- Authoring: content can be created either by 'developers', 'experts' or 'automatically', following specific-application ('algorithmic') or domain 'knowledge-based' rules. 'Automatic' techniques improve development costs and industrial implementation while limiting content adaptability. Besides, 'knowledge-based' techniques enrich content validity and 'operation' adaptability. 'Automatic knowledge-based' are the ultimate techniques. These create augmented content automatically from existing data according to 'operation' knowledge-domain rules. Nevertheless, they require detailed knowledge-domain representations which have only been achieved for 'small' assets in 'design' and 'assembly' operations.

- Context-Awareness: content can be modified based on 'single' or 'multiple' contexts (variables) according to specific-application ('algorithmic) or application-domain 'knowledge-based' rules. 'Multiple-context' provide more accurate contextualisation at the expense of higher development costs compared to 'single-context' techniques. Besides, 'knowledge-based' provide a wider contextualisation but less accurate than 'algorithmic' techniques. 'Multiple-context knowledge-based' are the most advanced. These use knowledge-representations of maintenance 'operations' to contextualise augmented content and are often connected to 'automatic knowledge-based' Authoring tools that share the same knowledge-domain representations. These techniques have been achieved only for 'repair' and 'assembly' operations.

- Interaction-Analysis: user-content interactions can be analysed 'automatically' or 'manually'. The same approaches can be used to acquire the data necessary for the analysis. The more automatic these processes become, the more effective analysis is for further content modification at the expense of less content adaptability. ‘Automatic data acquisition and analysis’ are the ultimate techniques. Their automatic results from user-interaction analysis provide direct rules to improve content creation (Authoring) and adaptation (Context-Awareness). To obtain those direct rules, they require to share knowledge-domain representations with Authoring and Context-Awareness techniques. That is why these techniques have only been achieved for ‘assembly’ operations in ‘small’ assets.

2.4.2 What are the research gaps in Authoring, Context-Awareness and Interaction-Analysis for Augmented Reality maintenance applications?

Specific discussions about existing A (Subsection 2.3.5), CA (Subsection 2.3.6) and IA (Subsection 2.3.7) techniques identified research gaps about their development. Besides, Subsection 2.3.2 presents a detailed map of A, CA and IA techniques application in maintenance ‘operations’ for different ‘assets. This map recognised research gaps concerning A, CA and IA application in different maintenance scenarios. Although these research gaps have already been discussed, both kinds are summarised in the following subsections.

2.4.2.1 Research gaps in Authoring, Context-Awareness and Interaction-Analysis techniques

2.4.2.1.1 Authoring

Authoring aims to create augmented content. Research has focused on creating ‘automatic’ methods in order to reduce industrial implementation issues and development costs. Although they are more efficient, they lack of ability to adapt content to different scenarios (data formats, ‘tasks’, and ‘operations’). That is why ‘knowledge-based’ techniques were proposed. So, the content is directly related to the maintenance knowledge-domain covered by the AR application. These approaches enhance content adaptability to maintenance ‘operations’, but other issues on industrial implementation still unsolved (e.g. adaptability to existing data formats and data storages, ‘tasks’, etc.). These are more AR-focused challenges and require from a more fundamental perspective rather than application-centred:

- There is lack of research regarding what kind of content to create depending on the ‘tasks’ to provide and not only the ‘operation’. Further research is required to define knowledge representations for the Authoring domain. So, knowledge-based rules can be applied to Authoring additionally to those of the application (‘operation’ knowledge domain).
- There is lack of research regarding Authoring adaptability to data formats and databases. Most ‘automatic’ techniques rely on specific data formats to create content from existing data (Espíndola *et al.*, 2013; Erkoyuncu *et al.*, 2017). Further research is required to enable automatic data conversions that capture the necessary features for content augmentation (e.g. string arrays from .doc files, small 3D meshes from CAD files, etc.).

Besides, there are certain maintenance ‘operations’ which do not have knowledge representations to enable ‘automatic’ Authoring. These research gaps are discussed in Subsection 2.4.2.2.

2.4.2.1.2 Context-Awareness

Context-Awareness aims to contextualise content created according to certain real-time variables (e.g. modify animations for assembly training according to user’s expertise level). Research has focused on developing ‘knowledge-based’ techniques. So, knowledge about a certain domain (‘operation’) to identify valid variables to create contextualisation rules (e.g. to identify expertise level by understanding time for a given instruction and years of working experience). Nevertheless, these techniques have some drawbacks regarding industrial implementation and development costs:

- There is lack of research on automatic content contextualisation. Due to the difficulty to classify this contextualisation (e.g. degree of difficulty), most

Context-Awareness techniques require to duplicate content instead of modifying existing (e.g. different text instructions depending on difficulty). Further research is required in automatic contextualisation techniques to avoid content duplication.

- There is lack of research on automatic contextual data acquisition. The use of complex variables for contextualisation while level of accuracy derives on difficulties for acquiring data for those variables (e.g. time to conduct an instruction for calculating user's experience level). Further research is required on obtaining or calculating contextualisation variables or reducing their complexity.

- Although the previous research gaps can be considered 'operation' dependent (contextual variables depend on the specific 'operation' domain), they can be considered from an AR-centred perspective. Further research is required to understand and describe the knowledge domain of Context-Awareness.

Besides, those two gaps still requiring from research in knowledge-domain representations regarding the context of certain maintenance 'operations'. These domains are in listed in Subsection 2.4.2.2.

2.4.2.1.3 Interaction-Analysis

Interaction-Analysis aims to analyse user-content interaction for enhancing augmented content effectiveness. Research has focused on creating 'automatic' techniques which can capture interaction-performance data and analyse it. Moreover, some techniques are able to connect those results with 'automatic' Authoring and Context-Awareness techniques, enabling knowledge capture capabilities. Nevertheless, most of these techniques still quite application-specific and almost none achieve knowledge capture. There are various reasons for this:

- There is lack of research on AR interactions and their features. Although they are similar, interactions are defined specifically for each application, and so Interaction-Analysis methods are difficult to extrapolate from one application to another. Further research is required to describe AR interactions and define knowledge representations for the Interaction-Analysis domain.

- There is lack of research on how to connect IA with A and CA techniques. Although the results of IA techniques are supposed to affect A and CA, most papers do not present evidence on how to connect them. Further research is required to provide rules from IA results to connect them automatically with A and CA techniques.

Besides, knowledge-domain representations in certain 'operations' have not achieved to represent user performance, needed for Interaction-Analysis. These domains are in listed in the following Subsection (2.4.2.2).

2.4.2.2 Research gaps in Authoring, Context-Awareness and Interaction-Analysis implementation in maintenance 'operations'

Although some Authoring (A), Context-Awareness (CA) and Interaction-Analysis (IA) methods have been achieved for certain 'operations, they are not directly applicable. This can be because there are still some research gaps which have not been fulfil for those 'operations':

1. Lack of knowledge-domain representations. There is the need to define knowledge structures such as ontologies or taxonomies to describe those maintenance 'operations'. So, A, CA and IA existing techniques can be applied to those.

2. Lack of 'automatic' Authoring techniques. For those maintenance 'operations' where AR research still at a fundamental stage (e.g. hardware, tracking issues), Authoring still being made by AR developers.

3. Lack of 'knowledge-based' methods for Authoring and Context-Awareness techniques. There are 'operations' where knowledge representations exist but those techniques have not been achieved yet.

4. Lack of 'automatic data acquisition and analysis' Interaction-Analysis techniques. These can only be obtained when 'knowledge-based' A and CA

exist. Moreover, these ‘operations’ require of further methods to merge A, CA and IA techniques.

This list of research gaps establishes a road-map of A, CA and IA research needs depending on the maintenance ‘operation’ and the ‘asset’ being considered. The end is to achieve latest techniques in each area: ‘automatic knowledge-based’ Authoring, ‘multiple-context knowledge-based’ Context-Awareness and ‘automatic data acquisition and analysis’ Interaction-Analysis. This is the set that enables AR knowledge capture and has only been achieved for ‘assembly’ operations in ‘small’ assets. For the rest of ‘operations’, Table 2-8 presents at which research gaps from the previous list need (table letters correspond to the list).

Table 2-8. SLR results: map of A, CA and IA research gaps in AR applications for maintenance ‘operations’ classified by ‘assets’.

Asset	Assembly	Design	Diagnosis	Mgmt.	Repair	Training
‘Large’	2	2	1, 3	2	2	1, 3
‘Medium’	1, 3	2	1, 3	4	2	2
‘Small’	*	4	1, 3	2	4	1, 3

2.4.3 What are the relations between Authoring, Context-Awareness and Interaction-Analysis techniques and knowledge transfer?

A, CA and IA techniques are related to the augmented content creation, contextualisation and improvement. Augmented content is the AR vehicle for knowledge transfer, as it is this content what puts information into context. Nevertheless, how these techniques affect to knowledge transfer is not a trivial question. Findings from A (Subsection 2.3.5), CA (Subsection 2.3.6) and IA (Subsection 2.3.7) Analysis Subsections are summarised below:

- AR knowledge transfer is affected by Authoring through its ability to create content. If the augmented content is correct, then the knowledge transfer obtained is valid.

- AR knowledge transfer is affected by Context-Awareness through its ability to adapt content. If the content is more accurate according to its context, then the knowledge transfer is more effective.
- AR knowledge transfer is affected by Interaction-Analysis through its ability to improve content. If content's accuracy and correctness are increased, then knowledge transfer is more efficient.

A theoretical explanation can be given to these findings. Authoring creates content based on information. If the information used to create the content is not right, then the knowledge transfer is not valid (e.g. using the wrong 3D animation to explain an instruction). Context-Awareness adapts the content generated to the specific context (e.g. user expertise). If the content is not adapted properly or to the wrong context, then the knowledge transfer is not effective (e.g. using step-by-step 3D animations for simple instructions for an expert). Interaction-Analysis evaluates the user-content interaction (knowledge transfer performance) and provides improvements for content's creation and contextualisation. So, it affects knowledge transfer efficiency by increasing content's correctness and accuracy.

2.4.4 What knowledge types are transferable by Augmented Reality?

The analysis of the different types of 'knowledge' considered in each selected relevant paper have found an important outcome: *'only that knowledge that can be represented ('explicit' knowledge) is able to be transferred by AR technologies'.* This finding seems reasonable. Only knowledge that can be transcribed in information can be transferred using Augmented Reality. Another question is whether 'implicit' can be converted into 'explicit' knowledge. Experts in knowledge management and conversion identify the SECI model (Nonaka, 1994; Becerra-Fernandez and Sabherwal, 2010) as a valid method to achieve that. Nevertheless, an interesting question for future AR research is whether AR can help to enhance the use of that model in organisations.

Two types of explicit knowledge where identified in the thematic categorisation (Table 2-6): ‘procedural’ and ‘declarative’. These two are similar to categories to those described by knowledge experts (Becerra-Fernandez and Sabherwal, 2010). Besides, findings from the Analysis Subsections (2.3.3 and 2.3.4) how these types relate to AR and maintenance applications:

- There appears to be a relation between ‘tasks’ and ‘knowledge’ types provided by AR applications.
- ‘Declarative’ knowledge is more common in ‘simulation’ and ‘monitoring’ tasks and ‘procedural’ in ‘guidance’ tasks.
- There appears to be lack of ‘tasks’ to support certain ‘operations’: ‘guidance’ for ‘design’ and ‘diagnosis’.
- Therefore, further research is required to explicit ‘procedural’ knowledge in ‘design’ and ‘diagnosis’ operations.

2.4.5 What potential applications of Augmented Reality for knowledge transfer are in maintenance contexts?

Knowledge transfer can be considered a primal objective of any AR application. As discussed in previous sections, as long as there are Authoring within an AR application, knowledge transfer is enabled. Then, this knowledge transfer can be more effective and efficient through Context-Awareness and Interaction-Analysis, respectively. In the case of maintenance applications, Table 2-8 identifies the road-map in AR research to achieve effective and efficient knowledge transfer for each maintenance ‘operation’ and ‘asset’. Throughout the analysis conducted in this chapter, it has been considered that AR technologies are able to transfer knowledge from data-repositories or experts to AR users. But little has been said about the opposite direction, where knowledge is transferred from AR users to data-repositories. This opposite direction has been described by experts as knowledge capture.

Knowledge capture is defined as *“the process of retrieving either explicit or implicit knowledge that resides within people, artefacts, or organizational entities”*

(Becerra-Fernandez and Sabherwal, 2010). In the case of AR, it seems reasonable to narrow this definition to the retrieving of explicit knowledge from people (AR users). Unlike knowledge transfer, AR knowledge capture capabilities require more from A, CA and IA techniques to be enabled. This can be theoretically explained through the connection between content, user and knowledge. In knowledge transfer, knowledge is delivered through the content. Instead in knowledge capture, knowledge should be obtained from the content. Therefore, content should be created by the user in order to capture the information and the context of use. That is why Interaction-Analysis and Context-Awareness are also required, so the context and the interaction can be analysed to check whether the content is correct and accurate. Only few evidences (Bleser *et al.*, 2015; Ramirez-Amaro, Beetz and Cheng, 2017) of AR knowledge capture have been found in this SLR. These recognised that knowledge capture was achieved in 'assembly' of 'small' assets with certain techniques: 'automatic knowledge-based' Authoring, 'multiple-context knowledge-based' Context-Awareness and 'automatic data acquisition and analysis' Interaction-Analysis. Nevertheless, further research is required to specify how knowledge capture can be obtained with AR and what other knowledge management processes can be enhanced through the use of this technology.

2.5 Discussion

The results (Section 2.4) offer a discussed view of this SLR objectives and findings. Still, it is needed to assess to what extent its research methods and results are valid. That is the purpose of this section. Research methods are discussed according to their validity and objectivity in Subsection 2.5.1. Research results are examined regarding their quality, validity and applicability in Subsection 2.5.2.

2.5.1 Research methods validity and objectivity

In order to provide a suitable discussion, it is important to define validity and objectivity of the research methods utilised in this SLR. By validity, the author understands the extent to which the research methods achieve the research objectives. By objectivity, the author refers to the ability to avoid bias in and increase transparency and replicability of the research. In this Subsection, frameworks and methods used within each phase of the SLR (Figure 2-1) are examined against these two concepts:

- SLR method selection: SALSA is a framework that conceptualises the stages of a systematic literature review and identifies the most suitable methods for each phase according to its objectives (Grant and Booth, 2009). It has been selected, against others such as Kitchenham's (Kitchenham, 2004), Cochrane (Higgins and Green, 2008) or Xiao and Watson's (Xiao and Watson, 2017), due to the following reasons. First, its approach is generic enough to be applicable for different reviews. Second, it proposes validated research methods for each phase. Last but not least, several authors have already applied this framework for similar reviews in this research field (Corrêa dos Santos, Delamaro and Nunes, 2013; Palmarini *et al.*, 2018).
- Protocol: the PICOC framework (Booth, Papaioannou and Sutton, 2012) was used at this stage to determine this review's scope. The SALSA framework (Grant and Booth, 2009) identifies some frameworks to do this (PICOC, SPICE, CIMO, etc.). The selection of one depends on the type of research questions to define. For this SLR, the PICOC framework was chosen due to two reasons. First, it is a valid, well-contrasted tool that has been used in other reviews within the same research field (Corrêa dos Santos, Delamaro and Nunes, 2013; Palmarini *et al.*, 2018). Second, it helps to identify the main concepts to define the research methods required for following SLR steps (Booth, Papaioannou and Sutton, 2012). It is the author's believe that this selection helps to achieve further transparency and replicability for the systematic review proposed.
- Search: the definition of search parameters (databases and string) was made based on PICOC framework results and supported on the propositions of similar reviews (Corrêa dos Santos, Delamaro and Nunes, 2013; Bacca *et al.*, 2014; Palmarini *et al.*, 2018). Regarding the database selection, only those

which were relevant and provided necessary meta-data were included. Therefore, it is the author's believe that the database selection can provide a relevant sample of papers and the SLR results should not be affected. Regarding the search string, it was created following the guidelines from Booth et. al (Booth, Papaioannou and Sutton, 2012) and only included those terms within the research scope. Therefore, it is the author's believe that the string created covers the research population the SLR was aiming to.

- Appraisal: one important topic to discuss is the inclusion/exclusion criteria definition. The criteria were determined utilising the research scope (Table 2-1), the guidelines from similar works (Corrêa dos Santos, Delamaro and Nunes, 2013; Akçayir and Akçayir, 2017; Palmarini *et al.*, 2018) and the author's experience in the research field. These criteria can be classified in four categories:

  1. Criteria directly connected to the research scope (I1, I2, I3, E4, E5): these criteria were defined using the required concepts from the PICOC framework (Table 2-1): interactions and context. The aim was to select those papers searched by population which had certain relevancy for the topic being reviewed.

  2. Criteria indirectly connected to the research scope (E2): the author noticed that the AR community was not focus in A, CA and IA associated challenges prior to 2012. Either because these challenges were not clearly identified then or because the techniques their selves were not mature enough to be explicitly described yet. These arguments are supported by other authors (e.g. Nee et. al), who did not mention these topics in their reviews before 2012 (Ong and Nee, 2004; Ong, Yuan and Nee, 2008) but they did from then onwards (Nee et al., 2012; X. Wang, Ong and Nee, 2016).

  3. Criteria related with research/data accessibility (E1, E3, and E8): these help to ensure that selected-relevant papers are fully assessed. So, all the evidence required from them can be extracted.

4. Criteria related with research quality and validity (E6 and E7): these help to ensure that selected-relevant papers present complete, peer-reviewed conclusions. So, the findings used to draw conclusions in this review are correct.

These criteria are supposed to be able to narrow down the papers' population from the search results to a sample specifically related to this SLR scope. Therefore, it could be said that these criteria were complete and sufficient for the purpose of this SLR. Besides, it is also important to note the resulted number of selected-relevant papers. The criteria were applied in order to obtain papers that are relevant for the reviewing topics. So, the number of papers selected should not affect the validity of the SLR results. However, this number is comparable to those in similar reviews (Nee *et al.*, 2012; Corrêa dos Santos, Delamaro and Nunes, 2013; Akçayir and Akçayir, 2017; Palmarini *et al.*, 2018).

- Synthesis: there are two important topics to discuss: themes definition and thematic categorisation. The definition of the right themes is critical to extract the relevant and necessary evidence from papers regarding the SLR questions. Following experts suggestions (Booth, Papaioannou and Sutton, 2012; Corrêa dos Santos, Delamaro and Nunes, 2013; Palmarini *et al.*, 2018), these have been defined considering the concepts within the SLR scope (Table 2-1). Besides, the thematic categorisation helps to classify and analyse the data extracted from the papers. So, its outcomes directly influence the SLR results. Because no method have been found in similar works within the research field, a systematic, reproducible process was created following the guidelines from experts (Booth, Papaioannou and Sutton, 2012). Moreover, the resulted categories were validated and corrected by comparison with similar works (Corrêa dos Santos, Delamaro and Nunes, 2013; Bacca *et al.*, 2014; Palmarini *et al.*, 2018).

- Analysis: the analysis was conducted with a combination of narrative, tabular and graphical analysis. In order to increase the reproducibility of the research, it is needed to describe how graphs and tables where selected. This selection was part of the exploratory phase of the analysis. The patterns identified between categories (variables) were transferred to graphs for supporting the narrative analysis. Besides, although the results section (Section 2.4) includes a summary of the findings, the analysis (Section 2.3) was included for transparency and reproducibility of the results obtained.

- Report: within those suggested by the SALSA framework (Booth, Papaioannou and Sutton, 2012), the PRISMA methodology (Moher *et al.*, 2010) was the most mature tool identified for reporting. That is why it was selected to write the report and then create the chapter by excluding those parts which did not add value to the community. Thus, the author believes that the chapter includes all the information required to understand and replicate the SLR.

Based on the previous discussions, some improvements have been identified and are listed below:

- To reduce quality assessment biases and provide additional results to be used during the analysis.
- To enhance validation of thematic categorisation results.
- To provide standardised guidelines for the exploration within the analysis phase.
- To include a standard method for research methods reporting.

Besides these improvements, the application of the SALSA framework along with the research methods selected can be considered a valid approach. Therefore, further applications of it can also be considered within the same research field:

- To review A, CA and IA and their relations with knowledge transfer capabilities in other AR fields of application (e.g. medicine, marketing, manufacturing, etc.).
- To review knowledge transfer capabilities of other visualisation technologies (e.g. Virtual Reality).

2.5.2 Research results quality, validity and generality

If the research method assessment provides an evaluation of the internal validity of the research, the results assessment does if for the external validity (Booth, Papaioannou and Sutton, 2012). Three criterions are used to evaluate this SLR findings and results. By quality, the author refers to the value of the results according to the research questions. By validity, the author understands the effectiveness of results. By generality, the author considers the extent to which the results are applicable. Results quality and validity are discussed for each research question independently.

2.5.2.1 What is the state-of-the-art in Authoring, Context-Awareness and Interaction-Analysis techniques for Augmented Reality maintenance applications?

Part of the narrative analysis provided on AR applications (2.3.2) and A (2.3.5), CA (2.3.6) and IA (2.3.7) techniques is already a description of the state-of-the-art that can be considered a result. Besides, these results are summarised and further discussed in the correspondent results Subsection (2.4.1). These results explain the latest advancements of A, CA and IA techniques in AR maintenance applications and their benefits and drawbacks to the time when the analysis was done. Even though there could be techniques not covered in this research (e.g. newer papers not included), the categorisation proposed enables to introduce and classify new techniques. Therefore, this SLR results are still valid as long as they are updated with new evidence from latest relevant publications.

2.5.2.2 What are the research gaps in Authoring, Context-Awareness and Interaction-Analysis for Augmented Reality maintenance applications?

The rest of the narrative analysis provided on AR applications (2.3.2) and A (2.3.5), CA (2.3.6) and IA (2.3.7) techniques has been used to identify and discuss these research gaps. Due to the scope of this SLR, two types of research gaps have been identified.

One type refers to research gaps in A, CA and IA techniques. These are more AR-focused, fundamental research gaps rather than maintenance-centred. The validity of these results may be affected due to the narrow scope of the SLR considering AR. The gaps identified might have been solved in other papers which were not relevant to this SLR scope (e.g. other application fields, computer science papers, etc.). However, they seem relevant research gaps because no evidence of such papers has been found referenced within the selected-relevant sample.

The other gap type relates to the application of A, CA and IA in maintenance 'operations'. They identify research gaps within specific 'operations' compared to those most advanced. That is why a road-map for future research in AR-maintenance applications. Although these gaps have already been solved for different applications, it is still a contribution to apply those techniques to different scenarios. And so, those can still be considered research gaps.

2.5.2.3 What are the relations between Authoring, Context-Awareness and Interaction-Analysis techniques and knowledge transfer?

These relations have been declared within the results section (2.4.3). A theoretical discussion was conducted to establish those relations. Besides, the results from that discussion are supported by findings from the Analysis section (Subsections 2.3.3 and 2.3.4) which present evidence from the selected-relevant papers. Therefore, it seems reasonable to believe that the explanation given to this question is sufficient to understand those relations. Nevertheless, further research to understand those relations more in-depth would be advisable.

2.5.2.4 What knowledge types are transferable by Augmented Reality?

The results related to this part of the analysis (Subsections 2.3.3 and 2.3.4) demonstrate that only 'explicit' knowledge can be transferred by AR technologies. Besides, the categories identified (Table 2-6) within the SLR coincide with the definitions provided by experts in knowledge related research. Even though it could be argued that other knowledge classifications could be used, the one utilised is coherent and consistent with the results obtained. Therefore, it seems reasonable to say that the categorisation obtained is sufficient. Nevertheless, other research gaps along these matters have been identified and discussed within the correspondent results Subsection (2.4.4).

Those gaps are related to the support provided by AR 'tasks', the 'knowledge' type being transferred and the need to explicit that knowledge. Although the scope of these gaps is outside the scope of this SLR, it seems reasonable to say that they are important to AR related research. It is necessary to understand how certain 'knowledge' can be specified in order to comprehend whether AR can be a useful to transfer it.

2.5.2.5 What potential applications of Augmented Reality knowledge transfer are in maintenance contexts?

Rather than the application of AR knowledge transfer in maintenance, the section on this SLR question (2.4.5) discusses the idea of knowledge capture using AR. The idea is supported by evidence from selected-relevant papers and inferred during the Analysis phase of this SLR. The discussion (Subsection 2.4.5) provides a theoretical explanation of how AR knowledge captured is achieved with A, CA and IA techniques and how does it affect to the AR research gaps previously identified. Therefore, although the idea was not considered within the initial SLR scope, it seems reasonable to include it as an additional, relevant finding.

Besides, research results should also be discussed according to their validity. Summarising the answer to the SLR questions (Section 2.4), it can be said that A, CA and IA techniques and AR knowledge transfer capabilities have been reviewed within the context of maintenance applications. The context of maintenance applications has been narrowed to medium-long life complex assets. There were two reasons to establish this scope. First, it includes maintenance applications where knowledge transfer is required. Second, it helps to reject papers where maintenance of low value assets is considered, which can be different in terms of knowledge transfer requirements. Nevertheless, it could be argued that low value assets do not require complex maintenance operations and so these AR applications are closer to manufacturing rather than maintenance operations.

Besides, the results of this research cannot be validated to other AR fields of application, such as manufacturing, medicine, marketing, etc. Further research should be required to extend the description of AR knowledge transfer capabilities and the techniques that enable them for other fields of application. Moreover, there is also an idea behind this research that could be considered for other visualisation or knowledge transfer technologies under research and development. These results could be used as basics to understand by comparison how other technologies' techniques enable knowledge transfer capabilities.

The previous discussions on research method and results cover different perspectives to the results and discussions presented in other sections of this SLR. All can be used as basis to extract conclusions and future works regarding this research.

2.6 Conclusions and future works

Academic literature reviews in AR have not focused on A, CA and IA techniques for maintenance applications neither in AR knowledge transfer capabilities. Therefore, this chapter has aimed to describe the state-of-the-art in A, CA and IA techniques and their relations with AR knowledge transfer capabilities in maintenance contexts. In order to do so, a SLR of 74 relevant academic papers was conducted for the papers between 2012 and 2017.

The SRL research protocol was based on the SALSA Framework (Booth, Papaioannou and Sutton, 2012) and inspired in similar reviews (Corrêa dos Santos, Delamaro and Nunes, 2013; Akçayir and Akçayir, 2017; Palmarini et al., 2018). The protocol presented (Section 2.2) ensures both reproducibility and transferability of the study. The SLR comprised the search and appraisal of applied research in AR-maintenance journal articles. It also included a thematic analysis (Section 2.3) of 74 relevant-selected papers. The results of such analysis were used to answer five pre-defined research questions: (1) the description of the state-of-the-art of ACAIA techniques in maintenance contexts, (2) their research gaps, (3) their relations with AR knowledge transfer capabilities, (4) the types of knowledge transferrable and (5) potential applications for AR knowledge transfer. The answers and discussions related to SLR questions are reported in Section 2.4. Besides, research methods and results are discussed in Section 2.5 according to their validity and applicability.

Overall, this research has led to some conclusions in the area of AR in maintenance, A, CA and IA techniques, and knowledge transfer capabilities. Conclusions from Section 2.4 and 2.5 are summarised below:

- A, CA and IA techniques have achieved different levels of technological maturity in different maintenance applications ('operations' by 'asset).
- There appears to be a relation between these technological maturity levels and the existence of knowledge-domain representations for the maintenance applications ('operations') considered.
- Development costs, industrial implementation and data and maintenance 'operation' adaptability appear to be the research challenges in which A, CA and IA techniques are focused on.
- A, CA and IA affect respectively to AR knowledge transfer validity, effectiveness and efficiency.
- There appears to be a relation between certain 'tasks' and 'knowledge' types in AR applications.
- There appears to be a relation between advancements in Interaction-Analysis, and Context-Awareness and Authoring advancements classified for maintenance 'operations'.
- Latest advanced techniques in A, CA and IA are respectively: 'automatic knowledge-based', 'multiple-context knowledge-based' and 'automatic data acquisition and analysis'.

- These advanced techniques are capable to trigger AR knowledge capture capabilities.

The research gaps and the conclusions identified by this research also draw a map that points to various future research directions. A summary of all further research works mentioned within the results and discussion sections is listed below:

- SLR methods improvements: (i) standard methods for thematic categorisation; (ii) standard methods for thematic exploratory analysis; and (iii) standard reporting of narrative/tabular analysis.
- SLR methods applicability: (i) to conduct similar reviews in other fields of application within AR; and (ii) to conduct similar reviews to other knowledge transfer technologies.
- Authoring: (i) to define a knowledge representation of the Authoring domain; and (ii) to enable automatic data conversions considering special features of content augmentation.
- Context-Awareness: (i) to study automatic contextualisation for avoiding content duplication; (ii) to study automatic context data acquisition for enhancing contextualisation accuracy; and (iii) to define a knowledge representation of the Context-Awareness domain.
- Interaction-Analysis: (i) to study AR user-content interactions from a generic perspective; (ii) to define a knowledge representation of the Interaction-Analysis domain; and (iii) to study methods for automatically varying A and CA results depending on IA outcomes.
- Road-map for A, CA and IA in maintenance applications (Table 2-8): (i) to define knowledge-domain representations of 'diagnosis' and 'training' of 'large' assets, 'assembly' and 'diagnosis' in 'medium' assets, and 'diagnosis' and 'training' of 'small' assets; (ii) to apply 'automatic' Authoring in 'assembly', 'design', 'management' and 'repair' operations in 'large' assets, 'design', 'repair' and 'training' operations in 'medium' assets, and

'management' operations in 'small' assets; (iii) to apply 'knowledge-based' Authoring and Context-Awareness methods in diagnosis' and 'training' of 'large' assets, 'assembly' and 'diagnosis' in 'medium' assets, and 'diagnosis' and 'training' of 'small' assets; and (iv) to apply 'automatic data acquisition and analysis' Interaction-Analysis methods in 'management' of 'medium' assets and 'design' and 'repair' of 'small' assets.

- Knowledge types transferrable by AR: (i) to study knowledge depictions to explicit 'procedural' knowledge in 'design' and 'diagnosis' operations.
- AR knowledge transfer applications in maintenance contexts: (i) to study how knowledge capture can be obtained in AR applications; (ii) to study where AR knowledge capture can have value within maintenance 'operations'; and (iii) to study which other knowledge-management processes can be enhanced by the use of AR in maintenance-related organisations.

Augmented Reality has been revealed as an impactful technology for organisations to transfer knowledge (from information systems to users) and capture and discover it (from users to information systems). Nevertheless, there are still some questions that have not been answered yet: what are the requirements to achieve it in real-world scenarios? What other technologies could be integrated with AR to achieve more powerful applications? What would be the role of AR in industrial organisations? What would be the role of AR in knowledge management? Even though AR is a maturing technology close to achieve real-life implementation, there are still many questions to answer about what its full potential is.

Chapter 3
Ontology-based diagnosis reporting and monitoring to improve fault finding in Industry 4.0

3.1 Introduction

Industry 4.0 offers improvements in operational efficiency, effectiveness, and safety through technological developments such as Cyber-Physical Systems (CPS), Artificial Intelligence (AI) and Augmented Reality (AR) (Pedersen *et al.*, 2016). These improvements are enabled through automatic data exchange between multiple devices such as sensors, 3D scanners, robots or Radio Frequency Identification Devices (RFID) (Nagy *et al.*, 2018). Industry 4.0 envisions a future based on integrated data management for real-time asset control and process optimisation (Angelopoulos *et al.*, 2020). Nevertheless, there are still data research challenges that impede such future (Rødseth, Schjølberg and Marhaug, 2017). These include heterogeneity of data formats (e.g. audio, video, etc.) and lack of structure of existing data sources (e.g. manuals, reports, etc.) (Vogl, Weiss and Helu, 2019).

In the Industry 4.0 era, a relevant example for unstructured data capture are experts' diagnosis reports. Expert diagnosis refers to knowledge-intensive human tasks aiming to identify a failure root cause by identifying all faults from its initial symptom (Zhou, Yu and Zhang, 2015). Although diagnosis systems are becoming more automated, experts are still required to conduct diagnosis procedures when these do not provide satisfactory solutions (Angelopoulos *et al.*, 2020). A relevant example is No-Fault-Found (NFF) scenarios (Rødseth, Schjølberg and Marhaug, 2017). NFF includes those events where the lack of robust failure modes impede a diagnosis system determine the cause of an identified failure (Khan *et al.*, 2014). When experts conduct diagnosis procedures, they are usually required to report them along with failure modes and conditions identified (Wan *et al.*, 2019). So, explicitly capturing experts' knowledge in structured reports can help to reduce the lack of robust failure modes (Nuñez and Borsato, 2017). Thus, improving maintenance efficiency and effectiveness as experts' structured knowledge can be reused by others (e.g. novices or diagnosis systems) or in different activities (e.g. repair or monitoring) (Pistofidis *et al.*, 2016; Nuñez and Borsato, 2017).

In maintenance diagnosis research, ontologies have been used to classify and correlate information from unstructured sources (Fernández-López, Gómez-Pérez and Juristo, 1997). Ontology-based methods have been widely researched for two applications: diagnosis decision support and data modelling for maintenance planning. Two examples are: (1) natural language processing of free-text diagnosis reports to recommend failure modes (Zhong *et al.*, 2018), and (2) fault propagation modelling for effective maintenance planning (Dibowski, Holub and Rojíček, 2017). Nonetheless, there is fewer literature evidence on ontology-based methods focused on capturing expert diagnosis knowledge and re-using it to enhance monitoring systems for improving efficiency of fault-finding tasks (Wan *et al.*, 2019).

Experts are capable of identifying failure modes from components' conditions identified through unstructured data such as incoherent signals (e.g. no-fault-found). If such conditions were to be structured, then monitoring systems could be improved by replicating the identification of those conditions. For example, an expert can be capable of identifying a gearbox failure through a sound (e.g. 'cranky') and a surface condition (e.g. 'corroded'). If it was possible to capture those quantitatively, then a monitoring system could control them using a microphone and a camera. Such approach would require AI to identify those attributes from heterogeneous data (microphone and camera), and also a method to structure expert's rationale. While AI for heterogeneous data capture is a trending research area, there is fewer academic on structuring experts' diagnosis rationale. That is why improving existing methods to capture diagnosis expert knowledge is a relevant research challenge within the Industry 4.0 scope (Longo, Nicoletti and Padovano, 2019; Wan *et al.*, 2019).

This research proposes ontology-based reporting and monitoring methods for expert diagnosis knowledge capture and re-use. This proposal includes the following contributions:

1. An ontology to represent expert diagnosis activities using failure modes and quantitative and qualitative measures similarly to monitoring rationale.

2. A cloud-based reporting method using ontology-inferred forms to capture expert diagnosis knowledge as structured maintenance reports.

3. A cloud-based monitoring method using real-time ontology inferencing to generate control rules from incoming expert reports.

This research aims to demonstrate that expert diagnosis knowledge can be captured with sufficient structure and precision to be re-usable by monitoring systems to improve efficiency of fault-finding tasks.

The rest of this chapter's structure is as follows. Section 3.2 presents a literature review on ontology-based expert diagnosis methods to detect current research gaps. Section 3.3 describes the methodology employed to identify, develop and validate this research's proposal. The above-mentioned contributions are described in Section 3.4. Section 3.5 presents these contributions validation using two complex assets with mechanic, electric and electronic NFF scenarios as cases of study. Validation methods include literature comparison, expert interviews and efficiency experiments. Their results are analysed and discussed in Section 3.6. Finally, Section 3.7 presents conclusions and future works. It covers the implications of expert knowledge re-use to enhance monitoring and other benefits of ontology-based approaches to improve heterogenous data exchange in Industry 4.0.

3.2 Literature review

Failure diagnosis can be described as the process to determine the causes of failures or abnormal behaviours for resolving unexpected or undesirable conditions of assets (Medina-Oliva *et al.*, 2014). In other words, it is a process conducted by maintainers for proposing and testing hypotheses about the symptoms and causes of failures in assets, until finding their root causes. Failure diagnosis involves concepts from several knowledge domains (e.g. failure modes, equipment disassembly) (Zhou, Yu and Zhang, 2015).

Ontologies can be really helpful for failure diagnosis applications (Rajpathak and Chougule, 2011) because of their ability to declare information and its relations about those domains. A relevant aspect mentioned in failure diagnosis ontologies literature is data granularity (Zhou, Yu and Zhang, 2015). Data granularity provides a qualitative measure to describe the level of detail to which information is described. In ontologies, data granularity can be described by the number of properties and relationships given for a class (Rajpathak and Chougule, 2011).

Academic literature presents a varying range of failure diagnosis ontologies and ontology-based methods, applied to diverse maintenance operations (e.g. decision support, prediction, etc.) and case studies (e.g. rockets, factories, etc.).

3.2.1 Knowledge domains in ontology-based diagnosis methods

An interesting aspect of failure diagnosis ontologies is the variety of knowledge domains they cover. Failure diagnosis includes not only generating hypothesis on possible failure modes that derive from a symptom, but also testing them (Zhou, Yu and Zhang, 2015). Hence, diagnosing also involves other maintenance operations such as assembly or repair that embrace other knowledge domains (Rajpathak and Chougule, 2011) (e.g. components geometry). Besides, different purposes of failure diagnosis ontologies may consider varying aspects of diagnosis. For example, classification of alerts to increase accuracy of maintenance planning need to evaluate the criticality of failures found (Bekkaoui *et al.*, 2017). As part of this research, the student has identified common knowledge domains included within failure diagnosis ontologies. These are the following:

1. Diagnosis techniques: this knowledge domain includes different activities conducted by maintainers to identify set of components' faults that generate failures. These faults are commonly classified as symptoms, traces or causes according to their time of identification within the diagnosis activity (Zhou, Yu and Zhang, 2015). Most papers used FMEA as method to describe links between faults, causes, and components (Rajpathak and Chougule, 2011; Medina-Oliva et al., 2014; Zhou, Yu and Zhang, 2015; Zhong et al., 2018). Others also established differences between the phenomena that produce the component's fault and its effect on system's functionality (Bekkaoui et al., 2017; Dibowski, Holub and Rojíček, 2017; Xu et al., 2018). A common trend is to use unstructured text to describe faults, causes and effects (Rajpathak and Singh, 2014). Besides, there is little mention about expert performance, including steps about unsuccessful hypothesis, or safety procedures.

2. Failure modes: failure modes refer to the physical phenomena that make components fail (Nuñez and Borsato, 2017). Some ontologies represented the relations between failure modes and relevant component parameters (Akbari et al., 2010; Khadir and Klai, 2010; Ebrahimipour and Yacout, 2015; Dibowski, Holub and Rojíček, 2017). Instead, others classified them according to their phenomenology (Zhou, Li and Zuo, 2009; Wang, Qin and Hu, 2012; Medina-Oliva et al., 2014; Zhou, Yu and Zhang, 2015; Mishra and Thaduri, 2016). For both cases, data granularity achieved was high at the expense of a higher dependency to case studies (e.g. electric or mechanic). Nevertheless, only those papers focused on fault propagation for specific systems managed to identify ontology-based rules for fault's control thresholds in components (Bekkaoui et al., 2017; Dibowski, Holub and Rojíček, 2017; Ferrari, Dibowski and Baldi, 2017; Behravan, Meckel and Obermaisser, 2019).

3. Sensor data: another knowledge domain present in most ontologies reviewed is sensor data and component-related parameters. Most authors aimed to utilise such data to establish thresholds for fault analysis (Akbari et al., 2010;

Ebrahimipour and Yacout, 2015; Bekkaoui et al., 2017; Dibowski, Holub and Rojíček, 2017; Ferrari, Dibowski and Baldi, 2017; Behravan, Meckel and Obermaisser, 2019). While others used it to demonstrate the usefulness of data integration for knowledge re-use (Rajpathak and Chougule, 2011; Medina-Oliva et al., 2014; Xu et al., 2018; Zhong et al., 2018). Nonetheless, only those papers that accessed real-time sensor data through their case studies used it. Besides, there was little evidence on possible qualitative or quantitative measures that could be provided by experts through direct measurements or subjective perception.

3.2.2 Applications of ontology-based diagnosis methods

In maintenance research, failure diagnosis ontologies have contributed to declare and record diagnosis-related information for different purposes. These purposes vary accordingly to case studies and maintenance operations researches focused on planning and prognosis. Failure diagnosis ontologies can be classified in two main groups:

1. Ontologies for data modelling: some authors utilised ontologies as models to represent the actual behaviour of existing systems. Such models worked as the basis for more advanced applications that go beyond diagnosis. For example, Ferrari et al. (2017) proposed a generic ontology-based model of ventilation systems to develope a probabilistic algorithm that determines uncertainties of interconnections for analysing fault propagation. A similar approach was taken by Dibowski, Holub and Rojíček (2017). They proposed SWRL (Semantic Web Rule Language) rules to analyse connections between the components of a system according to their main parameters and identify failure thresholds. Akbari et al. (2010) proposed an ontology model for power transformers and their failure modes and utilised it to build an artificial neural network algorithm to analyse dissolved gasses and identify their concentration values, which may produce failures in transformers. Also in fault propagation, Behravan, Meckel and Obermaisser (2019) proposed an ontology-based model of cars to analyse potential failures of interconnected electronic systems using

directed acyclic graphs. Other authors have used ontology-based models for other maintenance purposes. An example is given by Mishra and Thaduri (2016), who presented an ontological model of a roller and its context to enhance continuous monitoring by linking the roller's main parameters and relevant contextual features. Ontological models have also been used for maintenance planning. Bekkaoui et al. (2017) proposed an ontological model on a manufacturing plant that relates diagnosis tasks with repair works and triggering events. So, such models can be used to classify the different alarms present in the plant's monitoring system to improve maintenance planning according to the criticality of past similar works.

2. Ontologies for data capture: the most extended use of failure diagnosis ontologies is related with data capture. It focuses on capturing and organising sensor data to enhance diagnosis processes. Different authors have proposed failure diagnosis ontologies for diverse assets like aircrafts (Wang, Qin and Hu, 2012), wind turbines (Zhou, Yu and Zhang, 2015) or pneumatic systems (Ebrahimipour and Yacout, 2015). Using these to collect data from historical procedures, they then utilised SWRL rules to recommend previous experiences that can be relevant to diagnose a given failure. More advanced ontologies also proposed to include contextual data to support failure diagnosis decisions. That is the case of Rajphatak and Chougule (2011), who proposed an ontology development method to link failure diagnosis with other relevant scenarios such as equipment operation and repair. Besides, Medina-Oliva et al. (2014) considered as relevant same assets under different conditions. That is why they proposed a fleet-wide ontology for predictive diagnosis using SWRL rules that could help experts to identify equivalent failure cases in similar assets. On the other side, there are other authors who proposed more advanced applications to support maintainers in their diagnosis decision-making process. For example, Bekkaoui, Karray and Sari (2015) included experience models as part of their ontology to classify reported failures and determine the most suitable experts to diagnose new cases. However, the most common approach within advanced applications are case-based reasoning systems. Authors like Khadir and Klai (2010), Dendani, Khadir and Guessoum (2012), or Wang et

al. (2010) proposed case-based reasoning systems to suggest potential failures that could be occurring based on similarity of symptoms from past experiences. These approaches have found challenging due to the structure of data being collected. Low granularity levels make ontologies to record unstructured text difficult to classify (Zhong *et al.*, 2018). That is why latest advancements in failure diagnosis ontologies proposed applications for case-based reasoning systems utilising text mining methods. Authors like Zhong et al. (2018), Xu et al. (2018), Zhou et al. (2009), and Rajphatak and Singh (2014) proposed different text mining algorithms to classify unstructured text from reports, avoiding limitations of ontologies in data granularity.

3.2.3 Research gaps in ontology-based diagnosis methods

Subsection 3.2.1 summarised the knowledge domains presented in ontology-based literature for maintenance diagnosis. These included diagnosis techniques, failure modes and sensor data for a varying range of assets (e.g. aircrafts, factories, etc.). Failure modes and sensor data are more dependent on applied cases studies (Akbari *et al.*, 2010; Behravan, Meckel and Obermaisser, 2019), while diagnoses techniques were similarly declared by different ontologies (Rajpathak and Chougule, 2011; Medina-Oliva *et al.*, 2014; Rajpathak and Singh, 2014; Zhou, Yu and Zhang, 2015; Bekkaoui *et al.*, 2017; Dibowski, Holub and Rojíček, 2017; Xu *et al.*, 2018; Zhong *et al.*, 2018). Besides, data granularity was also found to be dependent on the knowledge domains covered. While sensor data and failure modes achieved higher data granularity (Akbari *et al.*, 2010; Dibowski, Holub and Rojíček, 2017), diagnosis techniques had lower levels and incurred on extended usage of unstructured text (Medina-Oliva *et al.*, 2014; Rajpathak and Singh, 2014; Zhong *et al.*, 2018) for achieving higher detail on the activities described.

Subsection 3.2.2 summarised the range of applications (data modelling and capture) proposed for ontology-based diagnoses methods. Data modelling applications were more specific, while data capture applications aimed to generally characterise diagnosis activities. Besides, literature evidence suggested a relation between the detail level to describe these steps and the usage of unstructured text (Rajpathak and Singh, 2014). Although variable in scope such as fleet-wide (Medina-Oliva *et al.*, 2014) or expert features (Bekkaoui, Karray and Sari, 2015), most authors focused their efforts to describe diagnosis activities as the series of fault components that comprise a failure path, from its initial symptom to its root cause.

The student found little academic evidence on ontologies that described maintainers' rationale for diagnosis activities. This can be considered a relevant research gap for various reasons. First, it can be useful to improve data granularity of diagnosis activities (Ebrahimipour and Yacout, 2015; Wan *et al.*, 2019) and reduce unstructured text (Roy *et al.*, 2016; Angelopoulos *et al.*, 2020) in current applications (e.g. experts support (Medina-Oliva *et al.*, 2014) or prognosis (Dibowski, Holub and Rojíček, 2017)). Second, capturing maintainers' diagnosis rationale can be also useful to enhance other, less-researched, maintenance diagnosis applications. One of them is condition monitoring. Although ontologies have been used in data modelling for enhancing monitoring, there has been little research on re-using experts' knowledge to model monitoring systems (Ferrari, Dibowski and Baldi, 2017). That can have a positive impact on efficiency of fault-findings tasks that monitoring systems support. It would allow to capture experts' knowledge on failures conditions and embed them in monitoring systems. Thus, bridging the gap for integrated data management in Industry 4.0. Hence, this research aims to propose ontology-based diagnosis reporting and monitoring methods to capture experts' diagnosis knowledge and demonstrate its impact to improve efficiency of fault-findings tasks supported by monitoring systems.

3.3 Methodology

This research aims to prove that experts' rationale in diagnosis activities can be captured and re-used to enhance condition monitoring. In order to do so, this chapter presents three contributions: (1) an ontology to describe such rationale, (2) an ontology-based, reporting tool for experts to describe diagnosis activities, and (3) an ontology-based monitoring tool that uses real-time inferencing to identify monitoring rules. Inspired by similar works (Rajpathak and Singh, 2014; Renu *et al.*, 2016; Nuñez and Borsato, 2017), this research applies Design Science Research (DSR) (Peffers *et al.*, 2008) a well-established methodology in ontology literature (García-Peñalvo *et al.*, 2012; Gong and Janssen, 2013). DSR methodology includes the following steps:

1. Identify objectives: "Define the specific opportunity and justify the value of a solution". This opportunity was identified from research gaps in a literature review, presented in Section 3.2.

2. Design solution: "Create a solution to satisfy the research opportunity". This research applied the NeON methodology (Suárez-Figueroa, Gómez-Pérez and Fernández-López, 2015) for ontology development. The resultant ontology and subsequent methods are described in Section 3.4.

3. Demonstrate solution: "Prove the use of the solution to solve the research opportunity". This research's contributions were built as software tools and implemented in two different cases of study for further experimentation. These are presented in Section 3.5.

4. Validate achievements: "Measure the impact of the solution in the research opportunity". To validate this research contributions, several experimental methods were applied to two cases of study. These methods include ontology structural analysis, expert interviews, usability surveys and efficiency experiments. Section 3.5 describes their protocols, while their results are discussed in Section 3.6. Finally, Section 3.7 presents this research's conclusions and suggests future works.

3.4 Ontology-based expert diagnosis reporting and monitoring

This research proposes ontology-based reporting and monitoring methods to capture and re-use expert diagnosis knowledge for improving efficiency of fault-findings tasks. These methods consist of the following:

1. An ontology that describes expert diagnosis activities and links them with asset conditions in the form of quantitative measures.

2. A cloud-based reporting method that uses ontology-inferred forms for capturing experts' knowledge as diagnosis reports.

3. A cloud-based monitoring method using real-time ontology inferencing to generate control rules from incoming expert reports and monitor them using latest sensor data.

Figure 3-1 presents the approach in which these methods are based upon. The reporting and monitoring methods work independently, using the ontology to infer the necessary data.

The reporting method (Figure 3-1 – left) works on demand. Experts can 'request' to generate a new report whenever they encounter a new failure. A report consists of a series of 'forms', which include data fields for experts to input their diagnosis actions. 'forms' are created using the ontology's schema as a template, using ontology's classes, attributes and relationships. When a form is completed, experts can 'submit' these to the ontology's knowledge base. If the expert reported a new form to be created (requesting for a new ontology's individual), then a new 'request' will be made, and a new form inferred. Thus, enabling navigation between 'forms' through ontology's relationships, until the report is completed.

The monitoring method (Figure 3-1 – right) works on real-time streaming. For a given asset, the method first infers the 'rules' to monitor the sensors controlling it. Then, it infers if any given 'expert' reports identify further logic for sensor control. Finally, it infers the latest 'sensor' data to be monitored. This process is repeatedly in loop, dropping inferred knowledge from previous loops to avoid multiplicity of comparisons. Besides, a separate loop is run by sensors to 'upload' data to the knowledge base as new individuals for every new read.

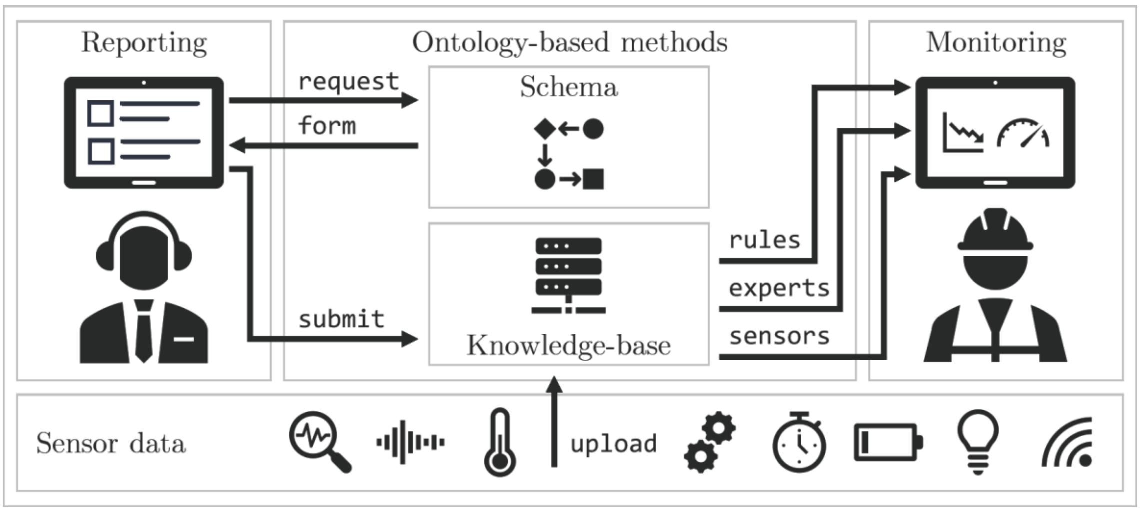


Figure 3-1. Overview of the proposed ontology-based expert diagnosis reporting and monitoring methods – It briefly describes data retrieval and inferencing steps to capture and re-use expert diagnosis knowledge in fault-finding tasks.

Following subsections explain each proposed method in-depth. First, Subsection 3.4.1 presents the ontology for expert diagnosis and condition monitoring logic based on asset's condition quantification. Then, Subsection 3.4.2 describes the inferencing rules proposed to create reporting forms based on ontology's elements. Finally, Subsection 3.4.3 explains the inferencing method proposed to convert expert reports in monitoring rules using sensor data to evaluate asset's condition.

3.4.1 Diagnosis rationale ontology (Diagont)

Data structures that can describe procedural knowledge in elementary steps are capable of linking that knowledge with quantitative measures. That can solve the current research gap of integrating experts' diagnosis logic in monitoring systems. The proposed diagnosis rationale ontology (diagont) was designed following the NeOn methodology (Suárez-Figueroa, Gómez-Pérez and Fernández-López, 2015). That enabled to utilise relevant academic literature and subject-matter expert opinions to declare relevant knowledge domains. First, the student conducted a literature review to identify relevant papers and specify their knowledge domains.

Second, this evidence served the student to build a conceptual ontology model. Finally, the student conducted expert interviews to refine the latter into a formal ontology model. Figure 3-2 presents the ontology model through its classes and relationships. It aims to explain experts' rationale for diagnosis and relate it to the logic of monitoring tools. The ontology's attributes are listed in Table 3-2.

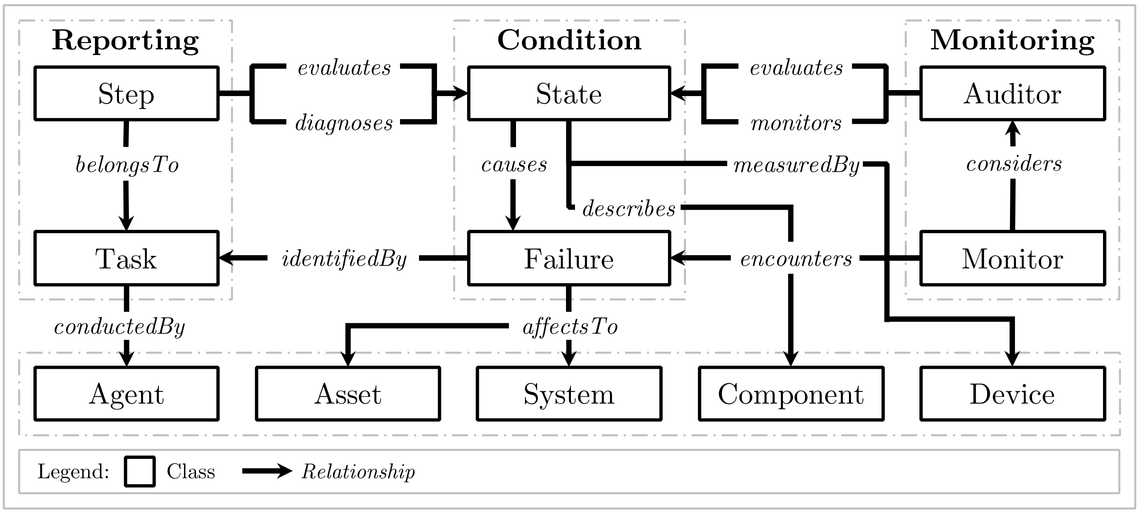


Figure 3-2. Depiction of class and relationships of diagont's ontology schema.

The ontology's classes that explain experts' diagnosis rationale are: *'Task'*, *'Step'*, *'Failure'* and *'State'*[2]. A *'Task'* summarises the procedure *'conductedBy'* an *'Agent'* for identifying a *'Failure'*. A *'Failure'* is the malfunctioning or unexpected behaviour of an *'Asset'*. It can be described through the set of faulty *'States'* of *'Components'* that *'causes'* it, from its initial symptoms to its root cause. Each of these *'States'* that conform a *'Failure'* are identified through 'Steps', which are the series of activities that *'belongTo'* a *'Task'*. When conducting a *'Step', an* *'Agent'* *'evaluates'* the current *'State'* of a *'Component'* and *'diagnoses'* it against a past *'State'* of the same *'Component'* which is known to be faulty. Thus, being able to demonstrate that the current *'State'* is also a fault. This comparison process is used by experts while diagnosing to demonstrate two things: (1) the current 'State' is fault, and (2) the current 'State' belongs to a 'Failure' path. Hence, the expert can report not only these faulty *'States'* but also the rationale used to ensure that such *'State'* is a fault.

[2] Words written in *'italic'* are those that represent classes and relationships in Figure 3-2.

The ontology's classes that represent the condition monitoring logic are: *'Monitor'*, *'Auditor'*, *'Failure'* and *'State'*. A *'Monitor'* is a control entity that aims to 'encounter' a *'Failure'*. In order to do so, it *'considers'* a series of *'Auditors'* that *'monitor'* different *'States'* of *'Components'*. Besides, the *'Auditors'* also *'evaluate'* current *'States'* against the monitored 'States'. These evaluations are made with a pre-specified logic (*'hasComparison'*). If those evaluations are all true, then it can be said the *'Monitor'* *'encounters'* a *'Failure'*.

Some of the ontology's attributes, such as *'hasComparison'*, are utilised by the reporting and monitoring methods to infer knowledge. These are presented in Table 3-1. As it can be seen, some specific datatypes have been determined in order to avoid extensive use of unstructured text in diagnosis reporting. These datatypes and their sets of values are listed in Table 3-2. The complete schema of diagont can be found at 10.17862/cranfield.rd.12279152. The inferencing rules for the reporting and monitoring methods are explained the following subsections.

Table 3-1. Depiction of classes and attributes of diagont's ontology schema.

Task		Failure		Monitor	
hasDescription	*string*	hasDescription	*string*	hasDescription	*string*
		hasImpact	*impact*		
		hasDomain	*domain*		
		hasPhenomenon	*phenomenon*		
		hasImage	*anyUri*		
		hasAudio	*anyUri*		
Step		**State**		**Auditor**	
isCritical	*boolean*	hasStatus	*status*	isValidated	*boolean*
isContributory	*boolean*	hasDomain	*domain*	hasComparison	*comparison*
hasObject	*object*	hasPhenomenon	*phenomenon*		
hasMethod	*method*	hasMeasureValue	*double*		
hasComparison	*comparison*	hasMeasureUnit	*unit*		
		hasMeasureDate	*date*		

Legend: Class | Attribute | *Datatype*

Table 3-2. Summary of diagont's proprietary attributes datatypes.

impact	status	domain	phenomenon		
local	*normal*	*mechanics*	*fracture*	*thermal shock*	*signal error*
global	*safely degraded*	*electrics*	*fatigue*	*thermal runaway*	*error*
	unsafely degraded	*electronics*	*corrosion*	*short circuit*	*material*
	faulty	*hydraulics*	*impact*	*open circuit*	*process*
		pneumatics	*blockage*	*electric loss*	
		humanics			
object	method	comparison	unit		
symptom	*inspect*	*equal to*	*metre*	*pascal*	*hertz*
trace	*measure*	*not equal to*	*degree*	*joule*	*watt*
cause	*repair*	*greater than*	*kilogram*	*mol*	*ampere*
	replace	*less than*	*second*	*kelvin*	*volt*
		less than or equal to	*newton*		*ohm*
		greater than or equal to			

Legend: datatype | *datavalue*

3.4.2 Ontology-based expert diagnosis reporting

The proposed ontology aims to describe experts' rationale for capturing their knowledge on diagnosis activities, including failure modes and conditions. Such knowledge capture process requires tools that allow experts to report diagnosis activities in real-time and minimising interference with them. Hence, ontology editing software (e.g. Protégé) may not be a feasible alternative, due to the complexity of interfaces and the need of ontology modelling experience. Instead, this research proposes a cloud-based reporting method to create web forms using ontology's classes as templates. Its aim is to create web forms that replicate the ontology's rationale to reduce intrusions on diagnosis activities being conducted.

Figure 3-3 details the reporting method proposed and its inferencing rules in SWRL notation. As previously explained, this reporting method creates forms on user's request. A user's request consists of an ontology class. In order to select it, a user can navigate through ontology's classes available on the cloud. These are organised hierarchically in menus using the SWRL rule "rdfs:subClassOf(?c)" (Figure 3-4). On request, the method infers attributes ("owl:DatatypeProperty(?a)") and relationships ("owl:ObjectProperty(?r)") asserted to the requested class ("rdfs:domain(?a,c)"). Besides, the method identifies all individuals asserted by all relationships retrieved ("?r(?x,?i)"). So, the user can either create new individuals or refer to existing ones while completing the form. With all this inferred data, the web template generates a form to be filled by the user (Figure 3-4). Once it is submitted, the reporting method identifies those individuals declared by the user as new. If any exists, then the method automatically requests their classes for new forms to be created in separate tabs. Thus, smoothing the navigation between forms within the report.

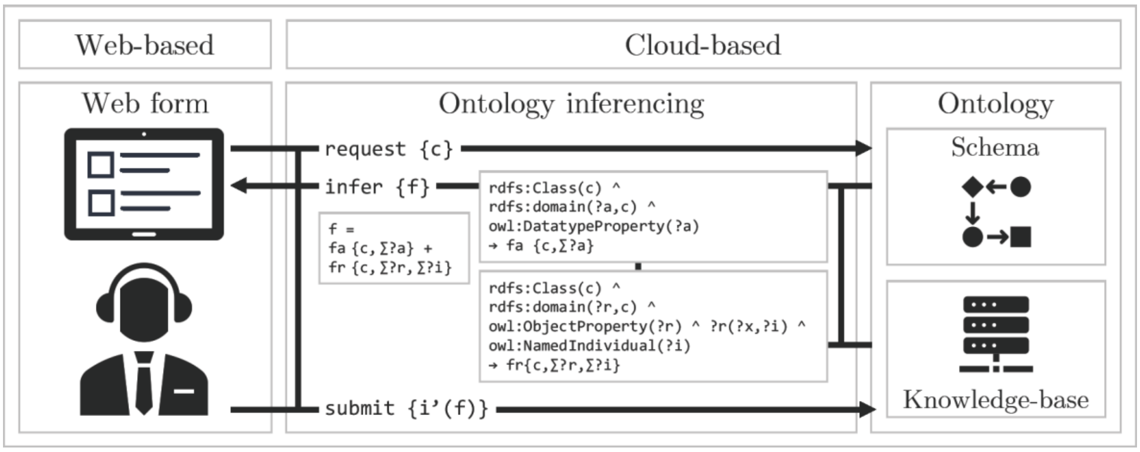


Figure 3-3. Ontology-based expert reporting method and its inferencing rules – Inferencing rules are presented in the form of SWRL rules utilised to identify datatype and object properties asserted to a given class, and individuals asserted through identified object properties for creating a reporting form upon user's request.

Figure 3-4 presents a simple example of resulted forms to report a simple diagnosis task. It shows a report that includes a *'Task'* and a *'Step'*. According to diagont's rationale (Section 3.4.1), an expert should report a diagnosis activity using this method as follows. First, report a *'Task'* about the diagnosis. Then, report one *'Step'* per fault (*'symptom'*, *'trace'* or *'cause'*) identified during the diagnosis. For each *'Step'*, declare the *'States'* (*'evaluated'* and *'diagnosed'*) used to identify it. Once the root cause is identified as *'Step'*, report the *'Failure'* found.

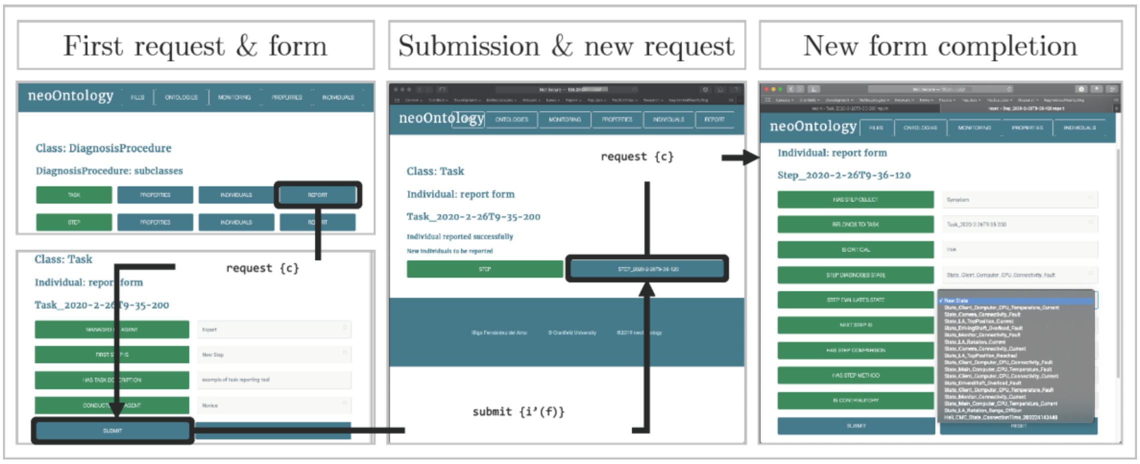


Figure 3-4. Pictorial demonstration of the ontology-based expert reporting tool – It shows several screenshots of different reporting steps as explained in Figure 3-3.

For each form, the user can declare all the attributes (Table 3-2) and relationships (Figure 3-2) asserted to the specific class in diagont, using these relationships to navigate between forms. The reason to use generic SWRL rules ("owl:" and "rdfs:") instead of specific ("diagont:") is such so the method can generate forms for any given ontology. Besides, the reporting method can identify when an attribute or relationship has not been asserted in a submitted form. Although it does not return it as an error in order to maintain the ontology's open-world assumption.

The following section explains how these diagnosis reports can be used to enhance condition monitoring systems with expert recommendations.

3.4.3 Ontology-based real-time monitoring and expert recommendations

Condition monitoring refers to the actions taken to identify abnormal behaviours of physical assets (Jardine, Lin and Banjevic, 2006). These actions normally consist of the following: data acquisition, data processing and decision-making. Monitoring methods proposed by literature are mostly data-driven or analytical with a smaller proportion of knowledge-based approaches (Angelopoulos *et al.*, 2020). It is also widely accepted that not "a single method is satisfactory in every respect" (Abele *et al.*, 2014). Most methods are quite specific to the monitored asset (e.g. wind-turbines or production plants) and limited in knowledge re-use

(Bekkaoui *et al.*, 2017; Liu *et al.*, 2019). Hence, a method that can combine the abovementioned approaches can help to resolve these challenges.

This research proposes a monitoring method that can combine data-driven, analytical and knowledge-based approaches. Its aim is to re-use expert knowledge from diagnosis reports by converting it into rules that can be monitored as those from data-driven or analytical approaches. This conversion is done through ontology inferences previously to the actions taken by monitoring methods: data acquisition, data processing and decision making.

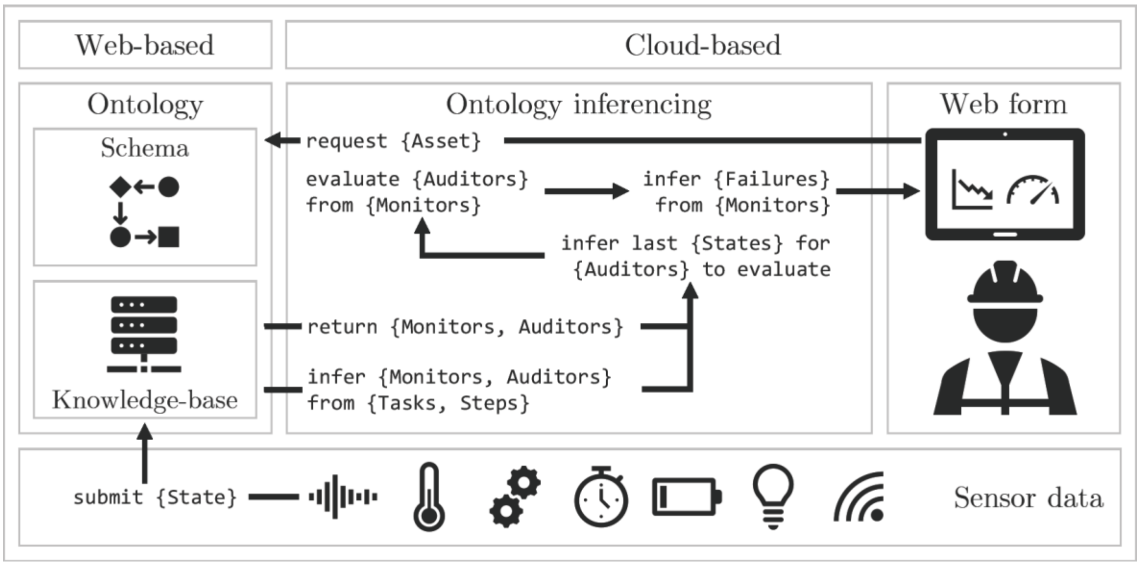


Figure 3-5. Ontology-based monitoring with real-time sensor control and experts' recommendations – It presents the data retrieval and inferencing steps utilised to infer monitoring rules from expert diagnosis reports and merge them with data- and analytical-driven monitoring approaches. These inferencing steps are described as SWRL rules in Figure 3-6.

Figure 3-5 presents the actions taken by the proposed monitoring method. First, this method returns the existing *'Monitors'* and *'Auditors'* for a requested *'Asset'*. Besides, it retrieves *'Tasks'* and *'Steps'* reported for the same *'Asset'* and infers additional *'Monitors'* and *'Auditors'* from those using a SWRL rule (Figure 3-6 – Monitor extraction). Thus, re-using experts' knowledge from diagnosis reports. Then, the method infers (Figure 3-6 – State identification) the latest *'States'* to be *'evaluated'* by each asserted or inferred *'Auditor'*. These *'States'* can be added either manually through expert reports or automatically by *'Devices'*. Once the *'evaluation'* and *'monitoring' 'States'* for each *'Auditor'* have been obtained, the method evaluates the resultant control monitoring rule (Figure 3-6 – Auditor evaluation). Finally, each *'Monitor'* is evaluated in order for the monitoring method to suggest a decision regarding the *'Failures'* being controlled. *'Monitors'* are evaluated using different SWRL rules (Figure 3-6 – Monitor evaluation) depending on the number of *'Auditors'* they *'consider'*. As a result, the web form presents the different evaluation results (Figure 3-7) classified as monitoring (data-driven or analytical rules) or expert recommendations (inferred from expert reports). These results visualise the decisions regarding the existence of failures, the evaluations made (at *'Auditor'* level), and the suggestions for maintenance.

Infer {Monitors, Auditors} from {Tasks, Steps}:

```
Failure(?a) ^
Task(?b) ^
Step(?c) ^
State(?d) ^
identifiedBy(?a,?b) ^
belongsTo(?c,?b) ^
isContributory
    (?c,true) ^
diagnoses(?c,?d)
→
Auditor(?c) ^
Monitor(?b) ^
considers(?b,?c) ^
monitors(?c,?d)
```

Infer last {States} for {Auditors} to evaluate:

```
Auditor(?a) ^
State(?b) ^ State(?c) ^
Device(?d) ^
Unit(?e) ^ Unit(?f)
monitors(?a,?b) ^
measuredBy(?b,?d) ^
measuredBy(?c,?d) ^
hasMeasureUnit(?b,?e) ^
hasMeasureUnit(?c,?f) ^
hasMeasureDate(?b,?g) ^
hasMeasureDate(?c,?h) ^
swrlb:equal(?e,?f)
swrlb:greaterThan(?h,?g)
→
sqwrl:select(?c,?h) ^
sqwrl:orderBy(?c,?h) ^
swrlb:first(?c) ^
evaluates(?a,?c)
```

Evaluate {Auditors} from {Monitors}:

```
Auditor(?a) ^
hasComparison
    (?a,?comparison) ^
evaluates(?a,?b) ^
hasMeasureValue(?b,?c) ^
hasMeasureUnit(?b,?d) ^
hasMeasureDate(?b,?e) ^
monitors(?a,?f) ^
causes(?g,?h) ^
hasMeasureValue(?f,?h) ^
hasMeasureUnit(?f,?i) ^
hasImpact(?f,?j) ^
hasDomain(?f,?k) ^
hasPhenomenon(?f,?l) ^
swrlb:?comparison(?c,?h) ^
swrlb:equalTo(?d,?i)
→
Failure(?newfailure) ^
resultIs(?a,?newfailure) ^
hasImpact(?newfailure,?h) ^
hasDomain(?newfailure,?j) ^
hasPhenomenon
    (?newfailure,?k) ^
hasDate(?newfailure,?e)
```

Infer {Failures} from {Monitors}:

Single auditors:

```
Monitor(?a) ^
Auditor(?b) ^
isUnique(?b,true) ^
resultIs(?b,?c)
→
encounters(?a,?c)
```

Multiple auditors:

```
Monitor(?a) ^
Auditor(?b) ^
isUnique(?b,false) ^
Auditor(?c) ^
isUnique(?c,false) ^
considers(?a,?b) ^
considers(?a,?c) ^
resultIs(?b,?d) ^
resultIs(?c,?e) ^
swrlb:equalTo(?d,?e)
→
encounters(?a,?d)
```

Figure 3-6. Inferencing rules for auditor evaluation and expert reports knowledge extraction – These are presented in the form of SWRL rules referencing to the data inferencing steps described in Figure 3-5.

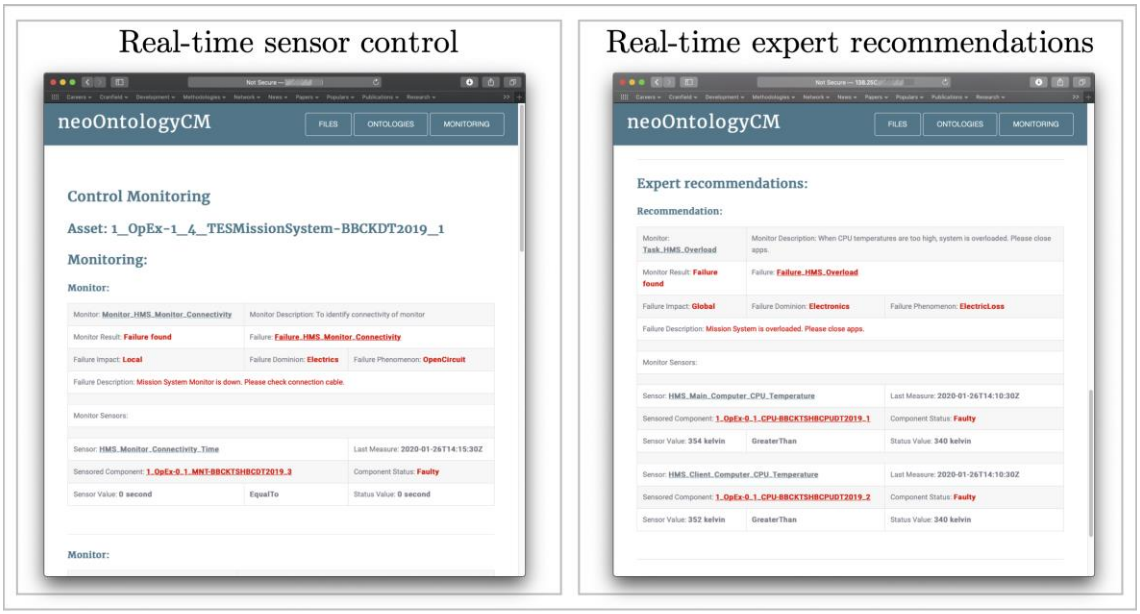


Figure 3-7. Pictorial demonstration of the ontology-based expert monitoring tool.

The proposed method is real-time and works in cycles of a pre-determined periodicity. Figure 3-8 presents an example of the inferencing results for "monitor extraction" from expert reports (Subsection 3.6.3). It shows the proposal's real-time implementation and a Protegé implementation used to confirm the real-time implementation's results. The Protegé implementation of diagont's ontology can be consulted at 10.17862/cranfield.rd.12279152. The abovementioned periodicity is set as the refreshing time of the web form that triggers the inferencing algorithm (Figure 3-7). So, it can be accommodated to the data rate required by the different sensors implemented in each case of study. Besides, the *'Auditors'* and *'Monitors'* inferred from *'Tasks'* and *'Steps'* are not asserted to diagont's knowledge-base after each cycle. This ensures that data is not duplicated to avoid consistency issues. Nevertheless, the monitor extraction inference step (Figure 3-6) may cause delays if the required refreshing time is too small. For those cases, another approach towards data consistency could be taken by to deleting inferred *'Tasks'* and *'Steps'*. Such approach has not been implemented due to other potential re-use applications of such knowledge (e.g. recommendation for novices' diagnosis reporting).

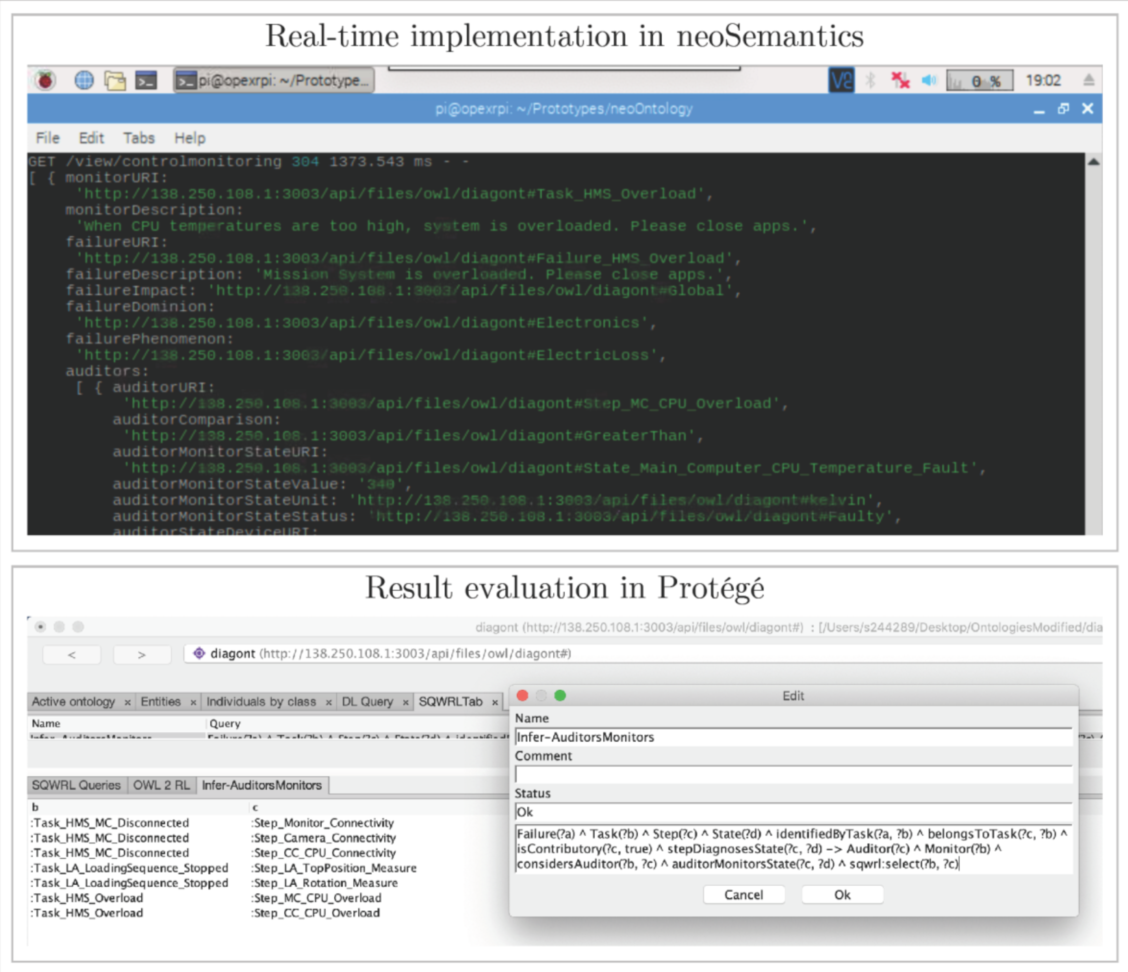


Figure 3-8. Depiction of inferencing results to extract 'Auditors' and 'Monitors' from 'Steps' and 'Tasks' from real-time implementation using neoSemantics and validation in Protegé.

The proposed ontology and monitoring and reporting methods have been implemented in a system prototype to evaluate their validity in two different case studies. This system implementation, case studies and the experimental protocol and its results are explained in the following section.

3.5 Validation protocol

3.5.1 System implementation

The proposed solution was implemented within a prototype system for experimentation. It comprises three components: (1) a cloud server for ontology storage and inferencing, (2) a web app for interfacing reporting and monitoring methods, (3) a web API for sensor data transfer. Figure 3-9 presents the languages and platforms utilised to code each component. The diagont ontology was designed in Protegé (Musen, 2015). Then, it was stored in Neo4j (Zhu, Zhou and Shao, 2019) using neosemantics (Barrasa, 2019) to separate the schema and the knowledge base. The cloud server was built with NodeJS (Surhone, Tennoe and Henssonow, 2010), using Cypher (Panzarino, 2014) and neosemantics to generate the SWRL rules. The web apps were built using EJS (Eernisse, 2015) to generate the HTML code. Finally, the web API for sensor data transfer was built using C++ (Plauger, 2002) to transfer sensor data to the cloud server by JSON objects.

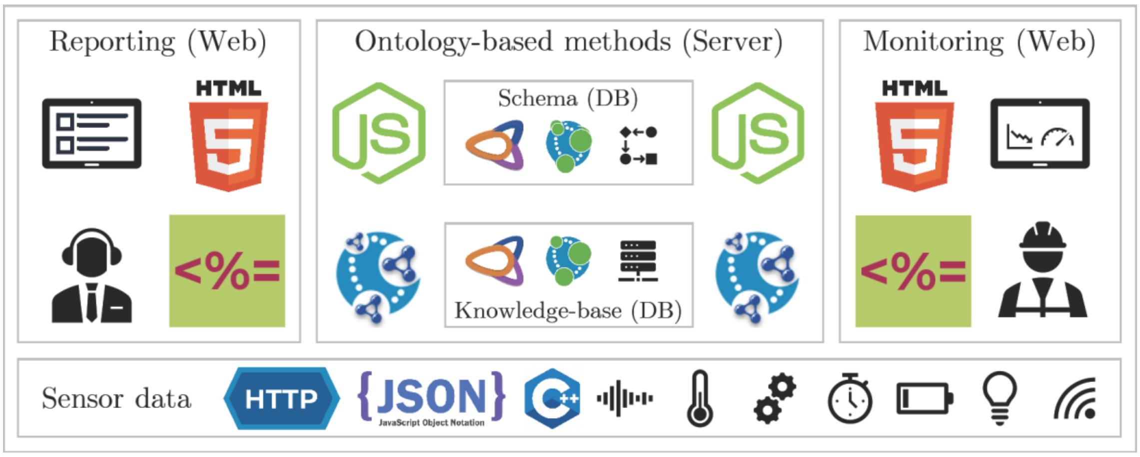


Figure 3-9. Overview of ontology-based methods implemented as a system prototype – It replicates the structure of Figure 3-1 to present the logos of each tool/language utilised to develop the prototype's components.

The proposed system prototype described above can be defined as a private cloud that provides Software-as-a-Service (SaaS) as described by (Jin *et al.*, 2010). The cloud server (1) executes ontology inferencing and storage operations, while the reporting, monitoring (2) and sensor data transfer (3) web services perform data collection and visualisation. The cloud server was built on a Raspberry Pi 3 Model B Plus using Raspbian 9 kept within the University's private network. There were two main reasons to choose this implementation as opposed to other alternatives such as Amazon Web Service's EC2 Infrastructure-as-a-Service (IaaS) or Microsoft Azure's Platform-as-a-Service (PaaS).

The first reason involves data security. This research made use of confidential data provided by the industrial sponsor, who required to keep it private and stored on British soil as per their regulations. Alternative cloud solutions such as IaaS and PaaS could meet the requirements. But they would have required to build additional security features (e.g. end-to-end encryption) into the prototype that were out of this research's scope.

The second reason relates to prototype development. The cloud server's prototype (1) was re-used not only for the research's contributions of Chapter 4 and Chapter 5, but also in consequent research works (Erkoyuncu *et al.*, 2020; Wali *et al.*, 2020) resulted from this PhD thesis discussed in Subsection 6.3.1. These required to share the cloud system prototype with researches and MSc students for conducting their work while keeping the security requisites explained above. The use of alternative cloud solutions (IaaS and PaaS) seemed a less conservative approach as modifications made by these individuals could compromise the prototype's security.

Overall, the decision to implement the cloud server prototype privately within the University's network was to ensure its security while allowing to share it for other research works.

3.5.2 Experiment design

This research validation aims to evaluate the proposed solutions' abilities to capture expert diagnosis knowledge and re-use it to improve efficiency of fault-finding tasks. That involves collecting and analysing data regarding criteria that can measure those abilities. For example, failure identification tasks' time and errors can serve to evaluate efficiency improvements given by monitoring tools. This validation's objectives include demonstrating the following hypotheses:

- The proposed ontology fairly represents experts' rationale for diagnosis tasks.
- The proposed reporting method enhances expert diagnosis knowledge capture.
- The proposed monitoring method has a positive impact on efficiency of fault-finding tasks.

According to similar research (Miguelañez *et al.*, 2009; Carrera *et al.*, 2014; Zhou, Yu and Zhang, 2015; McDaniel and Storey, 2019), there are quantitative and qualitative criterions considered appropriate to validate these hypotheses. Table 3-3 lists those utilised in this research validation. They are classified according to the proposed contribution they aim to validate, and the method utilised to collect and analyse data about them.

Table 3-3. Summary of validation methods employed in this research.

Contribution	Validation method	Quantitative criteria	Qualitative criteria
Ontology	Expert interviews	Schema changes	
	Structural analysis	OntoQA metrics	
Reporting	Usability tests		Cases of study (assets and failures)
	Usability surveys		Accuracy, completeness, etc.
Monitoring	Efficiency experiments	Time and errors	
	Usability surveys		Ease-of-use, visualisation

For these criterions to represent the proposed contributions' validity, there are some assumptions that must hold true:

- If an ontology describes a knowledge domain, then its schema represents the relations between its concepts and so its rationale. Hence, the proposed ontology schema can be evaluated about its ability to represent expert diagnosis rationale. Such evaluation can be done by comparison with similar ontologies and expert validation.

- Besides the ontology schema representing experts' rationale, the ontology's knowledge base must be able to explicitly capture experts' knowledge. A method to evaluate the validity of captured knowledge is to evaluate its ability to improve a certain task. Hence, it also seems relevant to evaluate each proposed method (reporting and monitoring) independently.

- The reporting method aims to capture expert knowledge on diagnosis for the monitoring method to use it. Besides evaluating capture knowledge's effect in monitoring-related tasks, the reporting method can have an impact on its usability. For example, if reporting is not intuitive enough, it can impact on reports' completeness and so completeness of captured knowledge. Thus, it seems relevant to validate the reporting method's usability with experts.
- The monitoring method aims to re-use expert knowledge to improve efficiency of fault-finding tasks. The main objective of monitoring tools is to identify abnormal asset behaviour (Jardine, Lin and Banjevic, 2006). Thus, evaluating its impact on fault-finding efficiency can be considered relevant to validate a monitoring method. If fault-finding tasks are given consistent quality, then completion time and errors can be a direct representation of their efficiency.
- Besides time and errors, the monitoring method's usability can also affect efficiency of fault-finding tasks' (e.g. misleading or not visual enough results). Hence, it seems relevant to evaluate that. Because usability is a subjective aspect of software tools, qualitative criteria based on testers' opinion seems more relevant for such evaluation.

Besides quantitative and qualitative validation criteria, Table 3-3 also lists the experimental methods used for data collection and analysis. These are presented in the following subsections.

3.5.2.1 Ontology expert interviews

The validity assessment of an ontology according to its intended purpose by its targeted users is a relevant step on its research validation (Aruna, Saranya and Bhandari, 2011; Bautista-Zambrana, 2015). These users can be represented by the experts that will use the ontology to report diagnosis activities. Prior to further experimentation with the reporting tool, experts were asked to state their opinion regarding diagont's ability to describe their diagnosis rationale. In order to collect those opinions, it seemed relevant to use semi-structured interviews, whose procedure and questionnaire are described in detail in Appendix A. These interviews followed the next protocol:

1. Presentation (10 minutes): interviewees were briefed in diagont, its purpose and how it describes diagnosis rationale.

2. Interview (40 minutes): interviewees were shown a presentation with all the classes, their attributes and relationships. Then, they were asked to propose changes for those to make them fit their reporting activities.

Interviewees were 9 subject-matters experts from two different maintenance organisations with at least 10 years of experience performing diagnosis activities. Their ages range from 30 to 60 years old and their roles included engineers, technicians and managers. They have previous experiences with other diagnosis reporting tools but no experience with ontologies nor ontology-based applications. The reason to select two different organisations was to reach cases of study of different natures (electro-mechanical and electronic).

3.5.2.2 Ontology structural analysis

Another relevant method to validate an ontology's fit for purpose is through evaluation of its structure in comparison with similar ontologies (Breitman, Casanova and Truszkowski, 2007; McGurk, Abela and Debattista, 2017). If an ontology represents a knowledge domain, then its schema describes the domain's rationale. Such logic can vary accordingly to the specific purpose for which an ontology is intended to (Breitman, Casanova and Truszkowski, 2007). Hence, comparing different ontologies' schemas from similar knowledge domains can validate an ontology against its intended purpose.

Several authors have reviewed existing ontology evaluation methods presenting quantitative metrics to compare ontologies according to specific objectives (García, García-Peñalvo and Therón, 2010; McGurk, Abela and Debattista, 2017). Most of them coincided that OntoQA is one of the most suitable tools for such aim, it metrics evaluate the "richness, width, depth and inheritance of an ontology schema" (Tartir and Arpinar, 2007). These metrics are:

- Relationship richness (RR): refers to the number of ontology relationships that are not hierarchical. These provide an understanding of the ontology's richness (Tartir and Arpinar, 2007). This means, the higher this number the easier the ontology is to describe an individual based on its relations with others, as follows:

$$Relationship\ Richness = RR = \frac{|P|}{|H| + |P|} \in [0,1] \tag{3-1}$$

where, P and H are the non- hierarchical and hierarchical relationships respectively.

- Attributes richness (AR): refers to the number of attributes, *att* per class, *C*. The higher this number, the more detailed an individual will be, but at the potential cost of higher levels of unstructured text (Tartir *et al.*, 2005), as follows:

$$Attribute\ Richness = AR = \frac{|att|}{|C|} \in \mathbb{R} \tag{3-2}$$

- Inheritance richness (IR): describes the ontology's width. The higher IR is, the more general the knowledge within an ontology is (Tartir and Arpinar, 2007).

$$Inheritance\ Richness = IR = \frac{\sum_{C_i \in C} |H^c\ C_1, C_i\ |}{|C|} \in \mathbb{R} \tag{3-3}$$

In order to evaluate these metrics for one specific ontology, it is necessary to compare them with those from other ontologies considering similar knowledge domains and purposes. The population of ontologies considered for evaluation is that from the literature review conducted for this research. From, a total of 35 papers found in this domain, a sample of 10 was selected based on the following exclusion criteria:

1. Ontologies that do not describe failure diagnoses activities.
2. Ontologies that do not describe failure phenomena and related measures.
3. Papers that do not include descriptions of ontology schemas.

The author utilised the remain ontologies schemas (Table 3-6) to calculate the abovementioned metrics.

3.5.2.3 Reporting usability tests

Besides ontology evaluation, this research validation has several requirements regarding the reporting tool. First, it is necessary for real-life experts to test the reporting tool for answering its usability questionnaires. Second, the monitoring efficiency experiments also require of cases of study that meet real-life conditions. In order to meet these two, the student decided to conduct usability tests for evaluating the proposed reporting tool. These tests were conducted under the following protocol:

1. Presentation: a brief presentation to introduce the reporting tool to testers. This included a description of the tool's operative and its different input-data formats.

2. Testing: testers were asked to complete two different reports on failures they have identified recently as part of their diagnosis work. Testers were handed over the reporting tool in a tablet device to evaluate it as they would use it in real-life conditions.

Testers were the same 9 subject-matter experts interviewed for the abovementioned schema validation. The reports collected as part of the testing procedure were later used to design the monitoring efficiency experiments. Reports data helped to identify the failures to be setup in the cases of study for experimental testers to perform failure identification tasks.

3.5.2.4 Reporting usability surveys

After proceeding with testing the reporting tool, real-life experts were asked to complete a questionnaire regarding its usability. The student considered this necessary to validate captured knowledge's quality. This seemed necessary in order to assess the ability of such knowledge to be re-used in condition monitoring. For example, if the tool's interface is very complex or its vocabulary is not clearly understandable, then the data collected may be incorrect or incomplete. Thus, having a negative effect on the monitoring tool's impact on condition monitoring efficiency.

A method to evaluate the captured knowledge's quality is through evaluation of the tool's usability from an ontological perspective (Brank, Grobelnik and Mladenić, 2005; Raad and Cruz, 2015). Several authors have discussed different validation criteria and ontological aspects to evaluate in ontology-based tools' validation (Gómez-Pérez, 2004; Brank, Grobelnik and Mladenić, 2005; Hlomani and Stacey, 2014; Raad and Cruz, 2015). Table 3-4 defines those that most authors agreed on identifying as the most relevant when validating usability of ontology-based tools (Vrandečić, 2009).

Table 3-4. Definition of relevant ontological aspects and validation criterions regarding ontology-based tools' usability as discussed by Vrandečić (2009)

Type	Name	Definition
Ontological aspect	Vocabulary	Refers to the language utilised to name ontology's classes, attributes and relationships. It considers the tool's ability to appropriately refer to the concepts described.
	Structure	Refers to the relations (taxonomical or semantical) between ontology's elements. It considers the tool's ability to navigate through the different ontology's classes (reporting forms).
	Context	Refers to the environment in which the ontology is being used. It considers the tool's ability to generate appropriate diagnosis reports using ontology elements to generate individuals.
Validation criterion	Accuracy	Refers to the ontology's ability to match the expert's knowledge domain. It evaluates whether the tool is capable of describing diagnosis reports similarly to expert.
	Completeness	Refers to the ontology's ability to fully cover the expert's knowledge domain. It evaluates whether the tool covers all necessary aspects to produce diagnosis reports.
	Conciseness	Refers to the ontology's ability to only cover the expert's knowledge domain. It evaluates whether the tool does not cover any unnecessary aspects to produce diagnosis reports.
	Consistency	Refers to the ontology's ability to not allow for any contradictions. It evaluates whether the tool can produce contradictions in the reports being produced.

These aspects and criterions are qualitative and so subject to opinion. Hence, the student found relevant to collect data about them from the real-life experts who acted as testers of the reporting tool. Questionnaires for data collection comprise several statements for each pair of validation criterions and ontological aspects.

This questionnaire, which can be found in Appendix B, asked experts to rank their agreement with its statements on a Likert Scale from 1 to 7. The reason to select such scale is based on the results presented by Weijters, Cabooter and Schillewaert (2010), who suggested that it maximises potential information transmission when surveying expert populations. Data collected was later utilised to analyse experts' opinion on the reporting tools usability.

3.5.2.5 Monitoring efficiency experiments

Experiments consisted of evaluating fault-finding tasks' efficiency to compare the impact of the proposed monitoring tool with respect to alternative solutions. Efficiency can be described by resources utilised for similar levels of effectiveness. With respect to failure identification tasks, it can be assumed that task's efficiency depends only in time and errors for a given failure and testers expertise levels.

In order to measure tasks' efficiency, experiments consist of identifying a given failure in an asset using a condition monitoring tool. Figure 3-10 presents the experimental setup, including the case study and the environment where the experiment took place. Additional screenshots of the monitoring tools (Table 3-5) utilised during the experiment can be seen in Figure 3-7. Besides, the case study failures (CNN and TEM) utilised for these fault-finding experiments are later described in Subsection 3.5.4. Figure 3-15 (CNN) and Figure 3-16 (TEM) present the relevant expert-recommended sensor values to identify the failure's root cause.

The quantitative variables used to measure efficiency are time and errors. Time can be defined as the number of seconds required by a tester from the experiment's start until the correct identification of the failure. Errors can be described as the number of incorrect failures hypothesised by the testers during the experiment.

Besides the abovementioned variables, there are several factors to consider. These factors are summarised in Table 3-5.

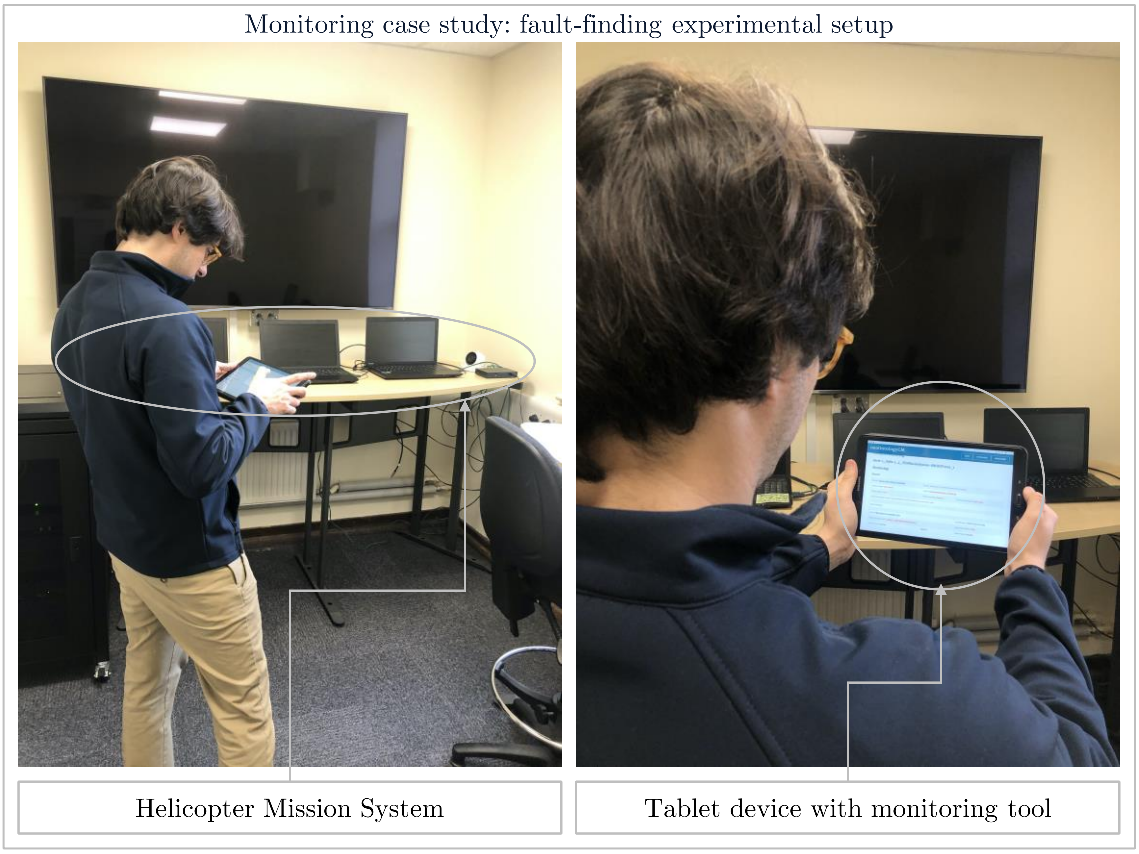


Figure 3-10. Description of monitoring efficiency experimental setup – The pictures present a tester with a tablet device looking at the monitoring tool to identify experimental failures, which are presented in Figure 3-15 and Figure 3-16.

Table 3-5. Definition of experimental effects, factors and their definitions from monitoring efficiency experiments.

Effect	Factor	Definition
Expertise level	IT	Tester has previous experience diagnosing electric and electronic failures
	NOIT	Tester has no previous experience diagnosing electric or electronic failures
Failure's nature	CNN	The nature of the failure is electric (Figure 3-15)
	TEM	The nature of the failure is electronic (Figure 3-16)
Monitoring tool	None	Tester has no support to diagnose the experimental failure
	KRD	Tester has support of a monitoring tool without expert knowledge
	KRE	Tester has support of the monitoring tool with expert knowledge

KRD = Knowledge Recommendations Disabled | KRE = Knowledge Recommendations Enabled

One factor is tester's expertise level. Unlike other experiments proposed in this PhD thesis, the student considered this factor relevant because the monitoring tool does not provide direct guidance of the tasks to be conducted. So, testers' expertise levels can affect their efficiency when identifying failures due to their previous knowledge or experience in similar tasks.

Another factor is the failure being diagnosed. Its effect relates to the case studies (Subsections 3.5.4) available for experimentation and its relation with testers' expertise levels. Testers' expertise levels are related to their previous experience regarding identification of failures from the same nature (electric or electronic) as the experimental ones. Failure's factors (CNN and TEM) presented in Table 3-5 represent the most reported failures in reporting usability tests for the case study's asset as described in Subsection 3.6.3.

The last factor taken into consideration is the monitoring tool that supports the failure identification task. Three different alternatives were considered:

1. None: no monitoring tool was used for the experiment test. This serves as baseline for further analysis.
2. KRD (Knowledge Recommendations Disabled): a monitoring tool that only includes common rules for condition monitoring (e.g. data-driven) as those presented in Figure 3-7 – Real time monitoring.
3. KRE (Knowledge Recommendations Enabled): the proposed monitoring method (Subsection 3.4.3) that includes common condition monitoring rules (Figure 3-7 – Real time monitoring) and condition monitoring rules from expert diagnosis knowledge re-use (Figure 3-7 – Real time expert recommendations).

These three factors and their levels identify the number of experimental groups to which allocate testers. This results in a total of 12 (2x2x3) different groups to which allocate testers to. Testers were allocated to one group only, so the experiment could be designed as between-subjects for further statistical analysis. Tester allocation was done randomly according to the failure's nature and the monitoring tool utilised, although it was ensured to include the same number of testers with different expertise levels at each subgroup for the other two factors.

"A priori" calculations can be made to design the sample size for this between-subjects experiment. The results are based on an F-test for a three-way ANOVA. This test considers a number of groups of 12 (factors levels), a variance of 0.3 (partial eta squared), a type-I error of 0.1 (alpha) and a power of 0.9 (1 – beta). The resulting required sample size is 50 people.

A total of 48 testers took part in the experiments. Testers have ages ranged from 22 to 30 years old and they were all enrolled in different MSc degrees relating with engineering. Those that had any additional working experience on electric and/or electronic failures were classified as 'IT', and those with none experience as 'NOIT'. The asset utilised as case of study for these experiments is explained in Subsection 3.5.4. The experimental failures, which resulted from the reporting tests for NFF scenarios, are described in Subsection 3.6.3. The results of these experiments are analysed in Subsection 3.6.5.

3.5.2.6 Monitoring usability surveys

After experimenting with the monitoring tool, testers were asked to complete a questionnaire regarding its usability. The student considered it relevant criteria due to the potential effects of the tool's interface on monitoring efficiency (e.g. complexity). Usability surveys aim at evaluating the perceived validity of the monitoring tool (KRE) to deliver information about asset's condition and failures to support fault-finding tasks in comparison with other alternatives (KRD). According to Nielsen (1993), usability is a qualitative attribute to assess *"how well users can use a functionality"* of a tool. There are many qualities of software tools that can be related to usability: learnability, simplicity, memorability, etc. Unlike other usability surveys proposed in this thesis, the time spent by testers with the monitoring tool cannot be considered long enough to evaluate many qualities of it. Therefore, the student proposed to select only few criterions from those proposed by Nielsen (1993). The criterions selected were those that are most related to the aim proposed above for the usability surveys. They are the following:

- Ease-of-use: ability of the monitoring tool to be easily utilised by a user. Refers to the tool's ability to present assets' condition and failures clearly.
- Effectiveness: ability of the monitoring tool to support a user with the task at hand. Refers to the tool's ability to help users conduct fault-finding tasks.

Usability surveys consisted of a questionnaire that includes several statements regarding the abovementioned criterions. The survey's questionnaire can be found in Appendix C. Testers were asked to state their agreement with each statement on a Likert scale from 1 to 5. The reason to select such scale is based on the results presented by Weijters, Cabooter and Schillewaert (2010), who suggested that it maximises potential information transmission when surveying non-expert populations. Unlike experts from reporting usability surveys (Subsection 3.5.2.3), MSc students who took part as testers in monitoring experiments cannot be considered an expert sample. Testers that utilised reporting tools (KRE, KRD) were asked to complete the survey's questionnaire after the monitoring efficiency experiments. So, they could state their subjective opinion regarding the tool's interface usability. It is also relevant to note that from the 48 experimental testers, only 36 of them used either one of the monitoring tools (KRE and KRD). Those testers which made the experiments without a monitoring tool did not take the usability survey.

3.5.3 Experimental protocol

The experimental protocol describes the necessary steps to collect and analyse relevant data for validating this research's proposals according to the validity criterions above. This protocol comprises the application of the validation methods presented in the previous section. It consisted of the following steps:

1. Data collection:

a. Ontology expert interviews (9 subject-matter experts): to capture qualitative data from subject-matter experts' opinions on of the proposed ontology (diagont) schema.

b. Ontology structural analysis (10 literature ontologies): to capture quantitative data from similar ontologies regarding OntoQA metrics (Tartir and Arpinar, 2007).

c. Reporting usability tests (9 subject-matter experts): to capture real-life data from physical assets and their identified failures in NFF scenarios for case study identification.

d. Reporting usability surveys (9 subject-matter experts): to capture subject-matter experts' opinions on the tool's usability for diagnosis reporting.

e. Monitoring efficiency experiments (48 testers): to capture quantitative data on the effect on efficiency of the proposed monitoring method (KRE) compared to other alternatives (None and KRD).

f. Monitoring usability surveys (32 testers): to capture qualitative data on tester's opinions regarding usability of the proposed monitoring tool (KRE) compared to other alternatives (KRD).

2. Data analysis:

a. Ontology expert interviews: to evaluate the number of changes proposed by interviewees in recurrent improved version of the ontology schema to validate its fit for purpose.

b. Ontology structural analysis: to analyse OntoQA metrics of the proposed ontology schema in comparison with similar ontologies to validate its fit for purpose.

c. Reporting usability tests: to evaluate collected reports on physical assets and related failures identified for designing experimental cases of study.

d. Reporting usability surveys: to analyse the reporting tool's usability regarding its effect on expert captured knowledge's validity.

e. Monitoring efficiency experiments: to study the effect on failure identification task's time and errors of the monitoring tool's compared to other alternatives and considering the effect of relevant factors such as testers' expertise levels and nature of the failure diagnosed.

f. Monitoring usability surveys: to analyse the monitoring's tool ease-of-use and effectiveness regarding its effect on failure identification task's efficiency according to testers' opinions.

The evaluation of quantitative criteria includes different types of analysis, including statistics. Ontology's expert interviews and structural analysis are simpler due to the smaller sample size and the lack of effects to consider. Usability surveys, both from reporting and monitoring tools, utilise simple statistics to analyse responses distributions (medians and quartiles) for each criterion evaluated. Efficiency experiments' analyses include more complex statistics due to the number of factors to consider for analysing the response variables. In efficiency experiments, response variables (time and errors) are assumed to be positively correlated and to be affected by experimental factors. According to Cohen (Cohen, 1992), those variables positively highly correlated should not be evaluated through multivariate analysis as their similarity can be assumed. Hence, the student proposed to analyse time and errors using independent three-way ANOVA between-subjects tests to analyse their correlations with experimental factors. So, it can be expected that the results of these test should be similar.

This protocol was applied to two different case studies from two different maintenance organisations and its results should be discussed within that context. These case studies, which resulted from the reporting tests, are presented in the subsection below to provide the necessary context for the analysis conducted in following sections.

3.5.4 Case studies

The abovementioned system prototype and experimental protocol served to validate the proposed methods using two case studies. These cases were selected because they were the two most referred within the diagnosis reports collected as part of the reporting tests. They embody two complex engineering assets of different natures (electro-mechanical and electronic). Figure 3-11 presents views of these two cases.

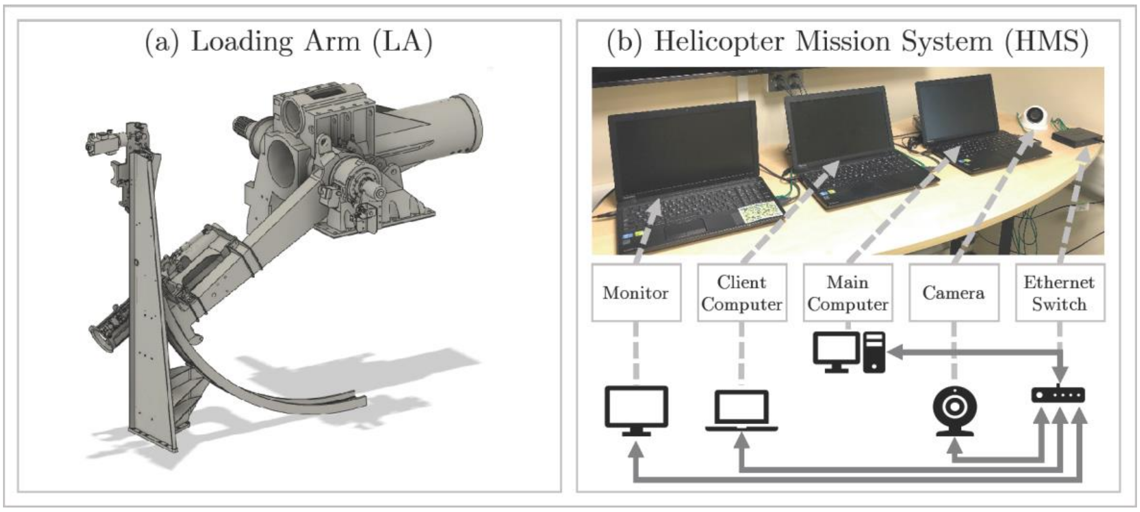


Figure 3-11. Depiction of complex assets utilised as case studies: (a) Loading Arm (LA), (b) Helicopter Mission System (HMS).

The first case study (Figure 3-11 – (a) Loading Arm) is an electromechanical system that belongs to an asset installed in a naval ship. It is a large piece of equipment design to move heavy loads at very high speeds. It comprises various mechanical components that rotate around each other, along with several actuators and sensors to control its behaviour. The latter are managed by an integrated control monitoring panel that is part of the system.

The second case (Figure 3-11 – (b) Helicopter Mission System) is a replica of an electronic system that aims to control the navigation mission of a helicopter. This replica was built with the same specifications as the original in order to enable laboratory experimentation. This system comprises three computers, one camera and an ethernet switch to connect them altogether. The first computer, called *'main mission'*, is used as controller for the rest of the elements and also controls the navigational parameters of the helicopter. The second computer, or *'client mission'*, is that used by helicopter pilots to set the navigation mission. The third computer, which acts merely as a *'monitor'*, is that from which pilots control the *'client mission'* computer. The *'camera'* is there to provide pilots with a visual of the terrain while handling the helicopter. Finally, the *'ethernet switch'* aims to connect the main mission computer to the client mission computer, the *'monitor'* and the *'camera'* for further control. The system comes with an integrated control monitoring system that evaluates electronic performance parameters from its different elements. Due to its criticality for piloting the helicopter, real-life maintainers are very careful when reporting diagnosis procedures on it. Besides, the control monitoring has some limitations regarding the electronic parameters it can control due to the system's configuration.

These systems' complexities make them suitable for the experimental methods proposed by this research. One example is the existence of sensors and actuators in these systems. They enable to provide reliable data sources for reporting numerical measures while conducting diagnosis. Besides, the systems' criticality within their assets can be assessed by the number of maintenance routines they need to pass through. Those require these systems to be inspected at least once every time the asset is set to operation. It was also noted by the tested experts that such criticality impacts on the detail required for reporting diagnosis procedures related to them. Another relevant aspect, emphasised by experts during the tests, is the complexity of their configurations. This makes them prone to incur in no-fault-found (NFF) scenarios more often than other systems in the assets. The reporting tests allowed subject-matter experts to report diagnosed failures in NFF scenarios. Those failures most reported were selected as cases of study for this research's experimental methods, whose results are analysed and discussed in the following section.

3.6 Results and discussion

This research's validation aim is to demonstrate the hypotheses from Subsection 3.5.2. The student conducted several experimental methods (Table 3-3) to collect data to do so. Their results are analysed in the following subsections to evaluate the validity of this research's hypotheses. The complete results datasets and analysis can be consulted at 10.17862/cranfield.rd.12279152.

3.6.1 Ontology expert interviews

During expert interviews (Subsection 3.5.2.1), experts were asked to propose modifications to the ontology schema. These changes included any variations to the ontology's vocabulary, as well as its classes, attributes and relationships and the axioms that relate them. The aim for recording these changes was to update the ontology schema according to subject-matter experts' opinions. So, the schema could fairly represent their rationale when conducting diagnosis activities.

These changes were recorded for each interview and applied iteratively between consecutive interviews. The only modifications applied were those that did not involve specific vocabulary from the interviewee's organisation. So, the rationale described by the ontology's schema could be kept generic. The changes proposed and applied by interviewee are shown in Figure 3-12. The total number of proposed changes is 85 with an average of 9 changes per interviewee, while the total changes applied are 47 (55% of total changes) with an average of 5 per interviewee.

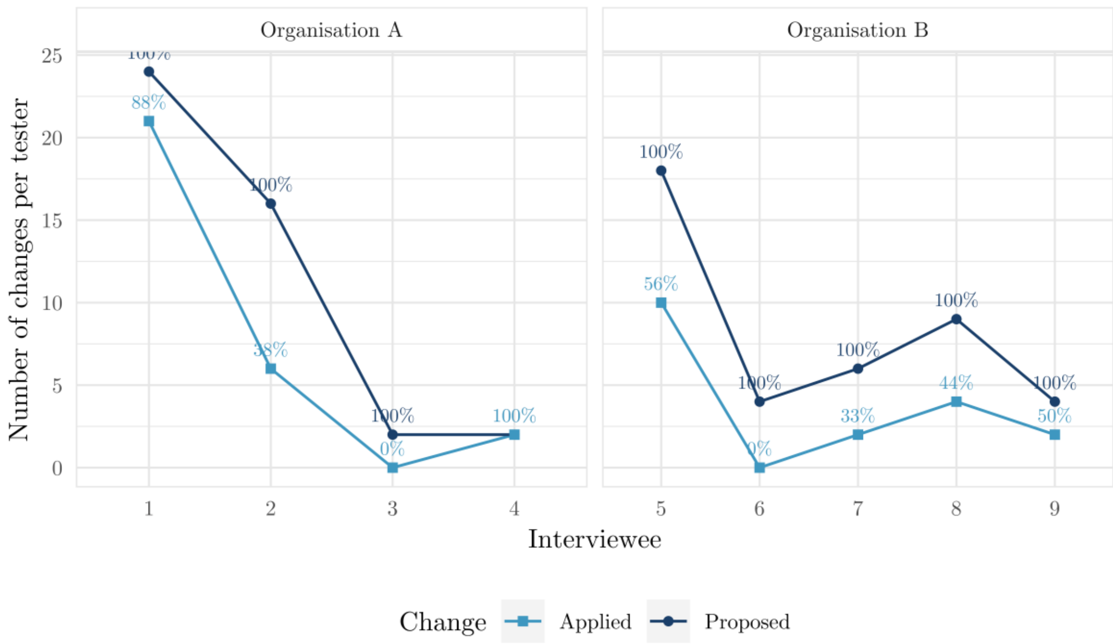


Figure 3-12. Results of proposed and applied design changes from expert interviews – The dark-blue line represents the number of changes proposed by interviewee while the light-blue line represents the number of changes applied.

Based on these results and Figure 3-12, there are a number of aspects relevant to note:

1. The total number of applied changes (47) is 55% of the total proposed changes (85). Number of changes can be considered low compared to the total number of changes that could have been proposed. A maximum threshold for this number can be taken at the total number of ontology's axioms (943). Based on that, total proposed changes represent a 9% while the total applied changes decrease to 5% of the total changes that could have been proposed.

2. Figure 3-12 shows a decreasing tendency on the number of changes proposed and applied by each consecutive interviewee per organisation, apart from the peak on changes in between organisations. This tendency can be interpreted as an alignment between the diagnosis rationale presented by the ontology's schema and the interviewees understanding of that knowledge domain.

Besides these interviews, experts also tested the reporting tool, which included the latest modified version of the proposed ontology after applying their changes. The results of these tests are described in Subsections 3.6.3 and 3.6.4.

3.6.2 Ontology structural analysis

The ontology structural analysis consisted of evaluating the proposed ontology in terms of OntoQA metrics (Tartir and Arpinar, 2007) through comparison with similar ontologies. This analysis aims to study the ontology's fit for purpose to describe experts' rationale on diagnosis activities. As described in Subsection 3.5.2.2, a total of 10 different ontologies were taken from literature for this comparative study. Table 3-6 presents the OntoQA metrics (Relationship Richness (RR), Attribute Richness (AR), and Inheritance Richness (IR)) results and rankings for the 11 ontologies studied.

Table 3-6. Results of ontology structural metrics evaluation.

Paper	Ontology	DOI	RR	Rank	AR	Rank	IR	Rank
1	DIAGONT	10.17862/cranfield.rd.12279152	0.636	5	0.577	6	0.769	7
2	IMAMO	10.3233/ao-2012-0112	0.500	9	0.360	8	0.760	8
3	MASTONT	10.3182/20120523-3-RO-2023.00124	0.917	2	0.182	10	0.910	10
4	MASONT	10.1016/j.jnca.2012.11.004	0.500	7	0.000	11	0.500	9
5	FMECAONT	10.1016/j.aei.2014.10.001	0.464	10	0.353	9	0.882	6
6	DTMONT	10.1016/j.compind.2013.03.001	1.000	1	1.000	3	0.000	11
7	EFMONT	10.1016/j.knosys.2014.02.002	0.361	11	0.958	5	0.958	4
8	FDONT	10.1109/TSMC.2013.2281963	0.675	3	1.000	4	1.000	2
9	AI2MS	10.1109/indin.2014.6945616	0.667	4	1.000	2	1.000	1
10	AHMK	10.1109/icqr2mse.2012.6246302	0.500	8	0.500	7	0.958	5
11	GOSP	10.1016/j.jlp.2012.10.001	0.521	6	1.639	1	0.972	3

The proposed ontology's (diagont) aim is to capture diagnosis experts' rationale avoiding the use of unstructured text. Compared to other ten similar ontologies, diagont's rankings are 5th in RR, 6th in AR and 7th in IR (Table 3-6). These metric results help to evaluate the achievement level of the diagont's aim:

- Diagont's RR is 0.636, which means that 63% of its relationships are non-hierarchical. Diagont is the 5th ontology in terms of knowledge's richness.

These results help to evaluate diagont's ability to capture knowledge (Tartir *et al.*, 2005). Hence, it can be said diagont captures a fair amount of knowledge regarding experts' diagnosis rationale.

- Diagont's IR is 0.769, which means that it is the 7th ontology in terms of knowledge generality (1st with 1.639). IR results help to analyse the compromise between generic and detailed knowledge (Tartir and Arpinar, 2007). So, it can be said that diagont provides detailed knowledge and attains some generality over the domain covered, compared to other ontologies.

- Diagont's AR is 0.577. This means it is the 6th ontology in terms of number of attributes per class. This result can help to identify diagont's ability to avoid unstructured text. While being the 6th ontology in terms of captured knowledge's detail, only 14% (3 out of 21 – Table 3-2) of it is described as unstructured text (string as datatype). Hence, it can be said that the level of detail proposed by diagont is high enough to not include unstructured text.

Figure 3-13 displays diagont's rankings in AR, IR and RR, helping to visualise the differences between RR, AR and IR metrics among ontologies. A relevant metric in Figure 3-13 is AR, whose difference between diagont and other schemas is higher compared to RR and IR. This difference resulted of the unstructured text attributes utilised by those ontologies, which is above 50%. This emphasises diagont's ability to achieve detailed knowledge while avoiding unstructured text.

The results on this section and the previous analysed the proposed ontology's fit for purpose. Nevertheless, it is still needed to evaluate what is the impact of its knowledge according to the tasks aiming to improve. Hence, the following sections presents validation methods results regarding the reporting and monitoring tools.

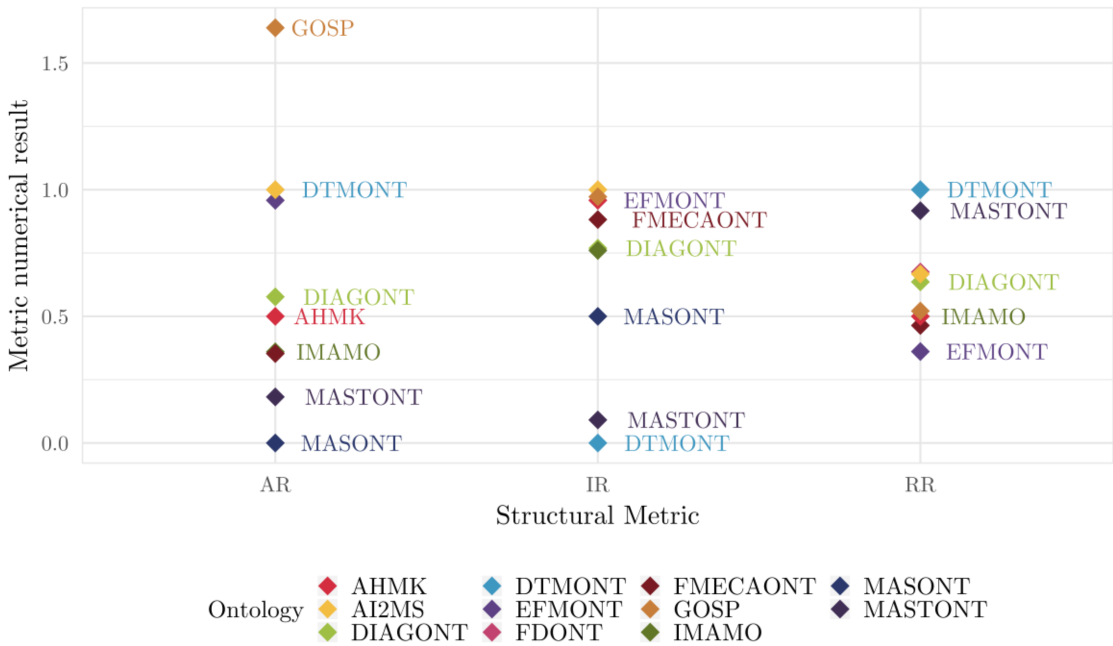


Figure 3-13. Ranked ontologies according to evaluation results of structural metrics: Attribute Richness (AR), Inheritance Richness (IR) and Relationship Richness (RR).

3.6.3 Reporting usability tests

The purposes of reporting usability tests were two. The first one was to collect sufficient data of identified failures in NFF scenarios that could be used as cases of study in monitoring experiments (Subsection 3.6.5). The second was for testers (subject-matter experts) to try the proposed reporting tool for providing their opinions regarding its usability for capturing knowledge on diagnosis activities. The results of the usability questionnaires are analysed in the following section. Instead, this section describes the failures and diagnosis tasks most reported during these tests.

The experts who took part on the ontology expert interviews (Subsection 3.6.1) also conducted the reporting usability test. The nine interviewees worked for two organisations (four and five each) from the maintenance industry. During these tests, experts were given a tablet device with access to the reporting tool through a web browser (Figure 3-4). For the purpose of testing the tool, testers were asked to complete a report regarding a failure they diagnosed in a NFF scenario.

As a result of these tests, data collected consisted of 6 failures reported (3 per organisation). For the first organisation, one of the three failures was reported twice. While for the second one, two failures out of the three failures were reported twice and both belong to the same asset. These three failures were the ones with the greatest number of reports. The assets to which the failures belong to were described in Subsection 3.5.4. Instead, the failures themselves are described below.

The most reported failure from the first organisation ("Loading Arm") is presented in Figure 3-14. This figure shows a simplified description of the failure using diagont's classes, attributes and relationships. The failure results of an incoherence in the control's logic that operates the "Loading Arm" caused by a deficiency with respect to the hydraulic medium inside one of its components. So, when the failure occurs it drives the control's module to a NFF condition.

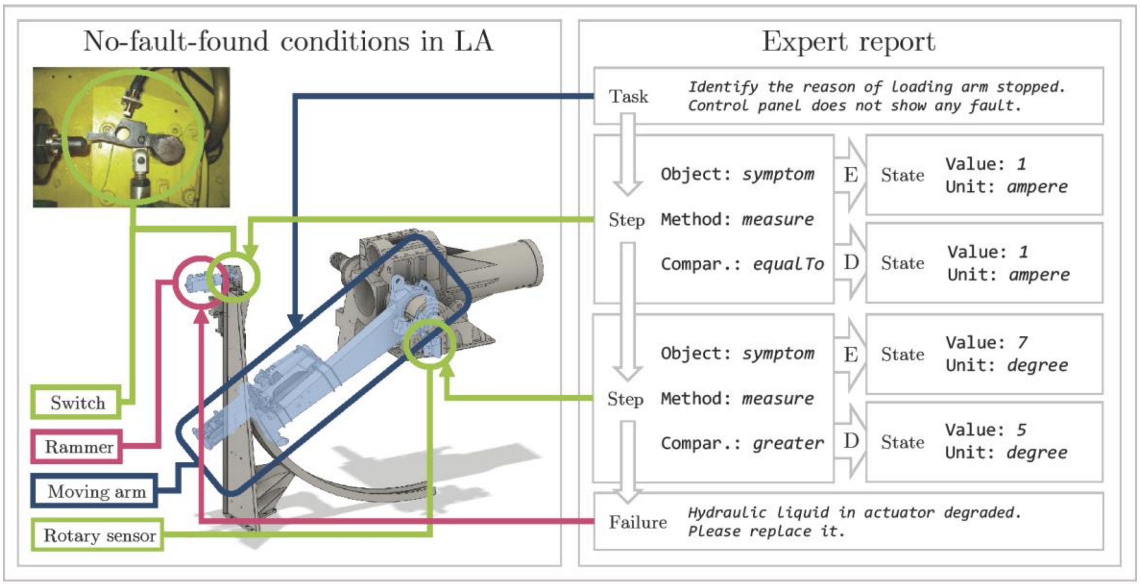


Figure 3-14. Description of NFF scenario in Loading Arm: Hydraulic degradation – Includes a graphical representation of the asset's components involved (left-side) as well as some diagont's classes' and object and datatype properties' assertions utilised by experts to report the failure (right-side).

The failure occurs when the "moving arm" (Figure 3-14) is at its horizontal position and it is asked to return to its vertical position. The system's control module controls that movement using two sensors. The first one is a "switch" that controls when the "moving arm" has left its horizontal position. The second one is a “rotary sensor” that controls the inclination of the “moving arm”. For the “switch” to indicate the “moving arm” has left its horizontal position, the “moving arm” has to start moving downwards and it also requires the “rammer” to move forward. The “rammer” contains a hydraulic liquid that allows it to move horizontally. However, if that liquid is degraded, the rammer won’t move. What happens then is that the “switch” does not show that the "moving arm” has left its horizontal position, but the “rotary sensor” also indicates the “moving arm’s” inclination. In that case, the control module incurs in a logic error, and stops the “moving arm” moving downwards. However, the control module does not show any indication of the failure but a simple logic error message. The numerical values from each sensor at which this situation occurs are shown in Figure 3-14. These values are shown as reported by the expert testers. Hence, there are *‘evaluated’* (“E”) and ‘diagnosed’ (“D”) ‘States’ that describe them.

Figure 3-15 graphically describes one of the two failures most reported in reporting tests for the “Helicopter Mission System”. This failure is referred as CNN in monitoring experiments (Subsection 3.6.5). The failure is caused by a fault in the cable that connects the “main computer” with the “ethernet switch”. This failure provokes a NFF condition because that cable cannot be monitored by the system’s control module. The control module is managed by the “main computer” and one of its limitations is that it cannot evaluate its own connectivity to the “ethernet switch”. So when this disconnection occurs, the control module shows the rest of the system’s components (“client computer”, “monitor” and “camera”) as not connected even though the connectors are in good condition. As shown in Figure 3-15, connectivity is measured by connectivity time, which is identified as zero by the control module when there is no connection. The tester’s report in Figure 3-15 also describes all the ‘Steps’, including ‘evaluated’ and ‘diagnosed’ ‘States’, taken by the tester to identify the ‘Failure’.

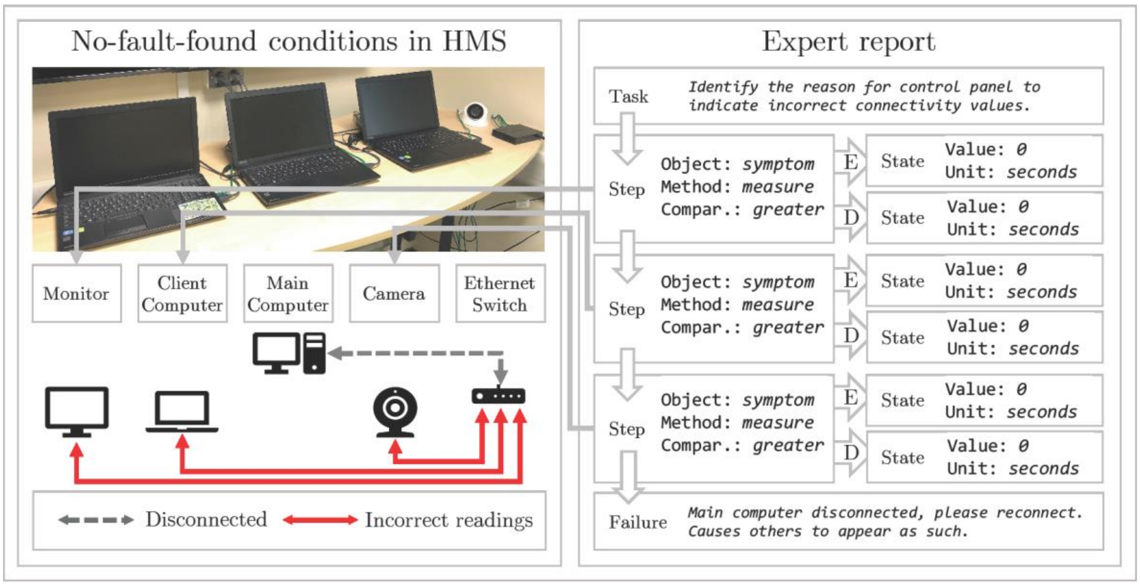


Figure 3-15. Description of NFF scenario in Helicopter Mission System: Main Computer disconnected (CNN) – Includes a graphical representation of the system's components involved (left-side) as well as some diagont's classes' and object and datatype properties' assertions utilised by experts to report the failure (right-side).

The other failure most reported for the "Helicopter Mission System" tests is described by Figure 3-16. It is described as TEM in monitoring experiments (Subsection 3.6.5) It is another example of NFF because the thresholds established by the system's control module are higher than the values that cause the failure. The failure consists of a hardware overload caused by too many software being run in the "Helicopter Mission System" simultaneously. When this occurs, both "main computer" and "client computer" reach CPU temperatures higher than 60º Celsius (~333 Kelvin). However, the system's control module cannot detect this issue because the CPU temperatures monitoring thresholds are set for each CPU independently at a temperature of 90º Celsius. Similarly to previous failures, Figure 3-16 shows a simplified version of the reports given by experts during tests about the failure.

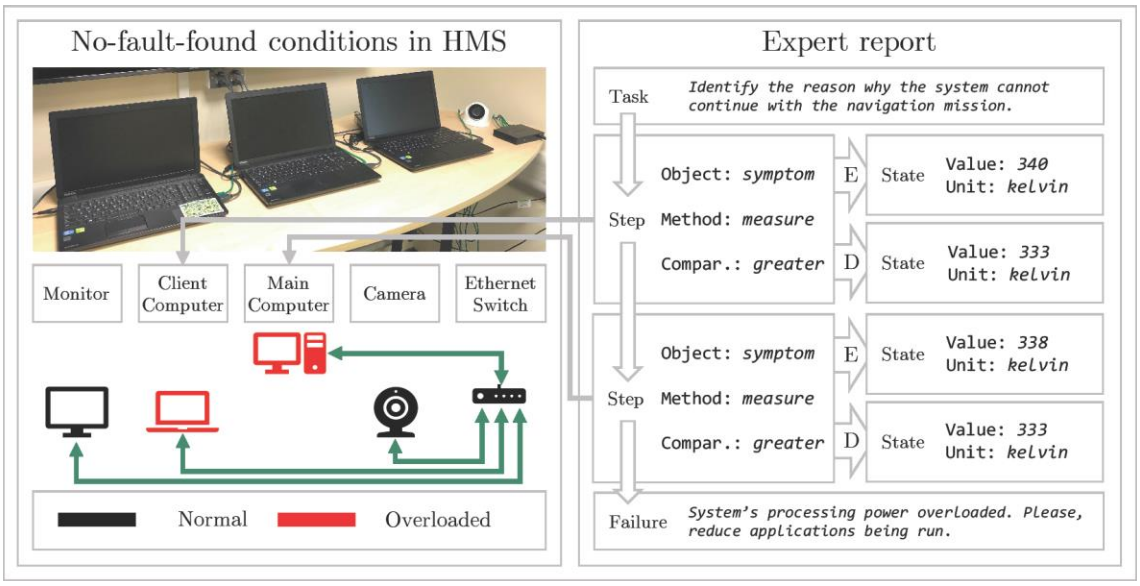


Figure 3-16. Description of NFF scenario in Helicopter Mission System: System processing temperature overloaded (TEM) – Includes a graphical representation of the system's components involved (left-side) as well as some diagont's classes' and object and datatype properties' assertions utilised by experts to report the failure (right-side).

These failures provided by reporting testers helped to complete the cases of study for the monitoring experiments. The failures chosen for evaluating the impact of the monitoring tool in failure identification's efficiency are the two referring to the "Helicopter Mission System". The reasons to use this system as case of study for the monitoring experiments are based on the need to conduct laboratory experiments. These were the following:

- Experiments required a system that could be setup for a period of time to the same failures' conditions consistently. This would have been difficult to achieve on a real-life working system ("Loading Arm") due to the disruptions that would have been caused to the equipment's operational availability.
- Experiments required to keep similar expertise levels in testers. Because real-life maintainers can have many variable expertise levels, the student decided to use MSc students instead, whose expertise levels regarding failure diagnosis can be categorised easier in fewer levels.

- In order for experiments' analysis to provide consistent conclusions, other factors that could have an effect on their results should remain constant (e.g. ergonomics, light conditions, etc.). Hence, the student decided to conduct laboratory experiments to control such factors.

An additional result of the reporting tests was that experts got a clear understanding of the proposed reporting tool's usability. The results of the consequent usability surveys are analysed in the following subsection.

3.6.4 Reporting usability surveys

Reporting usability surveys aimed at evaluating the reporting tool's usability to capture expert knowledge on diagnosis activities. Their results are presented in Figure 3-17. Besides, Subsection 3.5.2.4 explained the ontological aspects (context, structure and vocabulary) to analyse according to some validation criterions (accuracy, completeness, conciseness and consistency). These aspects and criterions can help to analyse experts' opinions regarding the tool's ability to capture their diagnosis rationale as well as the captured knowledge's quality.

Figure 3-17 shows a box and whiskers plot that describes the distribution of experts' responses per aspect and criterion evaluated. On average, all ontological aspects scored between 5 and 6 for all criterions considered. Besides, all aspects were considered highly complete and consistent according to the responses' variabilities. Thus, showing that experts' opinions suggest the reporting tool is very complete and consistent for capturing their knowledge. A relevant thing to note is the relatively lower responses obtained for the accuracy and completeness of the tool's structure. During the tests, experts noted that the tool would require additional user interface functions to improve its usability (e.g. simplicity of individuals instantiation). Another relevant thing to note is the higher variability on the scores for the vocabulary criterions. The specific terminology used by the tool, and so for the ontology, seems to reflect more disagreement than other ontological aspects. This could reflect the disagreement on terminology showed by experts from different organisations during the expert interviews (Subsection 3.6.1).

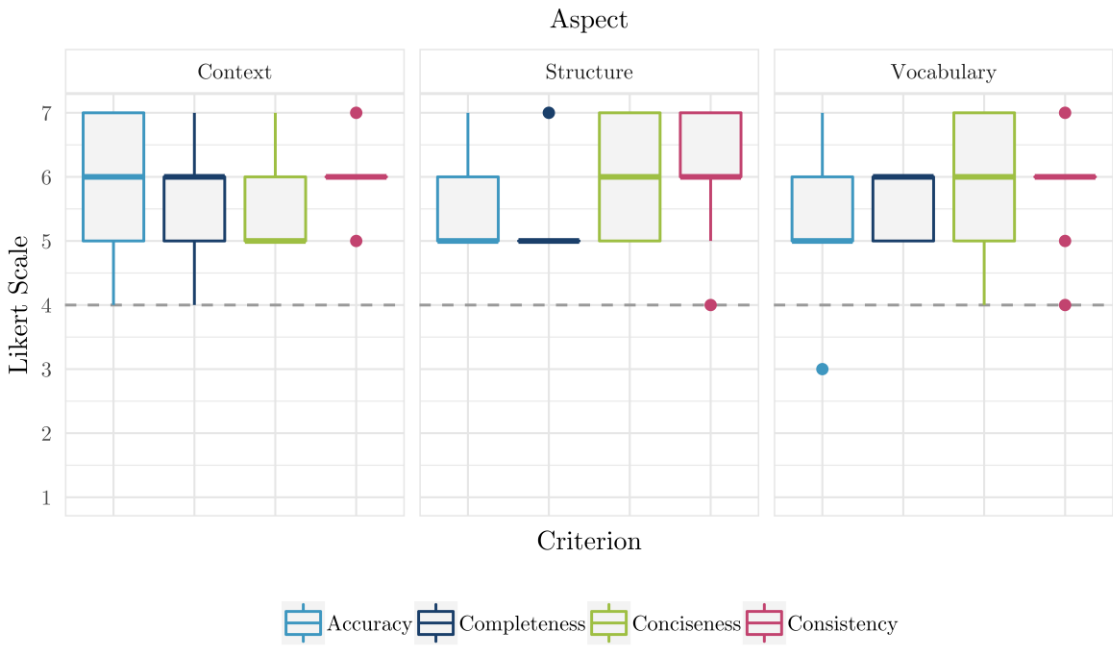


Figure 3-17. Distribution (maximum, upper quartile, median, lower quartile, minimum and outliers) of reporting tool's usability survey responses for each criterion and aspect.

The last two validation methods (Table 3-3) evaluated the monitoring tool's impact on failure identification tasks' efficiency and its usability for that purpose. These methods and their results are analysed in Subsections 3.6.5 and 3.6.6 according to the hypotheses they aim to demonstrate.

3.6.5 Monitoring efficiency experiments

Experiments aimed at evaluating the impact of the proposed monitoring method on failure identification tasks' efficiency in NFF scenarios compared to other non-knowledge-based methods. For a given failure, tasks' completion time and errors can be a direct representation of efficiency. Hence, these two are considered the response variables for the efficiency experiments. Besides failure identification tasks, there are other factors that can have an effect on these variables. The student considered the following: nature of failure, tester's expertise and monitoring solution. These factors can have different levels, which were listed in Table 3-5. The failures utilised as cases of study are explained Subsection 3.6.3 and the assets they belong to in Subsection 3.5.4.

These experiments were designed to demonstrate the validity of these hypotheses:

- The monitoring solution has a significant effect on failure identification errors.
- The monitoring solution has a significant effect on failure identification time.

A simple exploratory analysis on the correlation between these response variables helped to determine the statistical method chosen for analysis. Based on Cohen's guideline's (Cohen, 1992), Pearson's coefficient ($r = 0.732$, $p = 3.511e\text{-}09$) suggests that these two have a highly positive correlation. Hence, the student decided to analyse each one independently. In order to analyse the effects of each factor over each response variable, the chosen test was three-way between-subjects ANOVA. Thus, using the comparison among both studies to refute their results.

3.6.5.1 Stopwatch errors study

Errors measure the number of incorrect hypotheses suggested by testers during experiments. They can reflect quality and efficiency of failure identification tasks'. Errors' results present a decreasing tendency with more knowledgeable monitoring solutions as Table 3-7 shows. Compared to no support (none), the KRD solution has an error reduction rate of 50% and the proposed monitoring method (KRE) has a rate of 77%. Although these results require further analysis to evaluate the effect of other experimental factors (failure's nature and expertise level).

Table 3-7. Mean and std. deviations of completion errors per solution.

Solution	Testers	Mean	Std. deviation
None	16	1.620	0.957
KRD	16	0.812	0.834
KRE	16	0.375	0.619

Figure 3-18 presents the errors' average per experimental group along with their standard deviations. This figure shows a difference on average errors per failure, expertise and solution factors' groups. Although it shows a decreasing rate per solution, the rates differ for each group. Errors for different expertise levels vary greatly. It can be seen that IT testers make less errors compared to NOIT testers and with lower variability. Besides, errors' with KRD and KRE solutions seem similar for IT testers. According to failures, errors' averages are also different. For IT testers, Figure 3-18 shows that they made more errors in the experiments with an electronic failure (TEM) instead of an electric one (CNN). That is opposite for NOIT testers, who committed more errors in electric failure's experiments (CNN). An exception is the KRD solution, whose NOIT testers' errors where relatively lower in electric experiments (CNN). The differences on errors' according to failure, expertise and solution explained above require of further analysis to demonstrate their significance.

Table 3-8 presents the ANOVA's results on the error's variable according to experimental factors (nature, expertise and solution). For a confidence interval of 95% (p-value $\leq$ 0.05), it can be said that solution, expertise and their interactions have a significant effect on errors.

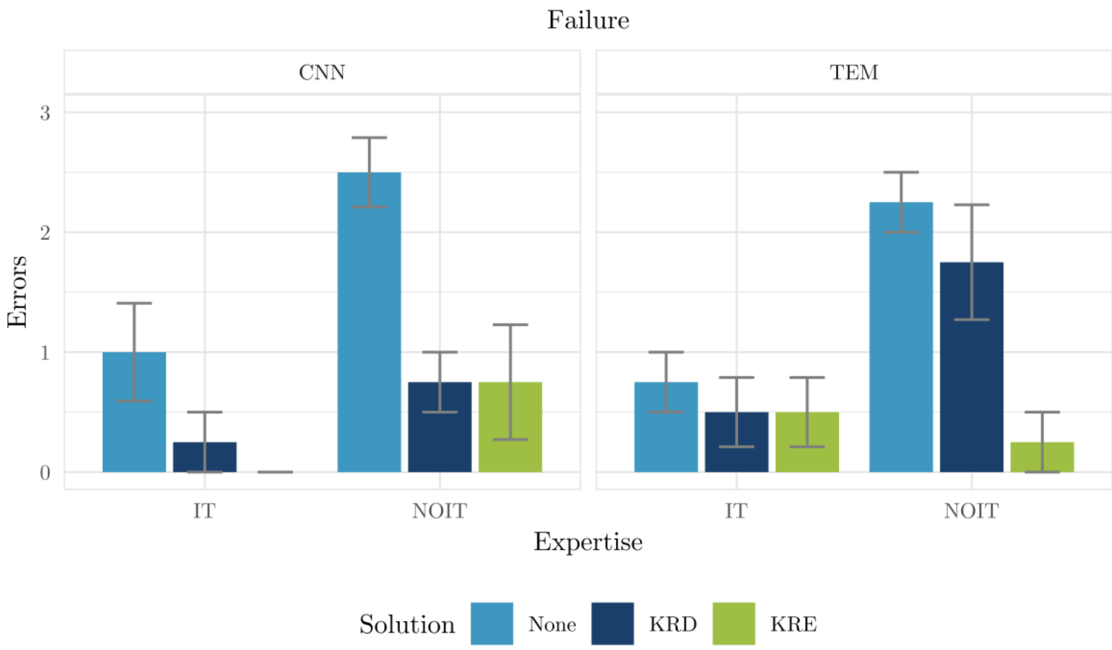


Figure 3-18. Average experiment completion errors (number of incorrect hypotheses) for None, KRD and KRE solutions per failure and expertise factor level, with the standard deviation represented by error bars.

Table 3-8. Summary of results of three-way ANOVA between-subjects test over experimental errors.

Effect	Df	Sum Sq	Mean Sq	F value	Pr (>F)	Sig (95% ci)
Solution	2	12.875	6.438	16.263	9.29e-06	Yes
Failure	1	00.188	0.188	00.474	4.96e-01	No
Expertise	1	09.188	9.188	23.211	2.62e-05	Yes
Solution:Failure	2	01.625	0.812	02.053	1.43e-01	No
Solution:Expertise	2	03.125	1.563	03.947	2.82e-02	Yes
Failure:Expertise	1	00.021	0.021	00.053	8.19e-01	No
Solution:Failure:Expertise	2	01.542	0.771	01.947	1.57e-01	No
Residuals	36	14.250	0.396	-----	-----	-----

In order to confirm these significances, a Tukey HSD test was conducted to post hoc evaluate the differences between factors groups. Table 3-9 shows the p-values for group factors interactions. These differences between interactions can be considered significant when p-values are equal or less than 0.05 with a confidence interval of 95%. Based on these results, the considerations made on Figure 3-18 according to relevant factors (expertise and solution) can be considered valid.

Table 3-9. Significance (p-value) results from post hoc comparisons Turkey HSD test on completion errors.

	A-C-E	A-C-F	A-C-G	A-D-E	A-D-F	A-D-G	B-C-E	B-C-F	B-C-G	B-D-E	B-D-F	B-D-G
A-C-E	-------	0.8629	0.5286	1.0000	0.9914	0.9914	0.0663	1.0000	1.0000	0.2184	0.8629	0.8629
A-C-F	0.8629	-------	1.0000	0.9914	1.0000	1.0000	0.0007	0.9914	0.9914	0.0035	0.0663	1.0000
A-C-G	0.5286	1.0000	-------	0.8629	0.9914	0.9914	0.0001	0.8629	0.8629	0.0007	0.0163	1.0000
A-D-E	1.0000	0.9914	0.8629	-------	1.0000	1.0000	0.0163	1.0000	1.0000	0.0663	0.5286	0.9914
A-D-F	0.9914	1.0000	0.9914	1.0000	-------	1.0000	0.0035	1.0000	1.0000	0.0163	0.2184	1.0000
A-D-G	0.9914	1.0000	0.9914	1.0000	1.0000	-------	0.0035	1.0000	1.0000	0.0163	0.2184	1.0000
B-C-E	0.0663	0.0007	0.0001	0.0163	0.0035	0.0035	-------	0.0163	0.0163	1.0000	0.8629	0.0007
B-C-F	1.0000	0.9914	0.8629	1.0000	1.0000	1.0000	0.0163	-------	1.0000	0.0663	0.5286	0.9914
B-C-G	1.0000	0.9914	0.8629	1.0000	1.0000	1.0000	0.0163	1.0000	-------	0.0663	0.5286	0.9914
B-D-E	0.2184	0.0035	0.0007	0.0663	0.0163	0.0163	1.0000	0.0663	0.0663	-------	0.9914	0.0035
B-D-F	0.8629	0.0663	0.0163	0.5286	0.2184	0.2184	0.8629	0.5286	0.5286	0.9914	-------	0.0663
B-D-G	0.8629	1.0000	1.0000	0.9914	1.0000	1.0000	0.0007	0.9914	0.9914	0.0035	0.0663	-------

Legend				
	Expertise	A = IT	B = NOIT	
	Failure	C = CNN	D = TEM	
	Solution	E = None	F = KRD	G = KRE

Table 3-10 presents errors' means and standard deviations for each group of relevant factors. Based on previous analyses, the following considerations can be considered significant:

- For NOIT testers, errors' reduction rates are 47% for KRD and 79% for KRE compared to none. The monitoring method proposed (KRE) reduces NOIT testers' errors (0.5 per test) significantly more than common data-driven approaches (KRD – 1.25 per test) compared to no additional support (None – 2.375 errors per test).
- For IT testers, errors' reduction rates are 57% for KRD and 71% for KRE compared to none. Although these percentages are very different, absolute values do not differ much (0.375 and 0.250). Both solutions reduced errors from almost 1 (None – 0.875) to almost 0 per test.
- Overall, the proposed monitoring method has a positive effect regarding errors reduction on failure identification tasks. Such effect is higher in the case of NOIT testers compared to IT.

These results provide an in-depth understanding of the impact of diagnosis knowledge re-use in fault-finding tasks through monitoring applications. For these results to be valid, they should be similar to the completion time analysis' results due to the correlation found between both response variables. The completion time results are analysed in the following subsection.

Table 3-10. Mean and std. deviations of completion errors per factors' group: expertise, nature and solution.

Expertise	Solution	Testers	Mean	Std. deviation
IT	None	8	0.875	0.641
	KRD	8	0.375	0.518
	KRE	8	0.250	0.463
NOIT	None	8	2.375	0.518
	KRD	8	1.250	0.886
	KRE	8	0.500	0.756

3.6.5.2 Stopwatch time study

Time measures the number of seconds taken by a tester from the experiment's start until the tester identifies the correct failure. Along with errors, time can help to reflect the efficiency of failure identification tasks. Table 3-11 presents an overview of time results according to solutions utilised during experiments. Compared to no support, KRD and KRE have time reduction rates of 61% and 70%, respectively. Thus, showing a similar pattern to errors' results regarding the effect of knowledge re-use. Nevertheless, further analysis is required to validate these considerations.

Table 3-11. Mean and std. deviations of completion errors per solution.

Solution	Testers	Mean	Std. deviation
None	16	66.000	25.631
KRD	16	25.625	13.579
KRE	16	19.938	12.551

Figure 3-19 presents the time's average per experimental group along with their standard deviations. It can help to identify the differences on average time per failure, expertise and solution. The figure shows a decreasing tendency on time per solution, although this decrease differs in between groups. For most groups, it can be seen that IT testers identified failures faster than NOIT testers. Only for the proposed monitoring method (KRE) both IT and NOIT testers achieve similar completion times. Besides, it also seems to be a difference on average time per failure. IT testers were faster on the electric failure (CNN), while NOIT testers were faster in the electronic one (TEM). The patterns described above seem similar to those presented for errors' results in Figure 3-18. Thus, emphasising the correlation between both response variables.

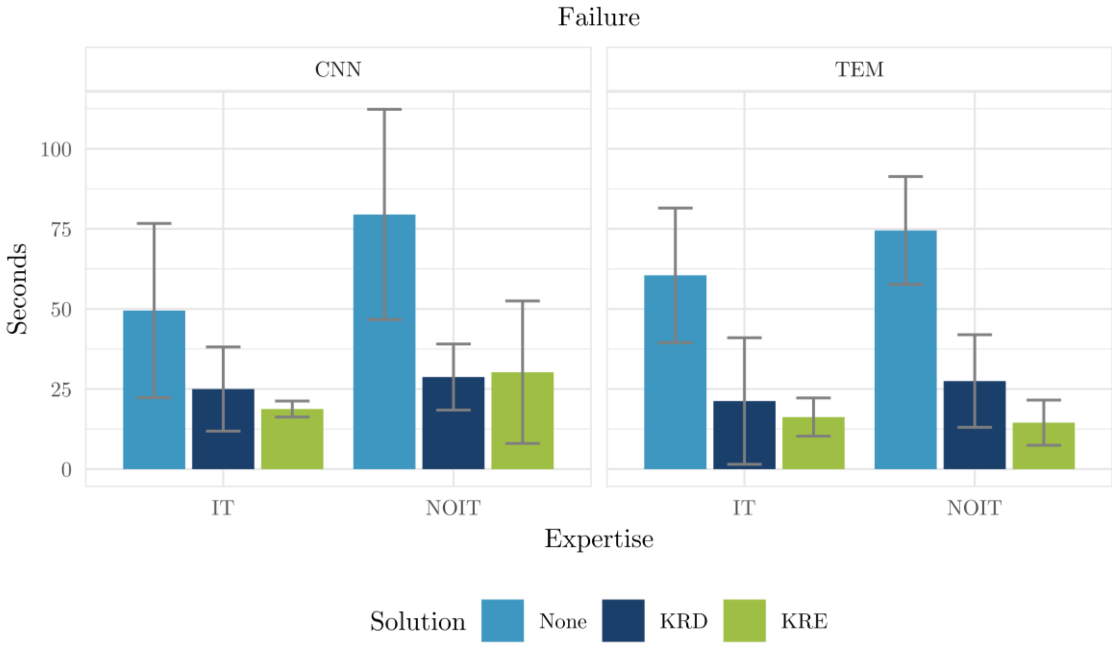


Figure 3-19. Average experiment time in seconds for None, KRD and KRE solutions per failure and expertise, with the standard deviation represented by error bars.

Table 3-12 summarises the results of the three-way ANOVA between-subjects test conducted to confirm the effects of the experimental factors in the time variable. For a confidence interval of 95% (p-value ≤ 0.05), it can be said that solution and expertise have significant effects on completion time, similarly to errors' results. Besides, Tukey HSD test post hoc comparison's results (Table 3-13) also suggest the same conclusion. Hence, the considerations on Figure 3-19 regarding effects of relevant factors on time results can be considered valid.

Table 3-12. Summary of results of three-way ANOVA between-subject test on time.

Effect	Df	Sum Sq	Mean Sq	F value	Pr (>F)	Sig (95% ci)
Solution	2	17109	8554	26.846	8.58e-08	Yes
Failure	1	00016	0016	00.050	8.24e-01	No
Expertise	1	00976	0976	03.064	8.88e-02	Yes
Solution:Failure	2	00583	0292	00.915	4.08e-01	No
Solution:Expertise	2	00398	0199	00.624	5.42e-01	No
Failure:Expertise	1	00107	0107	00.335	5.66e-01	No
Solution:Failure:Expertise	2	00127	0063	00.199	8.20e-01	No
Residuals	35	11153	0319	-----	-----	-----

Table 3-13. Significance (p-value) results from post hoc comparisons Turkey HSD test on completion time.

	A-C-E	A-C-F	A-C-G	A-D-E	A-D-F	A-D-G	B-C-E	B-C-F	B-C-G	B-D-E	B-D-F	B-D-G
A-C-E	-------	0.7269	0.4101	0.9990	0.5353	0.2999	0.9060	0.8805	0.9232	0.7025	0.8361	0.2350
A-C-F	0.7269	-------	1.0000	0.2184	1.0000	0.9999	0.0668	1.0000	1.0000	0.0172	1.0000	0.9994
A-C-G	0.4101	1.0000	-------	0.0778	1.0000	1.0000	0.0217	0.9996	0.9985	0.0045	0.9999	1.0000
A-D-E	0.9990	0.2184	0.0778	-------	0.1204	0.0490	0.9997	0.3638	0.4342	0.9922	0.3101	0.0349
A-D-F	0.5353	1.0000	1.0000	0.1204	-------	1.0000	0.0345	1.0000	0.9999	0.0078	1.0000	1.0000
A-D-G	0.2999	0.9999	1.0000	0.0490	1.0000	-------	0.0134	0.9970	0.9922	0.0026	0.9988	1.0000
B-C-E	0.9060	0.0668	0.0217	0.9997	0.0345	0.0134	-------	0.1232	0.1548	1.0000	0.1011	0.0095
B-C-F	0.8805	1.0000	0.9996	0.3638	1.0000	0.9970	0.1232	-------	1.0000	0.0366	1.0000	0.9910
B-C-G	0.9232	1.0000	0.9985	0.4342	0.9999	0.9922	0.1548	1.0000	-------	0.0490	1.0000	0.9805
B-D-E	0.7025	0.0172	0.0045	0.9922	0.0078	0.0026	1.0000	0.0366	0.0490	-------	0.0286	0.0017
B-D-F	0.8361	1.0000	0.9999	0.3101	1.0000	0.9988	0.1011	1.0000	1.0000	0.0286	-------	0.9958
B-D-G	0.2350	0.9994	1.0000	0.0349	1.0000	1.0000	0.0095	0.9910	0.9805	0.0017	0.9958	-------

Legend	Expertise	A = IT	B = NOIT	
	Failure	C = CNN	D = TEM	
	Solution	E = None	F = KRD	G = KRE

Table 3-14 presents time's means and standard deviations for each group of relevant factors. Based on the analysis above, the next considerations can be considered statistically significant:

- For NOIT testers, completion time with KRD is 63% and KRE is 71% faster than no additional support (none).
- For IT testers, completion time with KRD is 58% and KRE is 68% faster than no additional support (none).
- Moreover, the proposed monitoring method (KRE) achieve similar completion times for both, IT and NOIT testers.
- The patterns recognised for time results are similar to those recognised for errors' results. Thus, showing a positive correlation between both variables.
- Overall, the proposed monitoring tool has a positive effect on completion time for failure identification tasks. This impact is higher for NOIT testers, who achieve similar completion times to IT testers using the proposed solution.

These results provide deeper insights on diagnosis knowledge re-use's impact for condition monitoring applications. Besides analysing experimental results, another important aspect to validate of the proposed solutions is their usability. Hence, the next subsection analyses the usability surveys that testers completed after the experiments.

Table 3-14. Mean and std. deviations of completion time per relevant factors: expertise and solution.

Expertise	Solution	Testers	Mean	Std. deviation
IT	None	8	55.000	23.250
	KRD	8	23.125	15.661
	KRE	8	17.500	04.440
NOIT	None	8	77.000	24.302
	KRD	8	28.125	11.643
	KRE	8	22.375	17.443

3.6.6 Monitoring usability surveys

Usability surveys aimed at evaluating testers' opinions regarding monitoring tool's (KRE) usability for supporting failure identification tasks in NFF scenarios compared to alternative tools (KRD). Based on Nielsen's (1993) proposal, the student considered two criterions for evaluating monitoring usability: ease-of-use and effectiveness. Surveys analysis aims to identify differences on monitoring tools' usabilities according to testers.

Figure 3-20 illustrates a box and whiskers plot to summarise responses of 36 testers on each usability criterion regarding the monitoring tool utilised according to other experimental factors (expertise and failure). On average, both monitoring tools (KRE and KRD) scored above 3 on ease-of-use and effectiveness for each experimental failure and tester expertise level. But variabilities of responses differ for these two factors. Regarding tester expertise, Figure 3-20 shows that NOIT testers responded with lower variability and their responses are very similar for both, ease-of-use and effectiveness. Instead, IT testers responses had relatively greater variability. Moreover, those IT testers that diagnosed TEM failures in their experiments gave a relatively lower score to the proposed monitoring tool (KRE) that to the alternative (KRD). Although the differences shown are not big enough to indicate a significant difference in perceived usability. A potential reason for this lower score can be that IT testers are more used to common solutions (KRD). Nevertheless, both monitoring tools (KRE and KRD) shared the same interface. So, the difference in responses from IT testers may be due to either lack of trust in failure recommendations or lack of visualisation of such recommendations. In fact, some IT testers suggested that the monitoring interface could be improved when displaying complex failure modes through sensor data visualisation using schematics rather than textual descriptions.

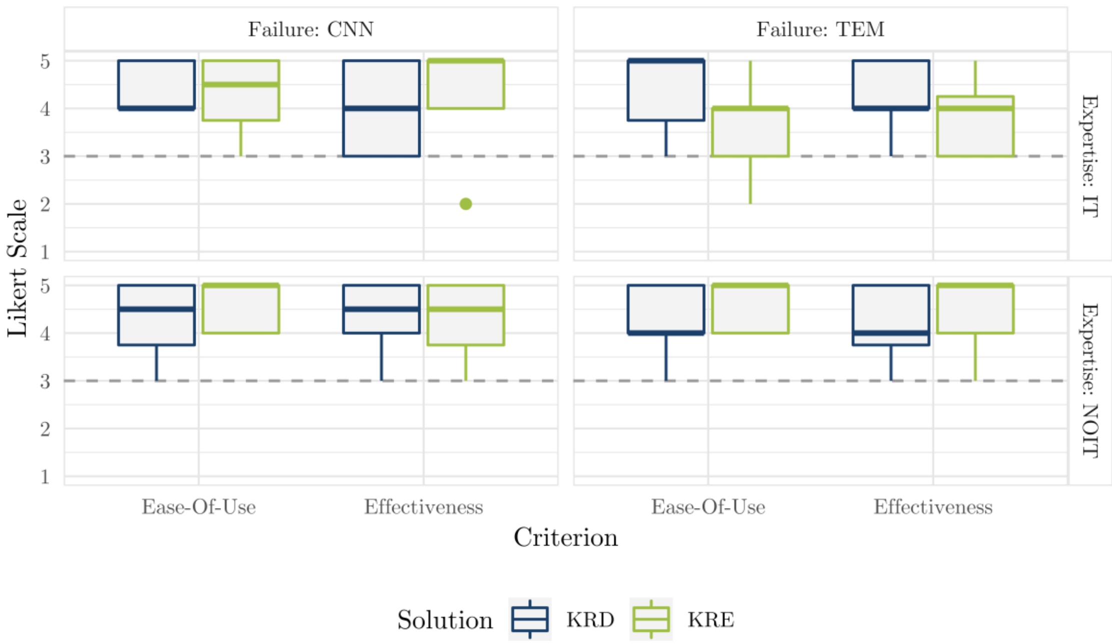


Figure 3-20. Distribution (maximum, upper quartile, median, lower quartile, minimum and outliers) of usability survey responses on each monitoring solution (KRD and KRE) for each experimental failure, expertise and criterion factor level.

Overall, Figure 3-20 indicates that IT and NOIT testers perceived a similar usability for both KRE and KRD monitoring tools, with lower variability and higher scores from NOIT responses compared to IT ones. Hence, it can be said that the proposed monitoring tool's (KRE) usability did not have a negative impact on fault-finding efficiency compared to alternative tools (KRD) according to testers.

3.6.7 Discussion

The experimental results analysed in previous subsections aim to demonstrate this research's hypotheses, which were presented in Subsection 3.5.2.

The first validation hypothesis stated that "the proposed ontology's schema fairly represents experts' rationale on diagnosis tasks". The methods to validate that hypothesis included expert interviews and ontology's structural analysis. Expert interviews' studies suggested that experts agreed with the proposed ontology's diagnosis rationale, based on the low number of changes (5%) and their decreasing rate per consecutive interviews. Besides, ontology's structural analysis indicated a rich, non-specific ontology that described diagnosis rationale without extensive use of unstructured text, based on OntoQA metrics comparison against other 10 ontologies. Hence, it can be said that the proposed ontology describes experts' rationale on diagnosis activities using qualitative and quantitative measures. However, these results should be considered valid within their context. Further studies could re-confirm these results to a wider context, extending the analysis to different organisations and other ontologies. Moreover, further research could also consider additional improvements, such as different datatypes to describe diagnosis measures (e.g. audio, image, etc.) or the relation between diagnosis and repair activities.

The second validation hypothesis affirmed that "the proposed reporting method is useful for capturing experts' knowledge on diagnosis tasks". The methods to validate it comprised expert tests and usability surveys. Subject-matter experts tested the proposed reporting tool on a tablet device to evaluate its usability and to collect reports on assets' failures occurred in NFF scenarios. These helped to design the case studies for further experimentation. Besides, reporting usability surveys' analysis suggested that experts found the tool accurate, complete, concise and coherent regarding certain ontological aspects for knowledge capture's usability. Although these results indicated the method's usability for the purpose of capturing knowledge, further experiments could increase their validity. They could focus on evaluating the reporting method in real-life conditions using quantitative criteria such as reporting efficiency and quality. Moreover, these experiments could also validate additional advances for the reporting method. These may include researching other information visualisation tools (e.g. Augmented Reality) and methods (e.g. recommendation algorithms) to improve reporting's efficiency and quality.

The third validation hypothesis stated that "the proposed monitoring method has a positive impact on fault-finding efficiency". If demonstrated valid, then it would imply that the proposed methods can actually serve for re-using diagnosis knowledge. Its validation methods included efficiency experiments and usability surveys. Experimental results analysis found a statistically significant correlation between fault-finding tasks' time and errors and statistically significant effects of expertise and solutions factors. Compared to usual data-driven methods (KRD), the proposed monitoring method (KRE) was found to improve error reduction rates by 32% for NOIT testers and by 14% for IT testers. The time reduction rate was found to improve by 8% for NOIT testers and by 10% for IT testers. Besides, usability surveys' analysis suggested that the proposed monitoring method's (KRE) ease-of-use and effectiveness had no negative impact on diagnosis tasks in comparison with alternative tools (KRD) according to testers' opinions.

The results discussed in the paragraph above indicate that the proposed monitoring method improves efficiency of fault-finding tasks. Although positive, the method's impact is variable with the user's expertise. Hence, future research could further improve it developing new methods to adapt knowledge re-use to user's expertise (e.g. recommender algorithms). Besides this factor, there are others that new algorithms could consider, which are also related to the proposed method limitations. One is the correctness of the knowledge being inferred (expert reports) that the current proposal assumes by default. Another is the similarity between failures of different assets or under similar environmental conditions. Besides recommendation methods, means for information delivery (e.g. Augmented Reality) can be another relevant area for future research to focus due to their impact on user's ergonomics and so, diagnosis efficiency.

The results' validity discussed above is relative to their experimental context. Reporting tests collected failure data of different natures (hydraulic, electric and electronic) from two different assets. Their results did not show any evidence of such nature having an impact on the method's usability. Monitoring experiments considered different nature of failures (electric (CNN) and electronic (TEM)) from one asset. Also, their results showed no significant effect on such factor. A reason for this can be that the proposed solutions were designed agnostic to it. However, future researches described above may require its consideration, and so their validation.

Another relevant aspect regarding this research result's validity is the laboratory conditions in which monitoring experiments were conducted. The student decided to perform laboratory experiments due to various reasons. First, experiments comprised testers conducting fault-finding tasks. They aimed to evaluate tasks' completion time and errors as a direct measure of efficiency. Hence, laboratory conditions were necessary to keep other factors with potential effects over the response variables (e.g. ergonomics or environmental) at constant levels. Second, experiments required of assets to which failure conditions could be setup repeatedly. If an operational asset was to be tested in real-life conditions, then it would have been unlikely to experiment with the same failure. If such asset was set down for experimental purposes, then it would have required resources that exceeded those available for this research at its initial stage. Third, experiments required testers with similar expertise levels to analyse its effect on diagnosis efficiency. If experiments were conducted in real-life conditions, they would not have been as useful without real-life maintainers. And maintainers would have been difficult to group in few expertise levels. So, the student estimated that such expertise' factor levels would require a bigger sample size of an order of magnitude higher than the current system's implementation. Overall, experimented asset and failures were replicated from real-life conditions and other potentially relevant factors (e.g. expertise) were considered. Also, the sample calculated was similar to that of similar experiments on maintenance efficiency (Zhou, Yu and Zhang, 2015; X Wang, Ong and Nee, 2016a). Hence, experimental results analysis can be considered valid in that context at this research's initial stage. Besides, future works described above may require including real-life conditions as part of their experiments when considering additional factors like ergonomics or environment.

3.7 Conclusions and future works

3.7.1 Conclusions

This chapter presented (1) an ontology to describe diagnosis rationale using quantitative and qualitative measures, (2) a reporting method to capture experts' diagnosis knowledge and (3) a monitoring method to re-use that knowledge. Their aim was to demonstrate that expert knowledge capture and re-use can improve efficiency of fault-finding tasks. These methods have been implemented in a cloud-based system prototype for validation with subject-matter experts, testers of different IT expertise, and assets and failures of different natures. Validation results indicated that they have a higher impact on efficiency of fault-finding tasks in terms of time and errors reduction than common data-driven monitoring approaches, and it is dependent on user's expertise. Besides, usability surveys helped to validate their ease-of-use and effectiveness in real working conditions.

The proposed solutions contributes to fill an important research gap towards the integration of expert knowledge in condition monitoring for improving fault-finding efficiency. For complex engineering assets, data-driven and analytical monitoring approaches have certain limitations regarding the failure modes they can identify and control, leading to NFF events. The proposed reporting method can capture experts' diagnosis knowledge in a structured manner to increase failure modes identified. Then, the proposed monitoring method can re-use that knowledge to increase failure modes being controlled. Hence, this research's proposal can be considered as an approach to bridge the gap for integrated diagnosis data management in the context of Industry 4.0.

3.7.2 Future works

Future works will explore further applications and improvements of these methods beyond the diagnosis focus in the context of Industry 4.0. A relevant aspect to consider relates to the kinds of measures used by the proposed methods for faults' evaluation ('States' of 'Components'). At present, the proposed methods focus on certain qualitative and quantitative datatypes including sets' elements and numerical values. There already exist software methods that monitor the condition of components using other datatypes such as audio or video (Henriquez *et al.*, 2014). Because experts can also diagnose faults using such measures, the proposed methods could be enhanced if those measures could be reported. Hence, future research should focus on enabling capture of other datatypes within the reporting process (e.g. audio capture).

Besides additional datatypes for reporting and monitoring, there are other aspects to improve in the proposed methods. One of them is interfaces' usability. Usability surveys' results indicated no negative impact of the proposed interfaces on knowledge capture (reporting) or diagnosis efficiency (monitoring). Although additional factors to consider in real-life experimentation (e.g. ergonomics) may refute this. Different information visualisation techniques, such as Augmented Reality, have been proven useful for monitoring operations (Zollmann, Hoppe and Kluckner, 2014); but its application into diagnosis reporting procedures has had limited literature coverage (Fernández Del Amo *et al.*, 2018). So, future works should evaluate the effects of such techniques regarding the quality and efficiency of diagnosis reporting tasks.

Another factor that experimental results found significant regarding diagnosis efficiency was users' expertise. The proposed reporting method was tested solely with experts, in order to be able to assume the correctness of the resultant reports. But monitoring efficiency experiments found that time and errors reduction rates were higher for low-expertise testers compared to higher-level ones. So, it seems reasonable to believe that knowledge re-use depends on previous experience. Future works should then try to improve the proposed monitoring method using adaption techniques (e.g. recommender algorithms) according to user's experience. They should also consider additional factors to rank failure's recommendations such as assets' similarity or environmental conditions. Besides, this approach could also serve to enable diagnosis reporting for non-expert users, which should be another relevant area for future research due to its ability to extend knowledge capture. Future works on recommender techniques for diagnosis reporting should also consider accuracy and correctness of reported diagnosis tasks as relevant factors. These future works described above consider additional factors than those evaluated within the experiments (e.g. ergonomics and environment). So, they should also include experimentation in real-life conditions to re-confirm this research's conclusions.

Previously proposed future works study additional improvements that could be achieved in diagnosis efficiency regarding the proposed methods. Because these methods aim to integrate unstructured data in Industry 4.0 context, future works should also investigate their applicability in other human-related tasks (e.g. repair or manufacturing) and the relations between those. For example, future research could develop ontology-based repair recommender methods based on diagnosis and repair reports. Thus, aiming towards the integration of causes and effects for improving maintenance efficiency. Overall, ontology-based approaches can be useful to integrate sources of heterogenous data and future research should investigate frameworks to apply them to resolve that relevant research challenges in the context of Industry 4.0.

Chapter 4
Programmable content and a pattern-matching algorithm for automatic adaptive authoring in Augmented Reality knowledge transfer maintenance applications

4.1 Introduction

Industry 4.0 is increasing the importance and criticality of maintenance operations (Angelopoulos *et al.*, 2020). Digital technologies like Cyber-physical systems or Artificial Intelligence (AI) aim to increase availability, safety and sustainment of assets to enable more intelligent and adaptive production systems (Gattullo *et al.*, 2019). Besides, humans are starting to attract more attention in Industry 4.0 due to their ability to use knowledge ubiquitously for effective decision making (Egger and Masood, 2020). Digital technologies such as Augmented Reality (AR) or Semantic Web (SW) aim to support them with real-time knowledge access for increased interaction with smart production systems (Longo, Nicoletti and Padovano, 2019). In maintenance operations, the human impact on efficiency for tasks like repair or diagnosis is driving digital technologies research to achieve improved asset availability (Bottani and Vignali, 2019).

AR enables effective knowledge transfer by embedding virtual information into the user's space in correlation with real-world objects (Palmarini *et al.*, 2018). With a predicted market size of $12.35bn by 2025 (Hall and Takahashi, 2017), AR industrial applications are wihtin the firsts closing the gap to commercial deployment (Capgemini Research Institute, 2018). Despite AR applications' maturity, there are some research challenges yet to be solved. A relevant one is the *"semantic understanding of real-world objects in large environments without emplaced infrastructure"* (Azuma, 2016). It involves the AR procedure to create augmented content that maintainers can visualise and interact with to understand their tasks at hand and related real-world objects (Bottani and Vignali, 2019). This so-called authoring process consists of utilising content formats (e.g. text, 3D models, animations, etc.) and embedding them with virtual information (e.g. instructions, sensor data, etc.) (X. Wang, Ong and Nee, 2016). AR authoring has recently fostered lots of research attention to investigate cost-effective and user-adaptive solutions (Akçayir and Akçayir, 2017; Bernhardt *et al.*, 2017; Bottani and Vignali, 2019).

In maintenance applications, AR research has mainly proposed two authoring approaches. One focuses on manually creating content with higher adaptiveness (e.g. user expertise or light conditions) to multiple tasks like monitoring and repair (X Wang, Ong and Nee, 2016b; Longo, Nicoletti and Padovano, 2019). The other aims to automate content creation for very specific tasks like assembly, reducing AR deployment costs at the expense of less adaptive content (Ramirez-Amaro, Beetz and Cheng, 2017; Chang, Nee and Ong, 2020). Both approaches propose content formats to support maintenance tasks at hand, although they require specific database schemas to create augmented content rather than those of existing maintenance information systems (Gattullo *et al.*, 2019). If authoring methods can embed virtual information from existing schemas in adaptive content formats, then they will be able to automate authoring without losing content adaptiveness. Thus, obtaining AR applications with adaptive content for multiple maintenance operations and reduced authoring costs.

In order to automate adaptive content creation in AR applications for diverse maintenance operations, this chapter proposes a real-time, ontology-based, pattern-matching technique. This proposal includes the following contributions:

1. A method to declare content formats in a programmable manner. So, content can be designed according to its visualisation and interaction requisites independently of the specific maintenance information it contains.
2. A pattern-matching algorithm to pair programmable content formats with ontology-based information sets according to data, user and environmental requisites. So, content can be produced automatically at real-time for any given ontology-based data structure.

This research aims to demonstrate that AR-maintenance applications can automatically produce adaptive content without pre-specified data structures that achieves sufficient levels of maintenance support compared to that of manual authoring approaches. For this reason, this research validates the contributions above through comparison with alternative authoring solutions for two different maintenance operations: repair and remote diagnosis.

The rest of this chapter's structure is as follows. Section 4.2 presents a literature review on authoring in AR-maintenance applications to detect current research gaps. Section 4.3 describes the methodology utilised to identify, develop and validate this research contributions, which are explained in Section 4.4. Section 4.5 presents the validation protocol, including experimental and survey methods, compared authoring solutions and their cases of study. Validation results are analysed and discussed in Section 4.6. Finally, Section 4.7 describes this research conclusions and future works, including the implications of automatic authoring for maintenance knowledge transfer in Industry 4.0 contexts.

4.2 Literature review

4.2.1 Authoring methods in Augmented Reality for maintenance

AR technologies rely on authoring methods to enable effective knowledge transfer (Palmarini *et al.*, 2018). Authoring refers to the set of software techniques aiming to create augmented content for AR applications (Ong and Zhu, 2013). These methods have the ability to automate AR deployment and to reduce its costs (Bottani and Vignali, 2019). For this reason, they have recently attracted lots of research attention for AR applications such as medicine (Bernhardt *et al.*, 2017), education (Akçayir and Akçayir, 2017), manufacturing (Bottani and Vignali, 2019) or maintenance (Palmarini *et al.*, 2018). In a previous research (Chapter 2), the author identified three relevant elements to describe authoring methods in AR-maintenance contexts: scope, applications and users. The following provides an update of that research based on latest publications from 2017 to 2020.

Authoring scope refers to the augmentation processes that authoring methods conduct (Palmarini *et al.*, 2018). These can include virtual content generation (e.g. 3D models or animations) and content behaviour declaration (e.g. step-by-step, multi-modal views, etc.) (Fernández Del Amo *et al.*, 2018). For example, Van Lopik, et al. (2020) presented an authoring method for shop-floor operators to generate AR assembly content in the form of text, images and video. A different approach is that from Blattgerste, Renner and Pfeiffer (2019), they proposed a web-based authoring tool to augment repair instructions using predefined 3D models to generate animations. Besides, Cao et al. (2019) proposed an advanced authoring method for repair tasks that enabled automatic animations' generation through real-time user tracking. Thus, embodying the authoring process within real-life operations. A common aspect of these methods is that they pre-determine the behaviour of the augmented content to be authored and so, simplify the content creation process (Flotyński and Walczak, 2017). The author encountered very few examples of authoring techniques that enable to create content with different behaviours. For instance, Wang et al. (X Wang, Ong and Nee, 2016a) proposed a context-aware authoring technique to adapt content formats (e.g. animations, text) for assembly tasks according to the user's cognition phase.

Authoring applications refer to the specific maintenance operations (e.g. repair or diagnosis) that authoring methods focus on. For example, Zubizarreta, Aguinaga and Amundarain (2019) presented a desktop authoring interface to generate instruction's animations for repair tasks. That is also the focus of Gattullo et al. (2019) proposal, who use 2D icons to simplify the semantic understanding of repair instructions. Instead, He, Ong and Nee (2019) described an authoring method to generate augmented annotations for users and sensors to register relevant monitoring data such as service records or usage logs. More advanced techniques have been proposed for more specific applications such as assembly or machine setup. Chang, Nee and Ong (2020) presented an authoring technique that utilises assembly planning algorithms to automatically generate augmented animations. A similar approach has been taken by Tzimas, Vosniakos and Matsas (2019) to provide instructions for setting up machine tools. Nevertheless, the student has been able to discover fewer authoring proposals that are applicable to more than one maintenance operation. Longo, Nicoletti and Padovano (2019) proposal is one example. They used AR authoring and Digital Twin technologies to enable repair task's description with virtual animations and real-time sensor data visualisation for diagnosis.

Authoring users refer to the targeted consumers of AR authoring methods. There are two common user types of authoring tools: application users and application-domain experts (Palmarini *et al.*, 2018). For example, the methods presented by Erkoyuncu et al. (2017) or Akbarinasaji and Homayounvala (2017) aimed for repair experts to produce augmented content. Meanwhile, there are others that aim for any AR user to produce content such as the proposals by Flatt, et al. (2015) or He, Ong and Nee (2019). Besides, latest research reviewed has aimed to further reduce AR deployment costs by developing automatic authoring methods like those from Chang, Nee and Ong (2020) or Wang, Ong and Nee (2016a). However, most automatic authoring methods are applied to very specific maintenance operations and do not propose further integration with others.

Overall, authoring methods aim to ease the content creation process for reducing AR deployment costs while enhancing users' understanding of maintenance tasks. Nevertheless, the author has found very few publications that include research on semantic computing for AR authoring (Chapter 2 - (Fernández Del Amo *et al.*, 2018)). Besides other semantic computing methods, ontologies are those that have shown more promising results for enabling semantic understanding of real-world objects through augmented content (Flotyński and Walczak, 2017).

4.2.2 Ontologies in Augmented Reality authoring

Ontologies are a set of statements and rules that capture and specify vocabulary within software applications. Thereby, making it understandable and processable by both humans and computers (Breitman, Casanova and Truszkowski, 2007; Flotyński and Walczak, 2017). In Augmented Reality research, ontologies have been used for various purposes in different contexts. These range from declaring direct relations between 3D models and real objects for augmenting museums' items (Walczak and Flotyński, 2019) to identify user's cognition in assembly tasks for adapting augmented content accordingly (X Wang, Ong and Nee, 2016a).

In the context of maintenance applications, AR research on ontologies has been mostly related to context-awareness techniques. For example, Wang, Ong and Nee (2016a) utilised ontologies to describe the assembly domain, including user's cognition as part of that description to enable context-aware authoring. Another approach was followed by Longo, Nicoletti and Padovano (2019) or Zhu, Ong and Nee (2015), who utilised ontologies to describe repair instructions and correlate them with specific, pre-built augmented content. A similar concept is presented by Vincent et al. (2017) and Akbarinasaji and Homayounvala (2017). They added ontological descriptions of monitoring tasks to infer the behaviour of augmented content for repair instructions. Park et al. (2013) took a different angle and proposed ontologies to model defect management in construction sites for enabling data collection by operators using AR.

Overall, ontologies in AR authoring have been used to model augmented content's behaviour according to the maintenance operation it supports. The same concept has also been applied in other areas of AR application such as tourism (Walczak and Flotyński, 2019) or education (Djordjevic, Petrovic and Tosic, 2019). Ontologies in AR authoring allow to separate the content generation process from its application's behaviour and to enhance its semantic understanding. However, these methods remain costly in terms of AR deployment as augmented content still needs to be created in advance, increasing sustainment of AR applications.

4.2.3 Research gaps

Latest research identified on AR authoring methods for maintenance applications aims to solve two main challenges: (1) semantic understanding of real-world objects (Azuma, 2016) and (2) reduce complexity and costs of AR deployment (Bottani and Vignali, 2019). Most authoring methods for AR maintenance applications focused on single operations (e.g. repair or diagnosis) and provide tools for users or application-domain experts to generate content (Flatt *et al.*, 2015; He, Ong and Nee, 2019). In order to further reduce AR costs, there are fewer proposals that enable automatic content creation, although they are very specific (e.g. assembly) and difficult to apply to other maintenance tasks. Besides, other authoring proposals used ontologies to enhance semantic understanding of maintenance operations (Zhu, Ong and Nee, 2015; X Wang, Ong and Nee, 2016a). Most of which described application domains and connected specific instructions with pre-built augmented content (Vincent *et al.*, 2017; Longo, Nicoletti and Padovano, 2019). Although these managed to divide content generation from its application behaviour, these methods still require for users or experts to produce augmented content in advance.

Based on the discussions above, the student identified several research gaps regarding AR authoring methods for maintenance. First, little academic evidence has been found on automatic authoring techniques that can integrate more than one maintenance operation into one AR application. Second, there is also little evidence of ontology-based authoring methods that re-utilise content formats for diverse maintenance tasks. If both research gaps were to be fulfilled by one authoring method, such technique should manage to: (1) declare programmable content formats that can load diverse maintenance datasets and (2) pair these formats with datasets to automatically create augmented content. Thus, reducing deployment costs and achieving sufficient augmented content understanding to attain operational efficiency improvements. The next section proposes a method that aims to fulfil these requisites to achieve automatic and adaptive authoring for multiple maintenance operations in one single AR application.

4.3 Methodology

This research aims to prove that AR authoring methods can automatically create user- and environment-adaptable content without pre-defined data structures that can provide sufficient support for multiple maintenance operations. Inspired by similar research works in the field (Gimeno *et al.*, 2013; Flotyński and Walczak, 2015; X Wang, Ong and Nee, 2016b; Longo, Nicoletti and Padovano, 2019) and well-established frameworks in design research (Nuñez and Borsato, 2017), this research conducted the following methodological steps (Peffers *et al.*, 2008):

1. Objectives identification: "define specific opportunities and justify the value of a solution". Section 4.2 presented a literature review to identify academic research gaps that followed the protocol presented by Booth (Booth, Papaioannou and Sutton, 2012).

2. Solution design: "create a solution to satisfy the research opportunities". Section 4.4 presents the proposed authoring-related software methods designed to satisfy AR-maintenance research contributions.

3. Solution demonstration: "prove the solution's use to satisfy the research opportunities". Section 4.5 presents this research application of feature-driven software development (Nawaz, Aftab and Anwer, 2017) to implement the proposed solution in an AR prototype for two experimental cases of study.

4. Research validation: "measure the solution's impact on research opportunities". Efficiency experiments and usability surveys evaluated the impact of this research's authoring technique on maintenance tasks through comparison against alternative authoring tools. Section 4.5 describes their validation protocol, while Section 4.6 discusses their results. Finally, Section 4.7 summarises this research conclusions and suggests future works.

4.4 Programmable content and pattern-matching for real-time automatic authoring

This chapter proposes a real-time, ontology-based, pattern-matching technique for automatic adaptive authoring in AR-maintenance applications. This technique consists of the following contributions:

1. A method to declare programmable content formats according to data, user and environment requisites. Programmable content formats aim to provide specific functionalities in terms of visualisation and interaction that can be used to augment varying maintenance information.

2. A real-time ontology-based pattern-matching algorithm to couple content formats with ontology individuals. It aims to automatically create augmented content for existing ontological datasets.

Figure 4-1 presents this technique compared to conventional manual authoring methods such as those presented by Longo, Nicoletti and Padovano (2019) and Zhu, Ong and Nee (2015). In conventional approaches, there normally are four software modules (Figure 4-1-a): information management, development interface, authoring interface and AR application. Experts utilise the information system to store all maintenance-related information, from repair tasks to monitoring data, using various interfaces. Programmers use development interfaces to generate content formats for augmented content generation and visualisation according to certain maintenance operational logic. Experts also use the authoring interfaces to create content from pre-programmed formats based on their own knowledge or other data stored in the information system. Finally, maintainers can use the AR application to visualise pre-created augmented content. In conventional methods, experts duplicate their efforts to input maintenance-related data since they need to use both information management and authoring interfaces. To eliminate that duplicity, one option would be for authoring interfaces to be set as unique points for information input, but that would have two limitations. First, maintenance experts would still need to acquire skills on AR visualisations and interactions for learning how to use content formats. Second, it would require major efforts to sustain AR applications and authoring interfaces if maintenance operations were modified, for example from corrective to preventive. The reason for it is that they implicitly correlate visualisation and interaction modes (e.g. animation) with maintenance operations' rationale (e.g. task).

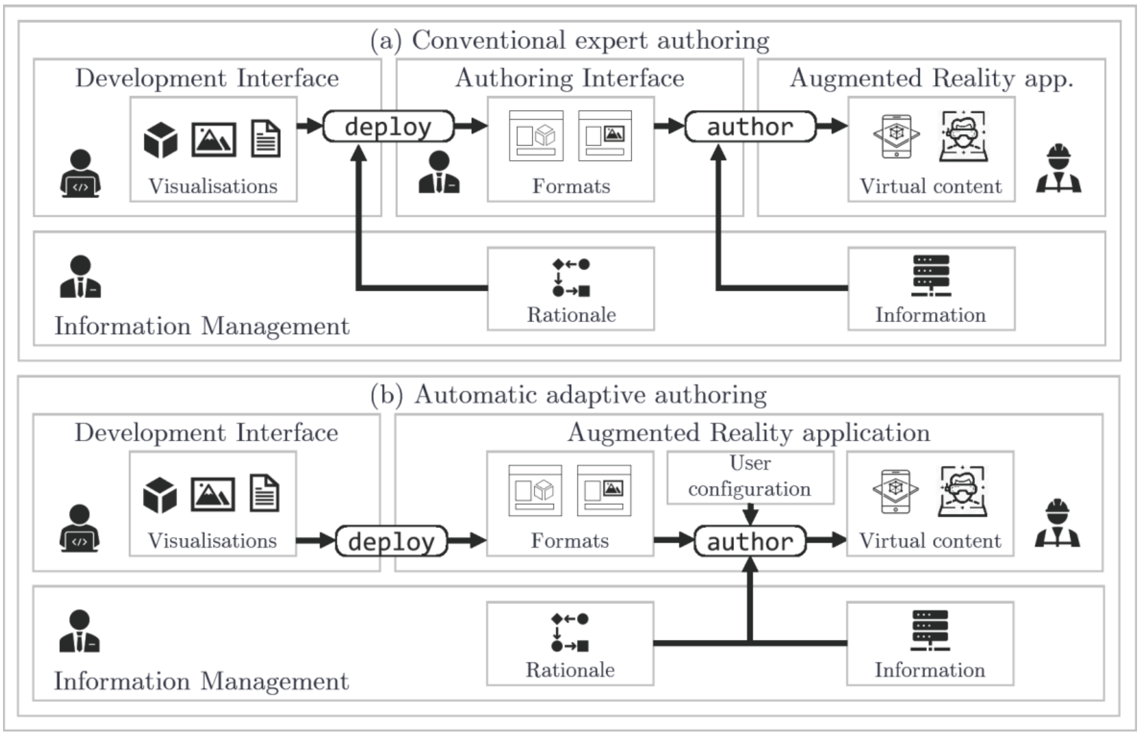


Figure 4-1. Overview of (b) the automatic adaptive authoring proposal for knowledge transfer compared to (a) conventional expert authoring approaches – It compares the relations among software modules (development interface, authoring interface, information management and Augmented Reality application) of both authoring approaches.

Another option to eliminate experts' efforts duplicity would be to automate the authoring process. However, this option would require to explicit the relation between visualisation and interaction modes and content's behaviour representing application's rationale (e.g. step-by-step). Thus, detaching authoring's content-creation and information management processes. That is the purpose of this research's authoring proposal, which uses maintenance rationale and information to author content formats for augmentation. Figure 4-1-b explains this research authoring proposal. Programmers develop content formats (e.g. image) but instead of linking them to specific datasets (e.g. error), they declare the requisites to which datasets must complain with (e.g. png file) to be augmented as such. These requisites, so-called facets, can be of diverse natures such as data, user or environmental. Experts use information management interfaces (Figure 4-1-b) to input maintenance information as well as to declare maintenance operation's rationale. Then, the proposed authoring method, from within the AR application, analyses the incoming maintenance dataset, identifies the most suitable content formats by facets to visualise it, and locates the resultant augmented content in the user's space. Figure 4-2 displays examples of augmented content automatically created by this research authoring proposal using various formats to represent data from diverse maintenance operations.

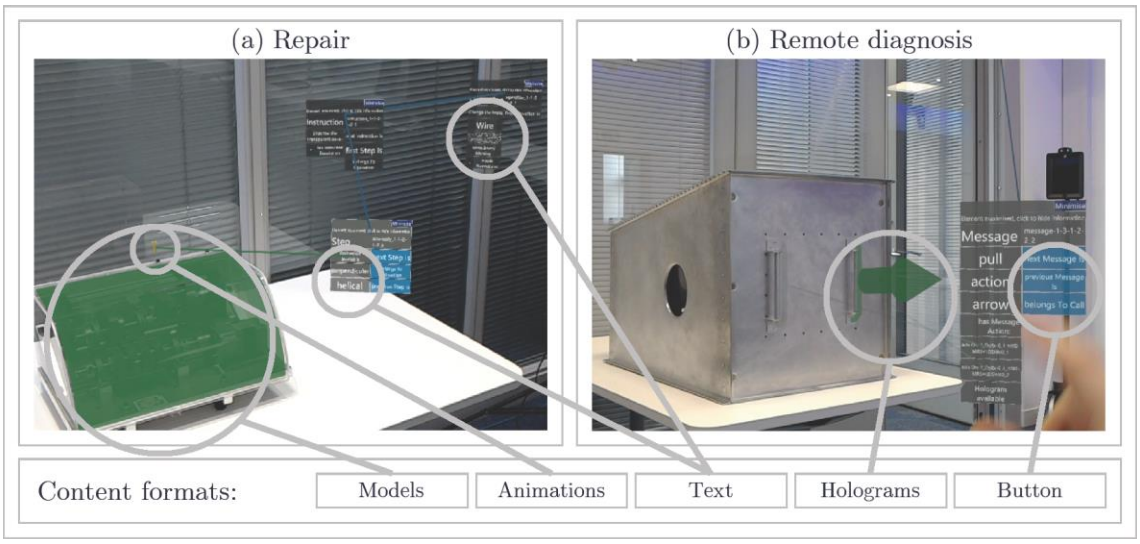


Figure 4-2. Examples of automatically created augmented content for repair and remote diagnosis operations – The different types of content formats are circled in the screenshots of the resultant AR application's prototype.

This research authoring algorithm runs in real-time because content is augmented on user's demand, so the user can navigate through an augmented version of the maintenance information system. The algorithm is also ontology-based because it uses some features of OWL and RDFS languages (Smith, Welty and McGuinness, 2004) (e.g. rdf:type) to standardise maintenance datasets structures for semantic analysis. Nevertheless, it does not use common ontology inferencing (SWRL rules) to enable datasets (ontology individuals) assignment to more than one content format, thus allowing for content adaptiveness. Therefore, if that standardisation is achieved by different means, then there is no reason to believe this technique can be applied with other information management techniques such as SQL or graphical databases. Finally, the algorithm can be considered pattern-matching as it pairs ontology individuals with multiple content formats for augmentation. The different variables used by the algorithm like formats, facets and individuals are declared in Subsection 4.4.1. Then, Subsection 4.4.2 describes the algorithm itself along with the rules proposed to assign, discard, and locate ontology individuals and content formats for AR authoring.

4.4.1 Programmable content and maintenance ontology individuals

Programmable content formats are templates with specific visualisation and interaction modes that can load diverse data sets to create augmented content. For example, animations are 4-dimensional visualisations with no user interaction that can represent component's movements (Erkoyuncu *et al.*, 2017). The code to generate animations' movements is the content format and the data defining the movement (e.g. degrees of freedom, component's CAD file, etc.) can be created separately. Unlike conventional authoring solutions, content formats aim to be reused for diverse maintenance datasets (Fernández Del Amo *et al.*, 2018). So, they declare different requisites that enable to load them automatically. For creating augmented content by loading of formats, this method uses maintenance data in the form of ontology individuals. These enable to standardise data for further semantical analysis.

Ontology individuals (in) are instances of classes that can be declared by the properties (pr_{in}) asserted to them.

$$in = (pr_1, pr_2, \cdots, pr_m, d_{cp}, d_{op}) \tag{4-1}$$

$$m = number\ of\ asserted\ properties\ to\ individual$$

Properties (pr) are axioms asserting attribute or relationship values of an individual, which can be described by a vector of resources linked to its assertion: name, range, value and type. Properties store data types like strings, numbers and URI's to other individuals or specific files (e.g. png, obj, etc.).

$$pr_{in} = (res_{in,m,na}, res_{in,m,ra}, res_{in,m,ty}, res_{in,m,va}) \qquad (4\text{-}2)$$
$$res_{in,m} \in URI \cup S \cup \mathbb{R}$$

Besides properties, this authoring proposal uses other individuals' features. Distances (d) are the minimum number of relationships connecting two classes inferred from ontologies' schemas. These quantitatively describe ontologies and enable contents' non-occlusive allocation in the augmented space.

$$d_{in,c}: C_{in} \times C_c \to [0, \infty) \qquad (4\text{-}3)$$
$$c \in \{component, operation\} = class\ to\ \text{measure}\ distance$$

Besides other ontological classes AR users may want to augment, the proposed authoring method assumes the existence of two classes in maintenance ontologies: component and operation. Component class refers to the equipment's component level at which maintenance operations can be done. Operation class describes the maximum level of abstraction for the maintenance operation to be augmented. For example, if a repair operation comprises several steps, then the class of the individual which groups all steps under on task will be the operation class. Distances to these classes helps the algorithm to determine the location of content in the augmented scene. The content allocation is explained in Subsection 4.4.2.

Content formats (ft) can be then declared by its interaction and visualisation augmentation modes and the requisites for the data they can augment. These features are referred to as facets and the proposed method considers three types: user, environment and data.

$$ft = \ uf, ef, df_1, df_2, \cdots, df_n \qquad (4\text{-}4)$$
$$n = number\ of\ data\ facets\ declared\ for\ a\ format$$

User facets (uf) designate formats' perception modes in terms of dimensions (dm) and descriptiveness (ds).

$$uf_{ft} = f(dm_{ft}, ds_{ft}) \tag{4-5}$$
$$dm_{ft} \in \{1D, 2D, 3D, 4D\}, ds_{ft} \in \{symbolic, literal\}$$

Environment facets (ef) describe formats' visualisation (vs) and interaction (in) modes in terms of human senses (sen).

$$ef_{ft} = f(vs_{ft}, in_{ft}) \tag{4-6}$$
$$vs_{ft} = (sen_{1,}, sen_2, \cdots, sen_p)$$
$$vs_{ft} \in \{text, image, audio, video, hologram, anim\}$$
$$in_{ft} = (sen_1, sen_2, \cdots, sen_q)$$
$$in_{ft} \in \{tap, hold, manipulate, keyboard, dictate\}$$
$$sen_n \in \{hearing, sight, kinaesthetic, touch, smell\}$$

Data facets (df) describe semantically formats requisites to load maintenance data as augmented content in terms of properties assertions (na, ra, ty, va).

$$df_{ft,n} = (rq_{ft,n}, dr_{ft,n,na}, dr_{ft,n,ra}, dr_{ft,n,ty}, dr_{ft,n,va}) \tag{4-7}$$
$$rq_{ft,n} \in \{source, optional\}$$

Data facets comprise data rules ($dr_{ft,n,o}$) for each property assertion (na, ra, ty, va), which determine if a property can be augmented with their format.

$$dr_{ft,n,o} = (qu_{ft,n,o}, set_{ft,n,o}) \tag{4-8}$$
$$qu \in \{\forall, \exists, \exists!\} = \ assignment\ method\ to\ rule\ set$$
$$set_{ft,n,o} = (s_{ft,n,o,1}, s_{ft,n,o,2}, \cdots, s_{ft,n,o,q})$$
$$s_{ft,n,o,q} \in S = string\ of\ format's\ behaviour\ feature$$

The proposed authoring method matches properties and formats to create fabrications (*fab*). They are the resultant augmented content that AR users can visualise and interact.

$$fab = (ft, pr_{in,1}, \cdots, pr_{in,s}) \tag{4-9}$$

Fabrications contain subsets of properties and elements (*el*) are groups of fabrications that include all properties of an individual.

$$el = \ \ fab_1, \cdots, fab_r, loc \ \ \rightarrow \forall pr_{in,m} \in in, \exists fab_r \supset pr_{in,m} \tag{4-10}$$

Elements can have diverse locations (*loc*) in the augmented space according to individual's distances. The proposed method defines three locations around the maintaining asset to augment ontology data: primary, secondary and tertiary.

$$loc_{el} = f(d_{cp}, d_{op}) = \begin{cases} primary, & d_{cp} \leq 1 \\ secondary, & d_{op} > 1 \\ tertiary, & otherwise \end{cases} \tag{4-11}$$

Formats comprise what can be defined as visualisation (e.g. 3D, symbolic) and interaction (e.g. tap, dictate) modes, which are the means that AR users perceive and act over augmented content. They can therefore be used to semantically classify fabrications with different behaviours that are similar in nature. Figure 4-3 presents some examples of resultant fabrications from several formats classified by their user facets (4-6). Colour coding has been used to differentiate behaviours of different formats' visualisations: green is used for static 3D (models or holograms), yellow for dynamic 3D or 4D (animations) and blue (buttons) and grey (text) for 1D. For the purpose of this research, content formats described in previous authoring-related researches (Erkoyuncu *et al.*, 2017; Fernández del Amo *et al.*, 2019) have been implemented. The reason was to validate that the proposed authoring method can embed content formats from multiple authoring approaches into a single one. Thus, allowing to augment adaptive content for multiple maintenance operations from within one single AR application and independently of the information structure being used. Implemented content formats their data, user and environment facets can be consulted at 10.17862/cranfield.rd.12213380. Besides, the following subsection explains the proposed authoring algorithm to matches formats with individuals to automatically create augmented content.

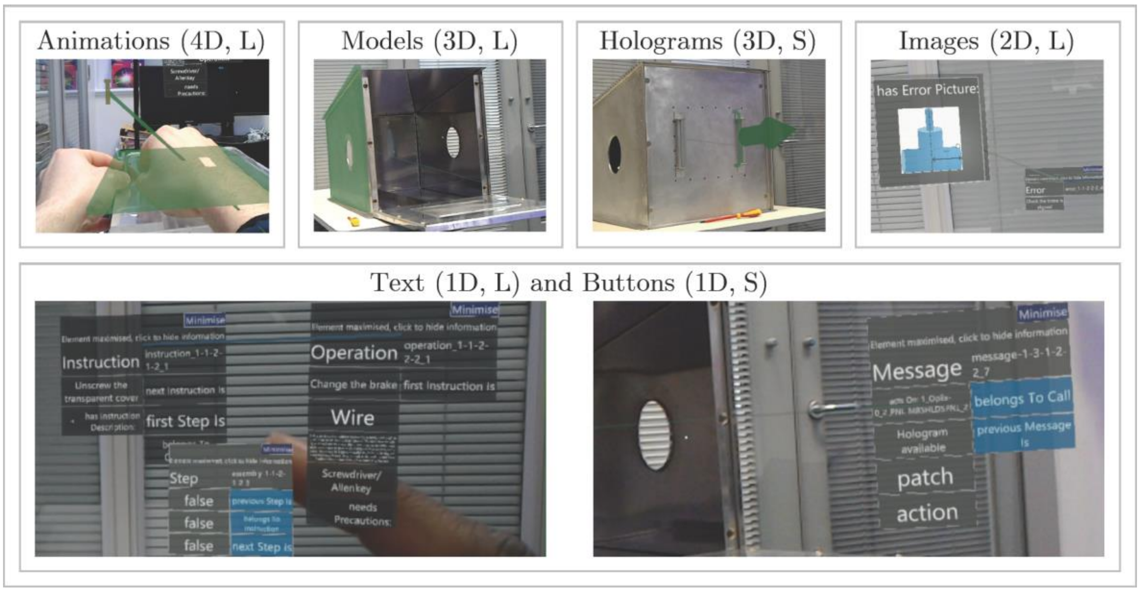


Figure 4-3. Examples of fabrications from implemented visualisation modes – These are shown in screenshots from the AR application's prototype and classified by user facets using the following notation *uf(dm, ds)* as in (4-5) where *dm* = dimensions (1D, 2D, 3D, 4D) and *ds* = descriptiveness (Symbolic, Literal).

4.4.2 Real-time ontology-based pattern-matching authoring algorithm

The proposed pattern-matching algorithm analyses facets and properties to pair formats with a given individual for automatically creating content (fabrications). It uses three types of rules to match formats and individuals: discard, assignation and allocation.

Discard rules ($fuef$) reject available formats that do not meet user-selected dimensions, descriptiveness, interaction and visualisation modes.

$$fuef = \begin{cases} ft, & \exists(sen_{ft,vs} \vee sen_{ft,in} \vee co_{ft,uf} \vee ds_{ft,uf}) \notin ac \\ \emptyset, & otherwise \end{cases} \tag{4-12}$$

Within the application, an AR user can modify algorithm's augmentation configuration (ac) in terms of dimensions, descriptiveness, interactions and visualisations. This allows the algorithm to match different formats to the same properties for adapting augmented content to user's requirements.

$$ac = (sen_{ac,1},,, sen_{aci}, co_{ac,1},,, co_{ac,j}, ds_{ac,1},,, ds_{ac,k}) \tag{4-13}$$

Assignation rules correlate fabrications' formats and properties that meet formats' data facets. For example, given a property of name "nextIs", a pairing rule (fdr) checks if the name includes either all ($\forall$), any ($\exists$) or one ($\exists!$) of the items in the rule's set (e.g. "next", "continue", etc.). If so, the rule will return true, otherwise false.

$$\begin{aligned} fdr &= f(dr_{ft,n,o}, res_{in,m,o}) = \\ &= \begin{cases} 1, & (qu_{ft,n,o})s_{ft,n,o,q} \in set_{ft,n,o} \to s_{ft,n,o,q} \subset res_{in,m,o} \\ 0, & otherwise \end{cases} \\ o &\in \{na, ra, ty, va\} \end{aligned} \tag{4-14}$$

If all the rules of a given facet are true, then the format's facet is assigned (fdf) to that property.

$$\begin{aligned} fdf &= f(df_{ft,n}, pr_{in,m}) = \\ &= \begin{cases} (df_{ft,n}, pr_{in,m}), & \forall o, \exists (dr_{ft,n,o}, res_{in,m,o}) \to fdr_{ft,n,o} = 1 \\ \emptyset, & otherwise \end{cases} \end{aligned} \tag{4-15}$$

These facet rules are run for all individual's properties. So, if all "source" facets are assigned, then a fabrication is created ($ffab$) by pairing format and matched individual's properties.

$$ffab = f(ft, pr_{in,1}, \cdots, pr_{in,s}) = \begin{cases} (ft, pr_{in,1}, \cdots, pr_{in,s}), & ** \\ \emptyset, & otherwise \end{cases} \tag{4-16}$$

$$** \ \forall df_{ft,n}, rq = source, \exists pr_{in,m} \to fdf_{ft,n} = (df_{ft,n}, pr_{in,m})$$

Table 4-1 presents the proposed algorithm. It utilises the rules described above to semantically analyse individuals and formats for generating fabrications. Besides, it considers additional rules to ensure that resultant fabrications are not equivalent to each other and that all individual properties are assigned.

Table 4-1. Pattern-matching algorithm for automatic authoring by semantic analysis of formats and individuals.

Algorithm 1:	Match formats to properties for creating elements and fabrications	
Inputs:	ac	user's configuration of augmented scene
	in	ontology individual to be augmented
	FT	set of available formats for augmentation
Outputs:	el	augmented element as combination of fabrications
Procedure:	$MatchFormatsToProperties(in, ac, FT)$	

01:	$FT' \longleftarrow \{\}$	»Assignable formats
02:	$FC' \longleftarrow \{\}$	»Assignable facets
03:	$FAB \longleftarrow \{\}$	»Assignable fabrications
04:	**for each** $ft \in FT$	»Resolve assign.
05:	$FT' \longleftarrow fuef(ft, ac)$	formats
06:	**for each** $pr_m \in in$	»Resolve assignable
07:	**for each** $ft \in FT'$	facets
08:	**for each** $df_n \in ft$	
09:	$FC' \longleftarrow fdf(df_n, pr_m)$	
10:	**for each** $ft \in FT'$	»Resolve assignable
11:	**for each** $(df_n, pr_m) \in FC'$	fabrications
12:	$FAB \longleftarrow ffab(ft, pr_{in,1}, \cdots, pr_{in,s})$	
13:	**for each** $vs' \in VS$	»Discard assignable
15:	**for each** $pr_m \in in$	fabrications with
16:	**find all** $fab' \rightarrow vs_{ft} = vs' \wedge fdf_{ft,req=source} \supset pr_m$	equivalent source
17:	**calculate** $\sum df$ **for each** fab'	properties and
18:	**remove all** $fab' \in FAB$ with $\sum df < \max(\sum df)$	visualisations
19:	**for each** $pr_m \in in$	»Assign default
20:	**if** $\nexists fab \in FAB \rightarrow fab \supset pr_m$ **then**	fabrications to non-
21:	$FAB \longleftarrow ffab\ dft, pr_m$	matched properties
22:	**calculate** $(d_{in,cp}, d_{in,op})$	»Calculate location dist.
23:	$el \longleftarrow FAB, loc(d_{in,cp}, d_{in,op})$	»Assign fabrications
24:	**return** el	and loc. to element

Once "assignable" fabrications are resolved (Table 4-1 – Line 13), the algorithm identifies for each property all assignable formats with that property assigned to their required data facets and equivalent visualisations. These fabrications may be very similar to each other (e.g. two types animations for the same component). So, the algorithm discards all fabrications but the one with the maximum number of data facets. Therefore, ensuring that only non-equivalent fabrications are augmented. Besides, the algorithm also identifies all the individual properties that have not been assigned to any fabrication (Table 4-1 – Line 19). These properties are assigned to fabrications with default formats to ensure that individuals are always augmented, although they haven't matched any more specific or adaptive formats. Default formats are those with no facets that simply overlay the individual's property as text or a button depending on their property type (datatype or object). Finally, the algorithm merges all assigned fabrications into one single element and uses the individual's distances (component and operation) to locate the element in a non-occlusive zone within the augmented space.

Figure 4-4 presents the proposed algorithm including its non-occlusive element-allocation method. It controls all augmented elements in the scene to ensure that they are overlaid non occlusively around the asset. In order to avoid augmented scene's overload, the algorithm assigns a total of 9 element spots in 3 different areas (primary (1), secondary (4) and tertiary (4)) that correspond to the locations calculated by the algorithm according to individual's distances. These locations reflect ontologies' logic visually for avoiding occlusion. For example, a primary location's element would be required to be as close as possible to the asset because it has direct references to its components. Instead, a tertiary element may not be necessary to be in direct sight at all times because it refers to the overall maintenance operation being conducted. The positions of location's spots are calculated according to the position, rotation and size of the asset (Figure 4-4) in the augmented scene. When an element is produced, the algorithm assigns it a spot at its calculated location in a first-in-first-out basis. The first element augmented in that location is therefore discarded when a new one is allocated and there are not free spots available. After an element has been assigned to a spot, the algorithm adjusts its scale according to the formulae presented in Figure 4-4. Element's width, height and length are applied a re-scaling factor based on asset's width to ensure that elements do not occlude at each other from any perspective.

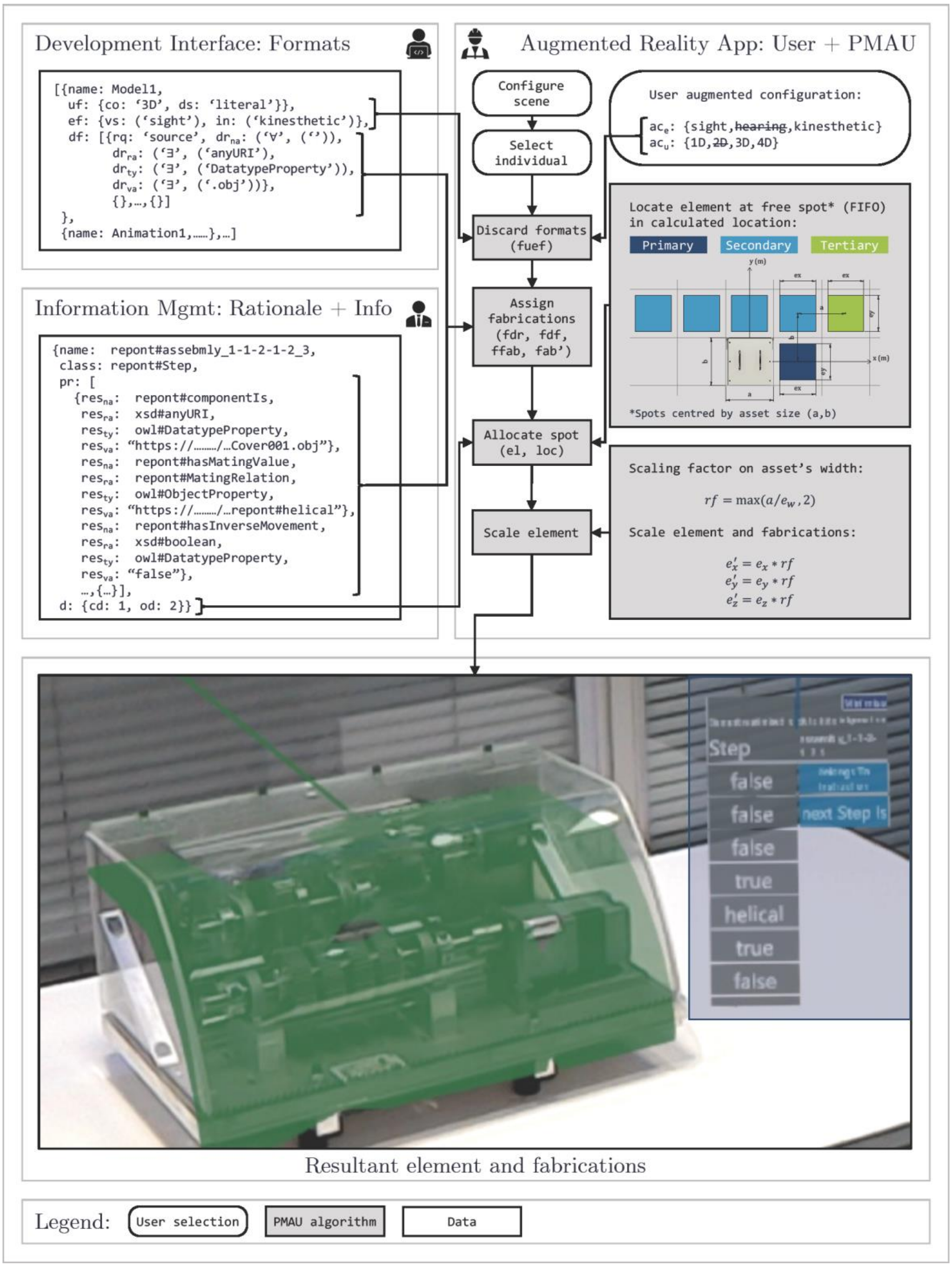


Figure 4-4. Pattern-matching algorithm and content formats for automatic adaptive authoring – The figure details the overview presented in Figure 4-1-b. It exemplifies how the PMAU algorithm (Table 4-1) uses an individual ($in(pr_m, d_{cp}, d_{op})$ – (4-1)), formats ($ft(uf, ef, df_n)$ – (4-4)) and augmentation configuration ($ac(ac_e(sen_i)$, $ac_u(dm_j, ds_k)$ – (4-13)) to automatically generate an element and its fabrications. The figure displays the individual (in), formats (ft) and augmentation configuration (ac)

as JSON objects to show their data. The figure also presents the different variables of the individual, formats and augmentation configuration used by each PMAU step: discard formats (*fuef* – (4-12)), assign fabrications (*fdr* – (4-14), *fdf* – (4-15), *ffab* – (4-16)), locate element (*el* – (4-10), *loc* – (4-11)) and scale element. The resultant exemplary element and its fabrications are displayed at the bottom of the figure. Each box containing data (*in, ft, ac*), PMAU steps and user selections have a different shape and colour as described in the legend.

The proposed algorithm requires real-time implementation because of its interactions with AR users and other AR technologies (e.g. tracking and registration). Therefore, Figure 4-5 presents an AR application's flowchart that embeds the proposed authoring algorithm. The algorithm is activated every time the user selects a button-like fabrication to augment a new individual through an element. Thus, enabling the user to augment maintenance data by navigating the ontology of choice through its asserted object properties. When the user selects a new individual, the algorithm downloads it in real-time along with any files that it refers to ("xsd:anyURI" datatype properties) and that may be needed to generate the fabrication. Besides the algorithm's, Figure 4-5 describes behaviour assumed necessary in AR-maintenance applications. That includes user-configurable real object (asset) registration and tracking as well as automatic content tracing (element creation time and related individual). '`Register Asset`' allows the user to select the equipment to be maintained and the algorithm to define the augmented space around it for element allocation. Automatic content tracing ('`save element creation time`') enables the algorithm to report the content visualised, including times when user selected each individual for visualisation and its creation.

Overall, the programmable content declaration method and the algorithm proposed aim to make the authoring process automatic and adaptive for different maintenance applications, information systems and content formats. To validate the proposed methods, the following section presents the proposal's system implementation along with the protocol for validation in two different cases of study.

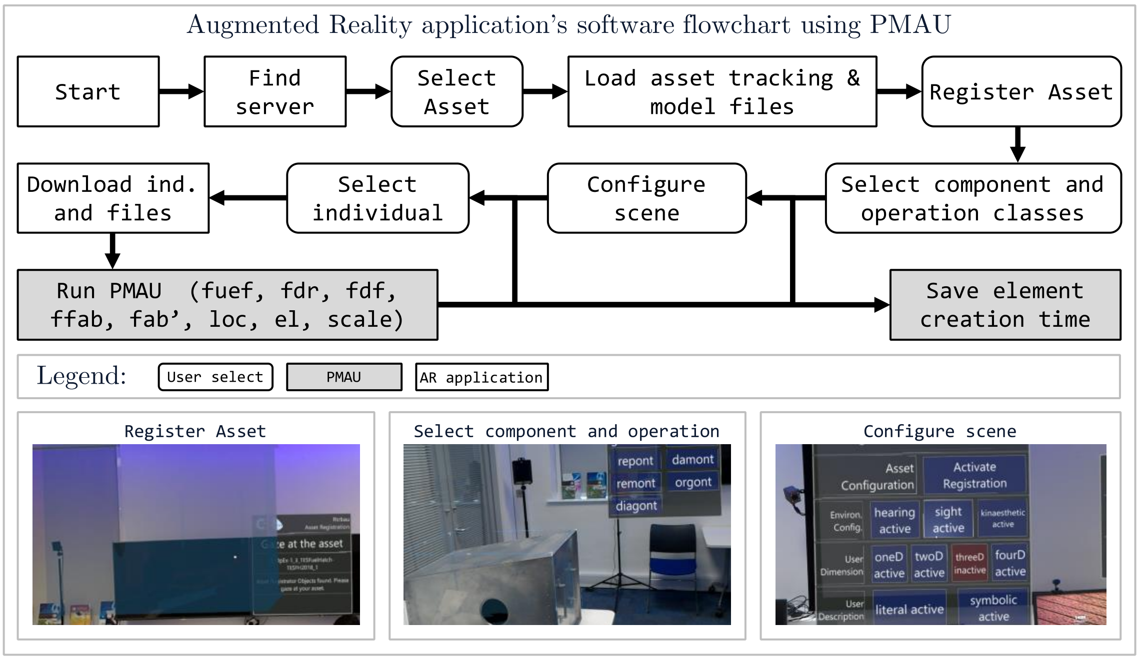


Figure 4-5. Description of the algorithm's real-time behaviour within a resultant AR application – It displays the working behaviour of an AR application using PMAU. The figure includes PMAU steps (grey), user selections (rounded white) and additional software capabilities (squared white) necessary for the PMAU algorithm to work as an AR application.

4.5 Validation protocol

4.5.1 System implementation

The proposed solutions were implemented within a prototype system for experimentation. This prototype consists of two subsystems: (1) a cloud server for maintenance ontologies storage and (2) a HoloLens-based AR application. Figure 4-6 presents the languages and platforms utilised to code each subsystem. The cloud server storage (Subsection 3.5.1) uses the graphical database Neo4j (Zhu, Zhou and Shao, 2019) to store maintenance ontologies, and Cypher (Panzarino, 2014) and neosemantics (Barrasa, 2019) to support data transfer through OWL and RDFS, respectively. Besides, the server incorporates a web-based application coded in EJS (Eernisse, 2015) for maintenance experts to input maintenance data, which has already been described in Chapter 3. And also a service provider built in NodeJS (Surhone, Tennoe and Hensonow, 2010) to transfer ontology data (e.g. classes, individuals, etc.) and related files (e.g. obj, png, etc.) using HTTP requests and JSON objects to the AR application. The HoloLens-based AR application has been coded and deployed using Unity Game Engine (Unity Technologies, 2019) and Visual Studio (Microsoft Corporation, 2019). The programmable content and pattern-matching algorithm have been coded using C# (Hejlsberg, 2011). The interaction through HoloLens has been enabled with MixedRealityToolkit (Microsoft Corporation, 2020). Besides, Vuforia (PTC Corporation, 2020) has been used to enable registration and tracking in the AR application and coded to use the JSON-based API to transfer from the cloud server necessary ontology-related files like Vuforia's model targets.

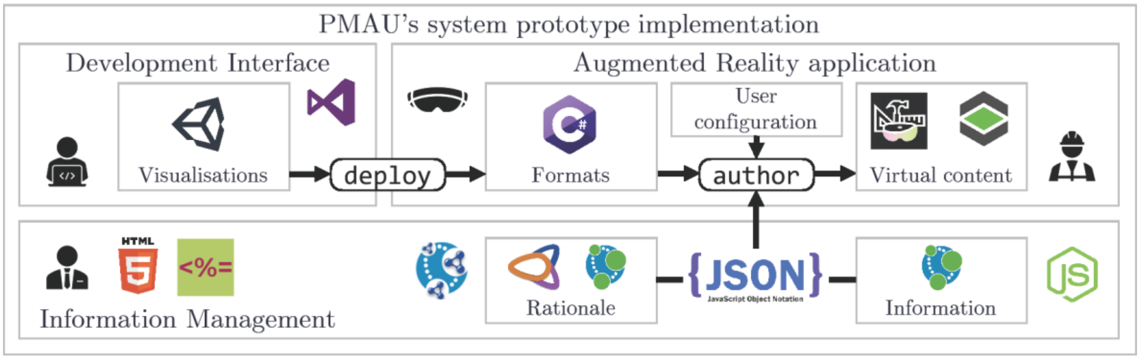


Figure 4-6. Description of the automatic authoring proposal's implementation as software system – It replicates the structure of Figure 4-1-b to present the logos of each tool/language utilised to develop the system's prototype.

4.5.2 Experiment design

This chapter proposes a real-time, ontology-based, pattern-matching technique for automatic adaptive authoring in multiple maintenance operations. The previous sections explained the methods utilised to separate authoring's content creation and information management processes and to automate the former. Hence, this research validation should aim to evaluate produced content adaptiveness to multiple maintenance operations.

In academia, usual approaches to evaluate content adaptiveness to a maintenance operation are comparisons of efficiency (time and errors) (Flotyński and Walczak, 2015; Longo, Nicoletti and Padovano, 2019) and usability (Gimeno *et al.*, 2013; X Wang, Ong and Nee, 2016a) effects of different AR and non-AR solutions. In this research, it is also necessary to evaluate such effects on multiple maintenance operations. For this reason, validation methods should compare this research proposal against alternative authoring and non-AR solutions in different maintenance operations.

The student identified two already-published, alternative authoring solutions available for experimentation. One is called ARAUM (Erkoyuncu *et al.*, 2017) and focuses on off-line context-aware authoring for repair operations. The other one, called SMAARRC (Fernández del Amo *et al.*, 2019), describes rule-based authoring for remote diagnosis. These solutions are later presented in Subsection 4.5.3 and further details can be found in Appendix D (ARAUM) and Appendix E (SMAARRC). In order to validate the proposed authoring method (PMAU) against these two, this research considers the following hypotheses:

- Completion errors do not vary significantly among authoring and no-AR solutions for each maintenance operation.
- Completion time decreases with authoring solutions compared to no-AR solutions for each maintenance operation.
- Completion time does not vary significantly among authoring solutions for each maintenance operation.
- Content usability does not vary significantly among authoring solutions for each maintenance operation.

For the abovementioned measures to be appropriate for evaluating these hypotheses, the following assumptions must hold true:

- Time and errors can be a direct representation of efficiency if a consistent quality is assumed at the experimented maintenance operations. In order to ensure so, this validation assumes pre-determined operations whose quality does not depend on the tester's performance.
- Usability of augmented content can affect maintenance efficiency if content is not compatible with maintenance environment or manual operations. Hence,

it is necessary to evaluate testers' perceived usability to evaluate maintenance operations' quality.

The student employed two different research methods to evaluate these hypotheses' validity according to the quantitative and qualitative measures described above. These are stopwatch time and errors studies and usability surveys, and they are described in the following subsections.

4.5.2.1 Stopwatch time and errors studies

Stopwatch time and errors studies aim to analyse the proposed authoring solution (PMAU) effect over maintenance efficiency on different operations compared to alternative solutions (ARAUM, SMAARRC, NOAR). It is assumed that AR-improved semantic understanding of real-world objects increases efficiency of maintenance tasks (Azuma, 2016). In such scenarios, it can be said that efficiency solely depends on time for similar levels of effectiveness (quality).

Time can be described by the number of seconds required by a tester to find, understand and complete a maintenance task. Quality, also understood as errors, can be defined as the number of tasks completed by a tester that deviate in form or result of what was pre-determined. Besides, semantic understanding is assumed to affect efficiency through the authoring solution utilised and the step of a maintenance operation being experimented.

Based on previous definitions, it can be said that if errors (quality) are invariable, then the effect of authoring solutions through semantic understanding over maintenance efficiency can be evaluated based on its effect on completion time. Such evaluation should be made over different maintenance operations to demonstrate the validity of this research contributions. If the assumptions above are correct, then it is reasonable to expect the following results:

- Errors do not vary with the use of different solutions for each maintenance operation.
- Time is reduced with the use of authoring solutions compared to non-AR solutions for each maintenance operation.

- Time does not vary significantly between authoring solutions for the same maintenance operation.

The study described above considers one response variables (time and errors), one control variables to test assumptions (step), and two independent factor variables (solution and operation). Table 4-2 defines these variables. Besides, each factor variable can have different levels, which are defined in Table 4-3.

Table 4-2. Description of response, control and factor variables for stopwatch studies.

Variable	Type	Definition
Time	Response	Time taken by a tester to identify, understand and complete a maintenance task
Errors	Response	Tasks completed with form or result deviations from its pre-defined target
Step	Control	Specific assignment to be undertaken by a tester as part of a maintenance operation
Solution	Factor	Authoring solution employed to generate augmented content support to conduct maintenance tasks
Operation	Factor	Nature of tasks being conducted which belong to a specific step in the maintenance process

These efficiency experiments are slightly different to the ones proposed in Chapter 3 (Subsection 3.5.2.5). Although they share similar response (time and errors) and factor (solution and operation) variables, these experiments do not consider IT expertise as a relevant factor. The student decided not the include this factor for two reasons. The first one relates to the nature of operations. Since the case study operations determine for these experiments are mainly mechanical, expertise with IT systems (electric and electronic) seemed to be not very relevant. The second one relates to the scope of the solutions. Although testers' expertise can have an impact on these experiments, the experimental case studies were designed for testers without previous maintenance experience. The reason for doing so was that alternative authoring solutions (ARAUM and SMAARRC) were prepared to produce content for non-expert users. Therefore, if the proposed authoring solution (PMAU) was to be compared against those two, it could only be done using non-expert testers to obtain a fair comparison.

Table 4-3. Description of factor levels for stopwatch studies.

Factor	Level	Description
Solution	PMAU	Use of this research proposal to generate AR support
	ARAUM	Use of an ad-hoc authoring solution for repair operations
	SMAARRC	Use of an ad-hoc authoring solution for remote diagnosis operations
	NOAR	Use of paper-based solutions to support maintenance operations
Operation	Repair	Set of tasks aiming to return equipment to its working conditions
	Diagnosis	Set of tasks aiming to identify the cause of an equipment's failure

These experiments aim to test the proposed authoring solution (PMAU) against other ad-hoc (ARAUM and SMAARRC) and paper-based (NOAR) authoring solutions in two different maintenance operations. In order to simplify the evaluation process, the tasks experimented at the ad-hoc authoring solutions researches (Erkoyuncu *et al.*, 2017; Fernández del Amo *et al.*, 2019) will be re-utilised for these experiments. These case studies comprising AR solutions, maintenance operations and associated equipment are presented in Subsection 4.5.3, Appendix D and Appendix E. It is worthy to note that the solution of one case study cannot be used for the operation of the other one and vice versa.

Each experiment consisted of a tester conducting the steps from one operation with an AR application of one authoring solution. Thirty testers (Subsection 4.5.3.3) were required to complete the two experimental operations. Results data from previous researches regarding the NOAR solution will be re-used as baseline comparators. Testers were randomly allocated to one of four groups according to the abovementioned procedure and factors. Table 4-4 presents these groups according to their operation and solution. Besides, experimental operations (Subsection 4.5.3) and solutions can be considered sufficiently different to not expect carry-over effects among experiments. So, the experimental design can be considered between-subjects for further statistical analysis.

Table 4-4. Description of experimental groups according to factors' levels.

	PMAU	ARAUM	SMAARRC	NOAR
Repair	A	B		C
Diagnosis	B		A	D

4.5.2.2 Usability surveys

After conducting efficiency experiments, testers were asked to complete a survey on the usability of the AR solutions employed. Usability surveys aims to evaluate the perceived validity of the proposed authoring solution to enhance semantic understanding compared to alternative authoring methods. Usability refers to the ability of the authoring solution's resultant augmented content to deliver information appropriately to the user regarding the maintenance operation to be conducted. Besides, it is a feature perceived by users and so subject to opinion. Therefore, it is necessary to use qualitative criteria for its evaluation. Unlike for previous usability surveys presented in this thesis (Subsection 3.5.2.6), testers spent reasonable time trying different AR solutions. Therefore, it seemed relevant to study AR solutions' aspects with an impact in usability more in detail (Nielsen, 1993). In order to enhance this survey's replicability, the student utilised usability criterions and AR solutions' aspects considered by previous researches (Chi *et al.*, 2012; Gimeno *et al.*, 2013; Lee and Lee, 2016; X Wang, Ong and Nee, 2016a; Dey *et al.*, 2018). Table 4-5 defines the usability criterions proposed for this survey, and Table 4-6 presents the AR solutions' aspects considered for each criterion.

Table 4-5. Description of usability criterions employed to evaluate AR solutions from authoring methods based on those by Nielsen (1993) and Lee and Lee (2016).

Criterion	Definition
Ease-to-learn	Ability of the AR solution to show its functionality by itself
Ease-to-use	Ability of the AR solution to be self-understandable
Accuracy	Ability of the AR solution to display augmented content correctly
Effectiveness	Ability of the AR solution to support a maintenance operation
Satisfaction	Tester's overall impression of the AR solution after using it

Table 4-6. Description of AR solutions' aspects considered for each usability criterion based on those presented by Gimeno *et al.* (2013) and Dey *et al.* (2018).

Criterion	Aspect	Scale
Ease-to-learn	Start, Finish, Intuitiveness	Likert 1-5
Ease-to-use	Gestures, Text, Buttons, Images, Models, Holograms, Animations	Likert 1-5
Accuracy	Overlay, Shaking, Occlusion, Visualisation, Latency	Likert 1-5
Effectiveness	Efficiency, Confidence	Likert 1-5
Satisfaction	Design, Feeling, Overall	Likert 1-5

Usability surveys consisted of separate section for each criterion including several statements for each aspect regarding the AR solutions tested in experiments. The survey's questionnaire can be found in Appendix F. Testers were asked to declare their agreement with questionnaire's statements in a Likert Scale from 1 to 5. The reason to select such scale is based on the results presented by Weijters, Cabooter and Schillewaert (2010), who suggested that it maximises potential information transmission when surveying non-expert populations (Subsection 4.5.3.3). The results collected serve to evaluate the authoring solution's usability compared to other authoring approaches. Besides, operational quality is also evaluated in terms of efficiency and confidence improvements. There are some assumptions to consider regarding these surveys:

- Errors are not evaluated in qualitative terms as they may be dependent on user expertise, which can vary for potential users of this solution.
- It is assumed that the quality is of consistent level for the stopwatch time studies if the results of the questionnaire provide a similar result to the experiments.

The protocol to collect and analyse experimental and survey data is described in Subsection 4.5.4. The following section presents the experimental cases of study along and testing sample.

4.5.3 Cases of study

The cases of study comprise two maintenance operations (repair and remote diagnosis) to be experimented in two complex-engineering assets. These cases of study were already presented and discussed in the two publications (Erkoyuncu *et al.*, 2017; Fernández del Amo *et al.*, 2019) regarding the experimental alternative authoring solutions. Further details can be found in Appendix D and Appendix E, respectively for each case study. In order to accommodate these case studies to ontology-based information systems, the mapping procedure from Cullot, et al. (Cullot, Ghawi and Yétongnon, 2007) was used. Figure 4-7 presents an overview of both cases of study, including equipment, resulted ontologies for PMAU application and views of alternative authoring solutions. The resultant ontologies produced to replicate the databases from previous researches can be consulted at 10.17862/cranfield.rd.12213380.

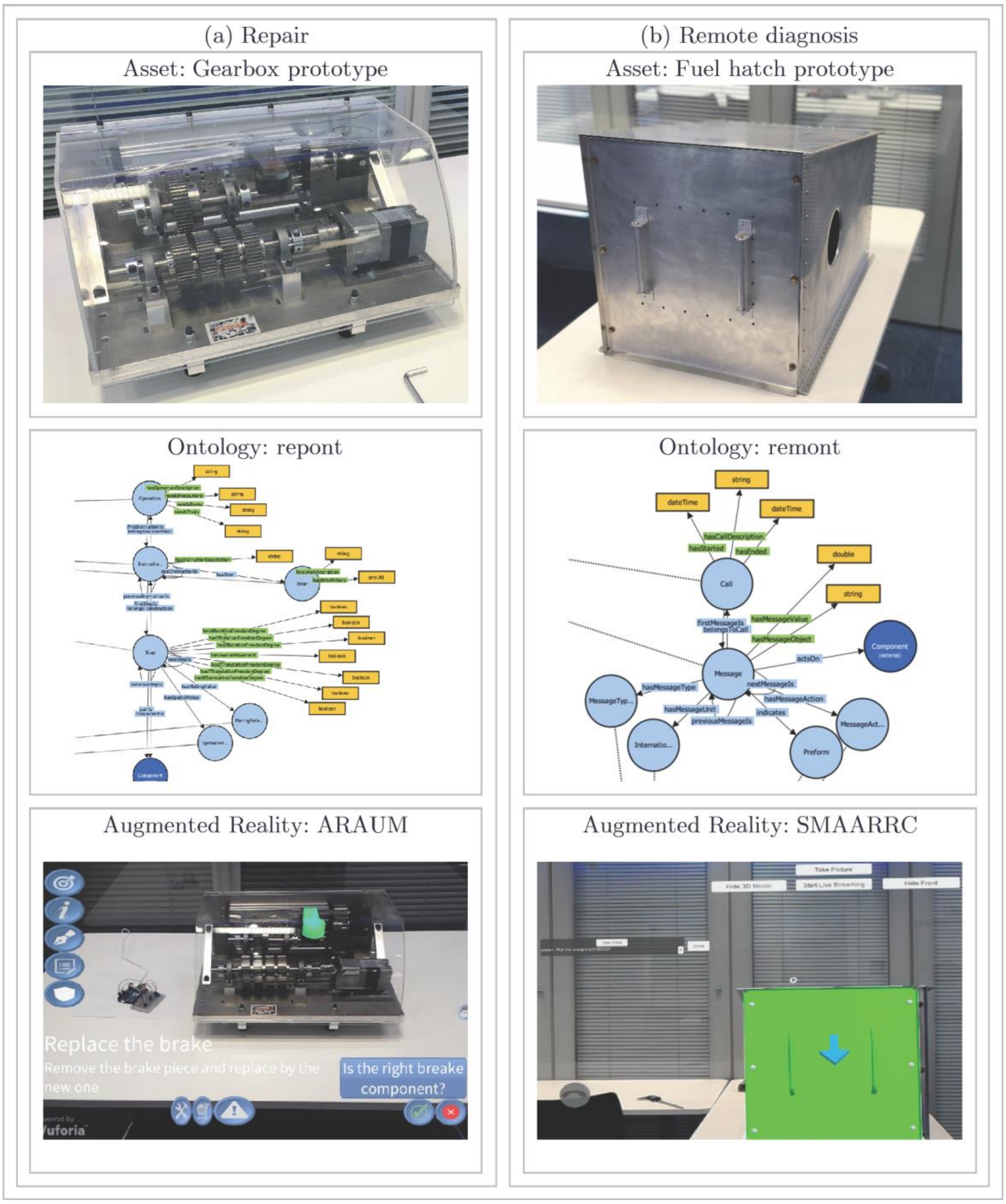


Figure 4-7. Description of repair and remote diagnosis case studies – The figure includes a view of the case study's asset (left-side), the resultant ontology proposed for the maintenance operation (centre) and screenshots of AR applications from case studies' authoring solutions (right-side).

4.5.3.1 Maintenance repair

The first case study operation is the same one described by Erkoyuncu, et al. (2017). Further details can be found in Appendix D. The operation represents a repair task in complex engineering assets for the Defence Industry that focuses in mechanical and electric systems and assembly and replacement procedures. The case study AR solution aims to develop effective guidance-support tools to enhance repair tasks in complex scenarios. The case-study equipment is a gearbox prototype utilised for studying gear-wheels degradation that represent real-life conditions of asset-repair scenarios. The experiments described in (Erkoyuncu *et al.*, 2017) focus on a specific repair operation composed of several assembly, disassembly and replacement steps involving mechanical components. The case study experiment, described in Appendix D, aimed to analyse the effect of an ad-hoc tablet-based authoring solution, called ARAUM, that aims to simplify the generation of augmented animations. The experimental repair scenario conducts an operation to replace a gearbox's component (brake wheel) that has been worn away. Figure 4-8 presents the case study's experiment setup. Figure 4-9 describes it operation's steps using PMAU content. Figure D-3 presents an example of ARAUM content for one of these operation's steps.

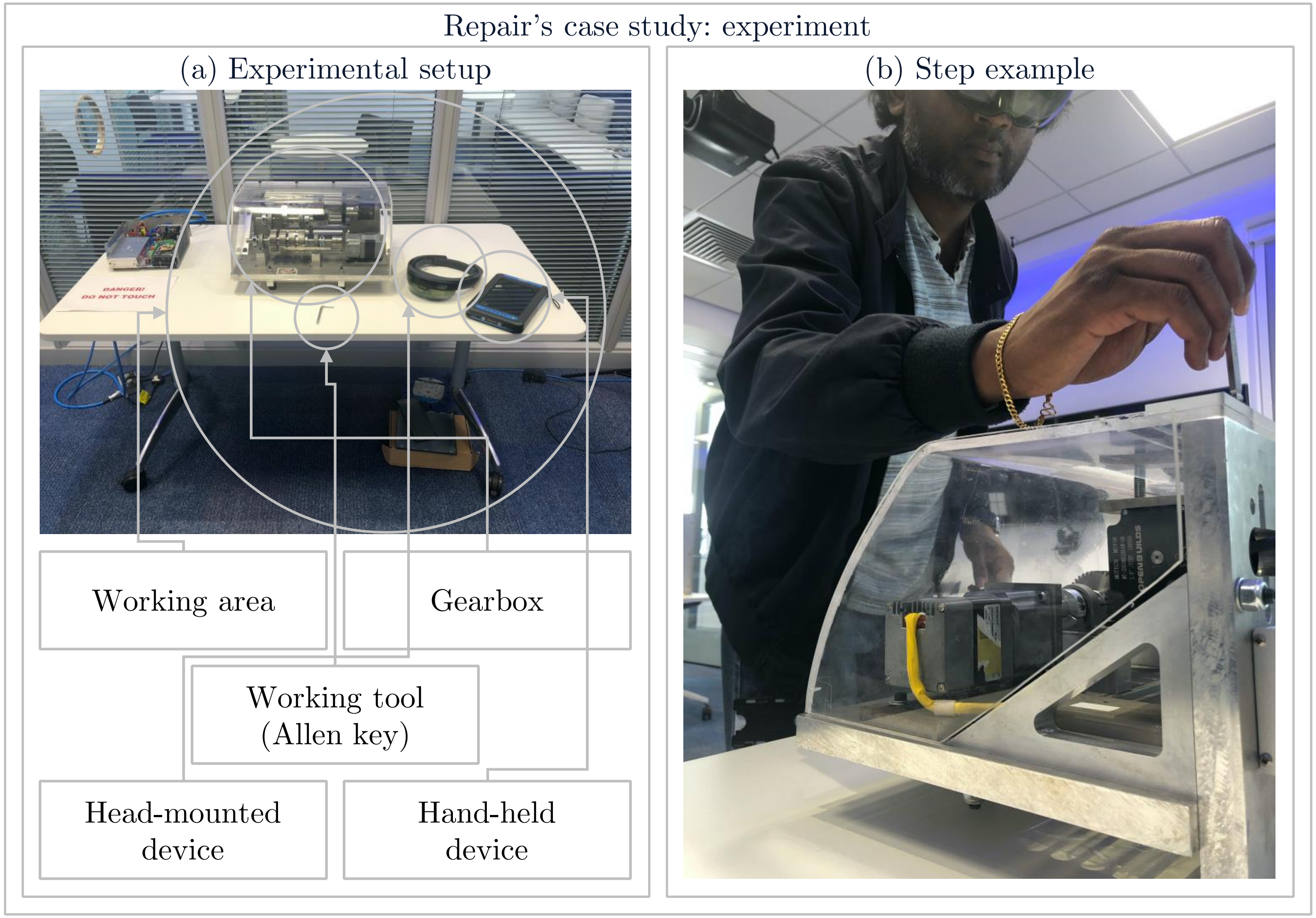


Figure 4-8. Description of repair case study experiment: (a) experimental setup and (b) experimental step example.

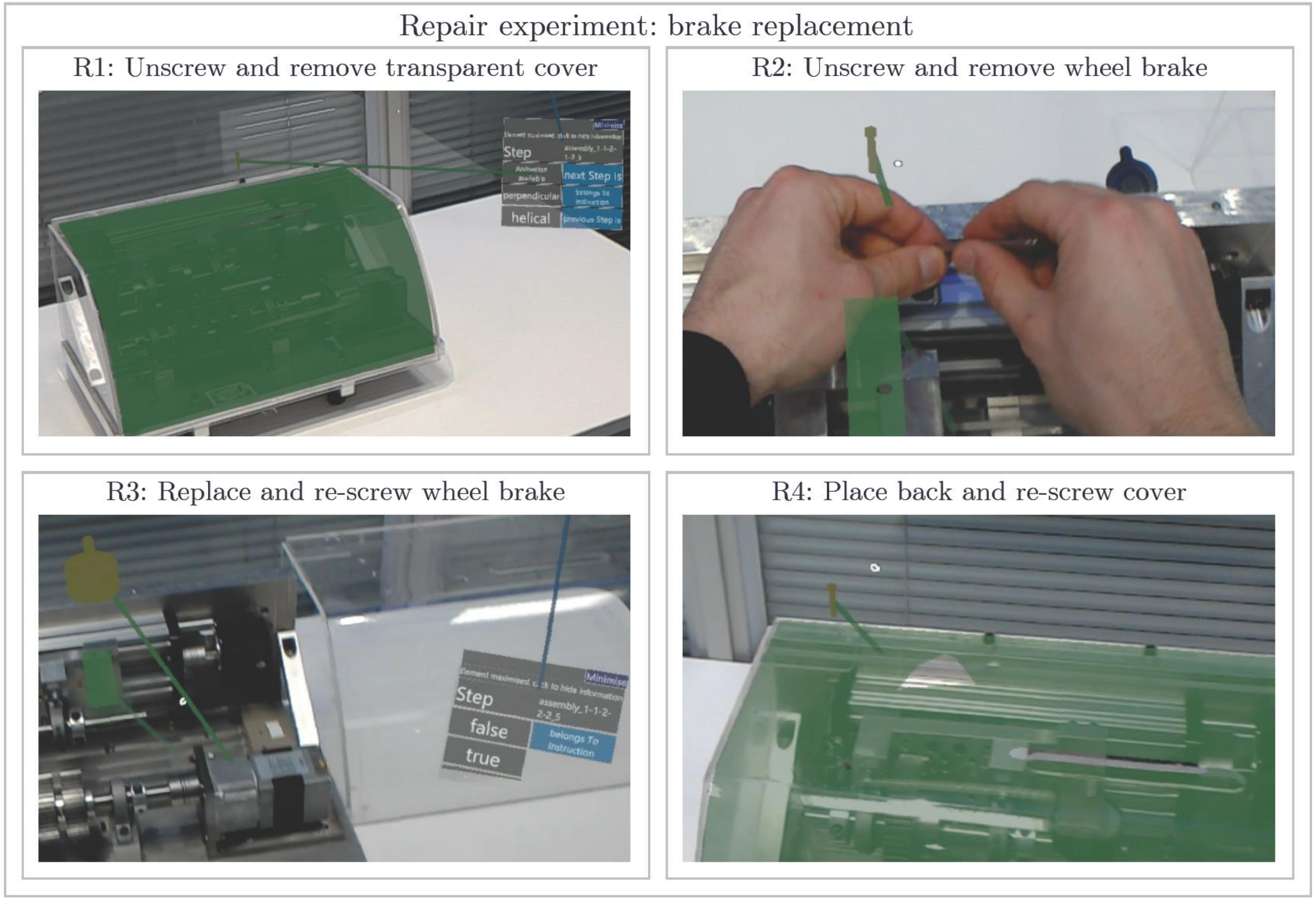


Figure 4-9. Description of repair experimental operation using PMAU content – Further details on experimental steps (R1-R4) can be found in Appendix D.

4.5.3.2 Remote diagnosis

The second case study operation is the same one presented by Fernández del Amo, et al. (2019). Further details can be found in Appendix E. The operation represents a remote diagnosis task for complex engineering assets in the Aerospace Industry and its focus is purely in mechanical systems. The case study AR solution aims to develop effective communication-support tools for enhancing remote diagnosis in 'decision-to-fly' scenarios. The case-study equipment is an aircraft's fuel hatch prototype with unidentified imperfections that are the diagnosis target. The case study experiment, described in Appendix E, focuses on a diagnosis operation that comprises inspection, measurement and repair of mechanical components. These experiments aimed to analyse the effect of the ad-hoc HoloLens-based authoring solution, called SMAARRC, that aims to simplify the understanding of complex messages. The experimental diagnosis scenario conducts an operation to identify several defects that the fuel hatch has and resolve them if necessary. presents the case study's experiment setup. Figure 4-11 describes this remote diagnosis operation's steps using PMAU content. Figure E-2 presents an example of SMAARRC content for one of these operation's steps.

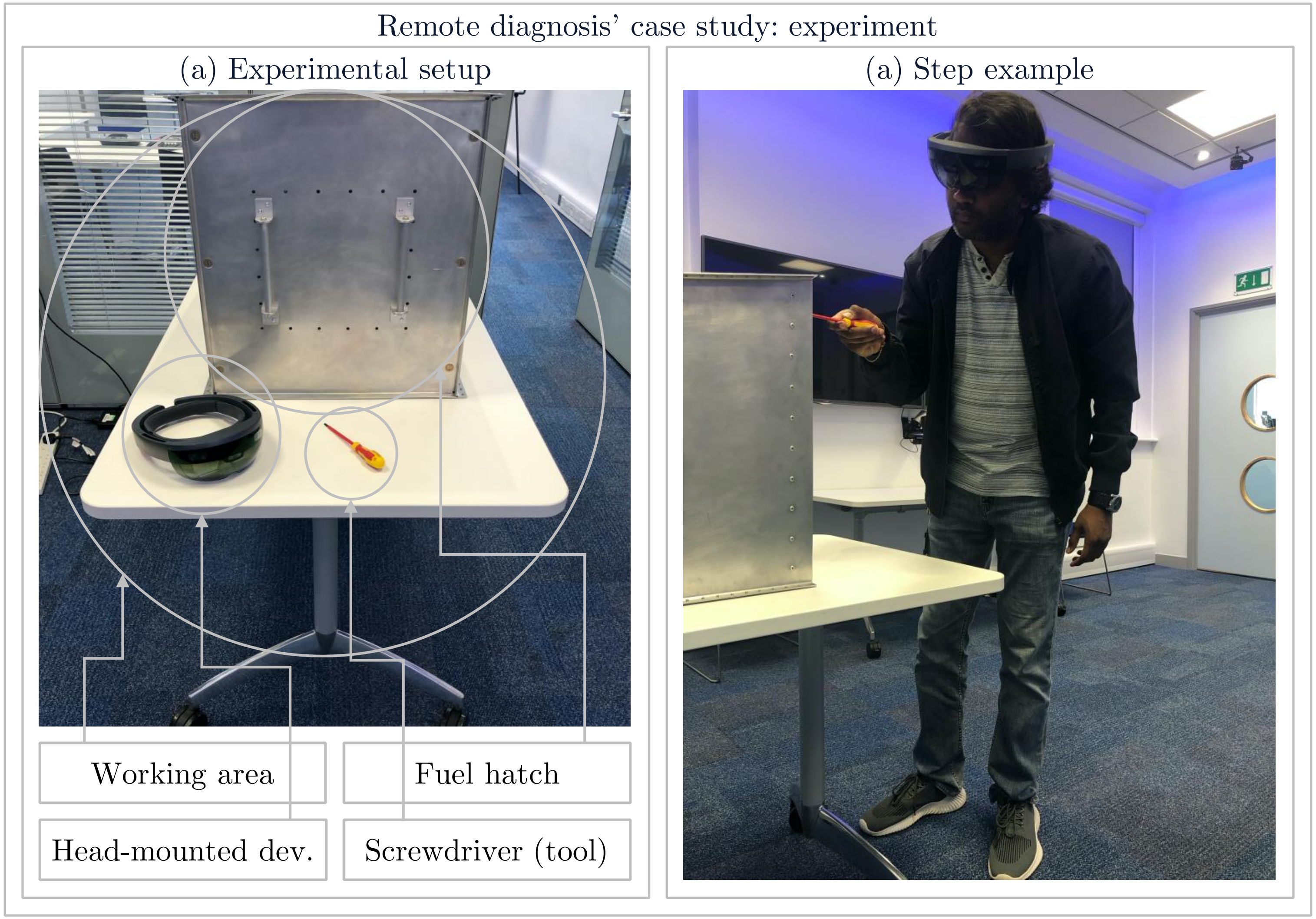


Figure 4-10. Description of remote diagnosis case study experiment: (a) experimental setup and (b) experimental step example.

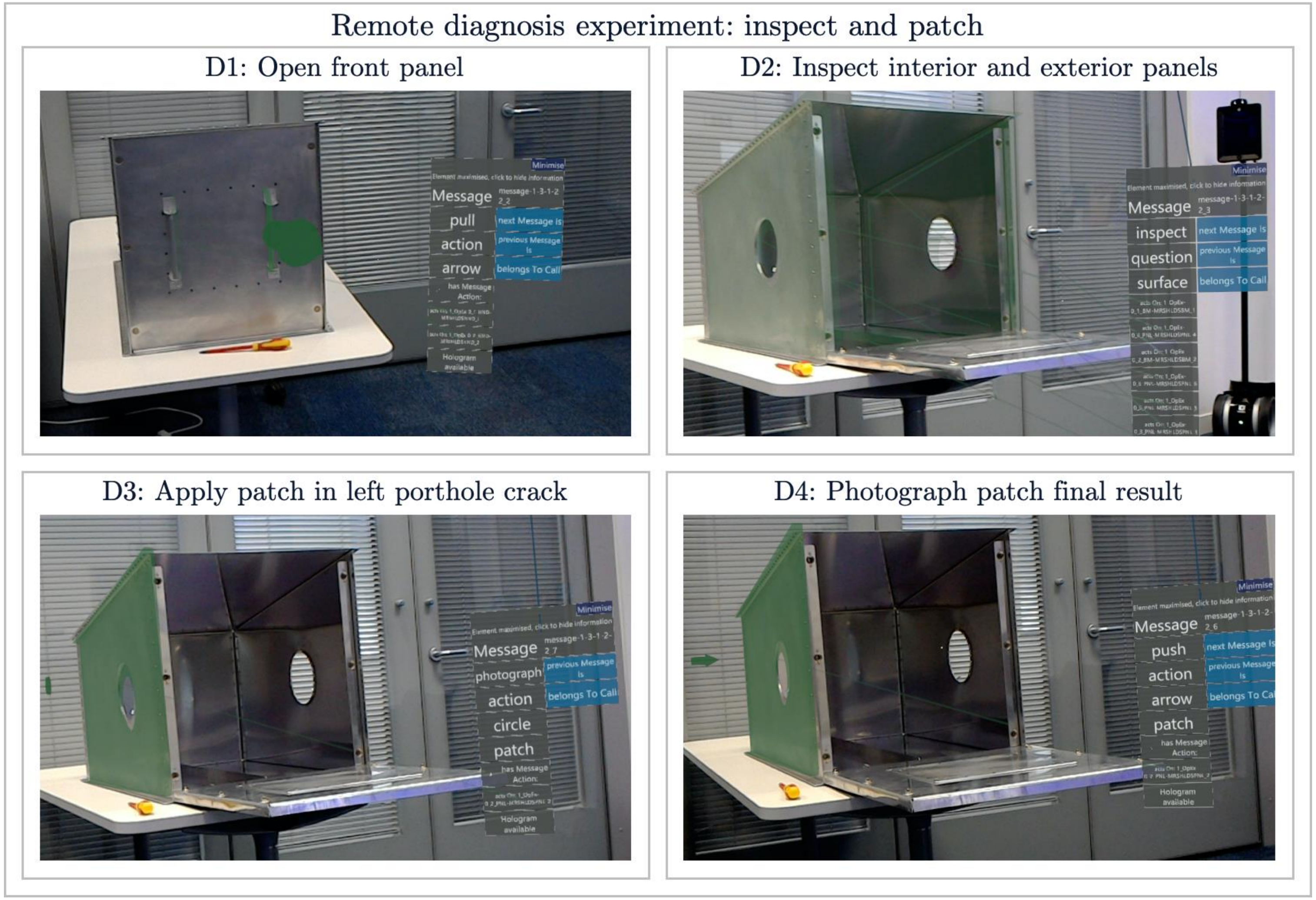


Figure 4-11. Description of remote diagnosis case study experimental operation using PMAU content - The complete experimental steps (D1-D4) are described in Appendix E.

4.5.3.3 Experimental population sample

A total of 30 MSc students (24 males and 6 females) participated as testers in laboratory experiments. Their ages range from 22 to 29 years and they are all enrolled in engineering-related MSc degrees. Although they have some basic knowledge in AR and maintenance due to their courses, they have no previous hands-on experience in any of them. This helped to ensure the validity of the experimental case studies, which were designed for testers without previous maintenance experience. Therefore, they were given a short training on AR devices right before experimentation to avoid the presence of any learning curves. Testers were randomly allocated to one of the two groups (A (15) or B (15) – Table 4-4) to avoid "carry-over" effects between operations while using two different authoring solutions. Besides, the results from previous researches (Erkoyuncu *et al.*, 2017; Fernández del Amo *et al.*, 2019) were re-used for NOAR solution's groups (C and D).

4.5.4 Experimental protocol

The protocol comprises the steps to collect and analyse experimental and survey data for validating this research proposal against its expected contributions. It implements this validation methods in the case study contexts described above. The following list summarises this protocol:

1. Data collection (30 testers per experiment):

a. AR-maintenance introduction: to briefly describe testers the purpose of experiments as well as the use of AR solutions in maintenance operations.

b. Stopwatch time and errors experiments: to capture quantitative data on the effect on efficiency of different authoring solutions for diverse maintenance operations.

c. Usability surveys: to capture qualitative data on tester's opinions regarding usability of the authoring solution proposed compared to other alternatives used within experiments.

2. Data analysis (45 testers per experiment):

a. Errors effect study: to ensure the validity on the assumption that quality is kept among experiments. Results should reflect that there are no significant differences on the errors made by testers using different solutions in maintenance operations. Basic statistics, one-way ANOVA between-subjects tests and graphical analysis will be used for this matter.

b. Time effect study: to analyse the correlation between the response variable (time) and considered factors (solution and operation). Results should reflect that the proposed authoring solution (PMAU) does not present significant differences on time compared to alternative authoring solutions (ARAUM and SMAARRC) in different maintenance operations. They should also reflect that these are significantly different to NOAR solutions. Experiments are set independently for each maintenance operation, and so the factors to consider in the analyses (Step and Solution). Due to the number of factors (2 - step and solution), a two-way ANOVA between-subjects analysis will be used to tests these hypotheses for each experiment. Moreover, additional post hoc (Tukey HSD) test comparisons will be used to evaluate interactions between factors' levels.

c. Usability study: to quantitatively evaluate testers' (30) opinions on the proposal's content usability. Results should reflect that usability does not compromise the effectiveness of the supported maintenance operation. Due to the quantitative nature of these results, basic statistics and graphical analysis will be used for this matter.

This experimental protocol aims to validate this research's proposal against its expected contributions. For this validation to be coherent, there are few assumptions to consider:

- In order to keep consistency with previous researches (Erkoyuncu *et al.*, 2017; Fernández del Amo *et al.*, 2019) the experiments were conducted in a laboratory environment in order to keep constant other factors (e.g. ergonomics or lighting conditions) that may affect the results. This enabled to reutilise results from previous research regarding the testing of NOAR solutions for the case study operations.

- Additional effects studied in previous researches are not considered in this protocol. The aim is to prove that the new authoring method achieves similar times to alternatives, so the contributions achieved with those should also be applicable to this new authoring method.

- Experimental sample size for the abovementioned statistical tests can be estimated "a priori". Such estimation can be done using a F test for the most requiring analytical test (two-way ANOVA between-subjects). With 12 factor groups (solution and step factor levels), a variance of 0.25 (partial eta squared), a type-I error of 0.1 (alpha) and a power of 0.9 (1 – beta), the resultant sample size is 51 people. That is quite close to the 45-sample size achieved: 30 testers from this research experiments and additional 15 testers results obtained from previous researches (Erkoyuncu *et al.*, 2017; Fernández del Amo *et al.*, 2019). Besides, these numbers are bigger compared to similar researches that achieved sample sizes of 30 testers (Gimeno *et al.*, 2013; X Wang, Ong and Nee, 2016a; Longo, Nicoletti and Padovano, 2019).

- As described above, testers are MSc students with none or very little experience in AR or maintenance. Although this ensures a baseline for measuring maintenance efficiency, further experiments should be required to corroborate laboratory results in real-life working conditions.

This protocol's results are discussed in the following section.

4.6 Results

The research validation aims to corroborate the hypotheses listed in Subsection 4.5.2 using the experimental protocol described in Subsection 4.5.4. Its results are analysed and discussed in the following subsections to evaluate this research hypothesis validity. The complete results datasets and analysis can be consulted at 10.17862/cranfield.rd.12213380.

4.6.1 Errors effect study

The aim of the errors effect study was to validate the following hypothesis: *"errors do not vary significantly among authoring and no-AR solutions for each maintenance operation"* (Subsection 4.5.2). Stopwatch experiments consisted of testers completing two maintenance operations: repair and remote diagnosis. Errors are defined as the number of tasks within operations' steps completed with form or result deviations from their pre-defined targets. Testers made use of AR (PMAU, ARAUM and SMAARRC) and NOAR solutions to support their selves with augmented information while completing operations' steps. So, if the information utilised was the same but in different content formats, then errors should not differ among solutions for each operation.

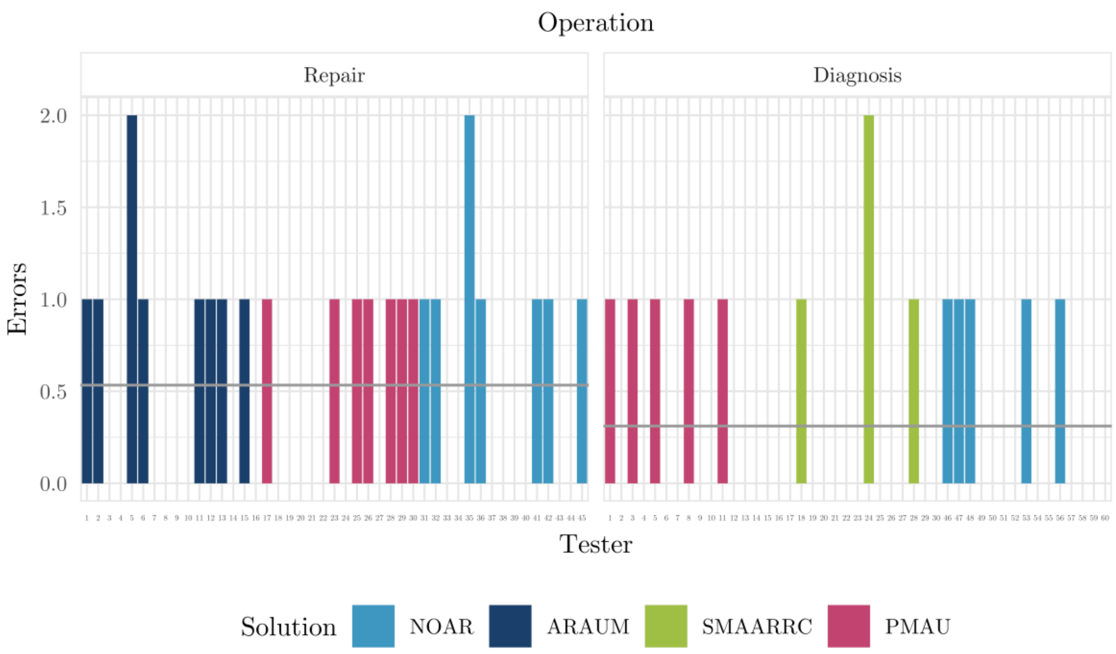


Figure 4-12. Number of experimental operation errors per tester according to case study operation and solution employed, with the mean per operation as a grey line.

Figure 4-12 and Table 4-7 present errors results per tester and on average grouped by operation and solution factors. A conservative estimate on errors for novice testers can be taken at 2 errors per experiment (50% error rate per experimental step). So, total number of errors can be considered low as only 3% of testers make 2 errors. On average, average errors grouped per operation and solution range from 0.267 to 0.6. In repair, average errors with PMAU (0.467) are the lowest, while ARAUM (0.6) is higher than NOAR (0.533). In diagnosis, PMAU and NOAR are equal (0.333), while SMAARRC (0.267) are the lowest.

Table 4-7. Mean and std. deviations on completion errors per operation and solution factors.

Operation	Solution	Tester	Mean	Std. deviation
Repair	PMAU	15	0.467	0.516
	ARAUM	15	0.600	0.632
	NOAR	15	0.533	0.639
Diagnosis	PMAU	15	0.333	0.488
	SMAARRC	15	0.267	0.594
	NOAR	15	0.333	0.488

Further analyses (Table 4-8) can identify significance of factors on errors results. One-way ANOVA between-subjects tests made on errors over solutions for each operation indicate that the solution factor is not significant ($p\text{-value} < 0.05$), with a p-value of 0.831 in repair and 0.923 in diagnosis. Besides, t-test results on errors over operations for all solutions suggest that the operation factor is close to be significant with a p-value of 0.059.

Table 4-8. Statistical tests on errors results per solution and operation factors.

Factor: Solution:Repair		Test: One-Way ANOVA between-subjects				
Effect	Df	Sum Sq	Mean Sq	F Value	Pr (>F)	Sig (95% ci)
Solution	2	00.133	0.067	0.186	0.831	No
Residuals	42	15.067	0.359			
Factor: Solution:Diagnosis		Test: One-Way ANOVA between-subjects				
Effect	Df	Sum Sq	Mean Sq	F Value	Pr (>F)	Sig (95% ci)
Solution	2	00.044	0.022	0.080	0.923	No
Residuals	42	11.600	0.276			
Factor: Operation		Test: t-test between-subjects				
			T	Df	p-value	Sig (95% ci)
			1.908	86.483	0.059	No

According to previous discussions, the following considerations can be considered valid:

- Number of errors per tester can be considered low with an average of 0.422 errors per test.
- There is significant variance on errors results per operation (p-value of 0.059).

Therefore, the validation's errors hypothesis can be considered true and so, task completion time can be understood as a direct measure of efficiency. The following subsection analyses the results on experimental completion times.

4.6.2 Time effect study

The aim of the time effect study was to validate the two following hypotheses:

- *"Time decreases with AR solutions compared to no-AR solutions for each maintenance operation".*
- *"Time does not vary significantly among AR solutions for each maintenance operation".*

Time is a response variable considered in the stopwatch experiments (Subsection 4.5.2.1). It measures the number of seconds taken by a tester to find, understand and complete an experimental operation step (Figure 4-9 and Figure 4-11). Because AR (PMAU, ARAUM and SMAARRC) and NOAR solutions did not show a significant effect on errors, time can be considered a direct representation of maintenance efficiency. Hence, time can be evaluated as the main effect of AR content support on maintenance operations through semantic understanding.

Figure 4-13 presents average time results per step and grouped by operation and solution factors. It displays a difference in completion times per step for each experimental operation. Besides, it shows a clear difference between NOAR and AR (ARAUM, SMAARRC and PMAU) solutions, but not among AR solutions. A relevant case is D1, which indicates that the effect of AR solutions is not significant. This case is similar to the findings presented by Fernández del Amo, *et al.* (2019), where the kind of step had an effect on AR impact.

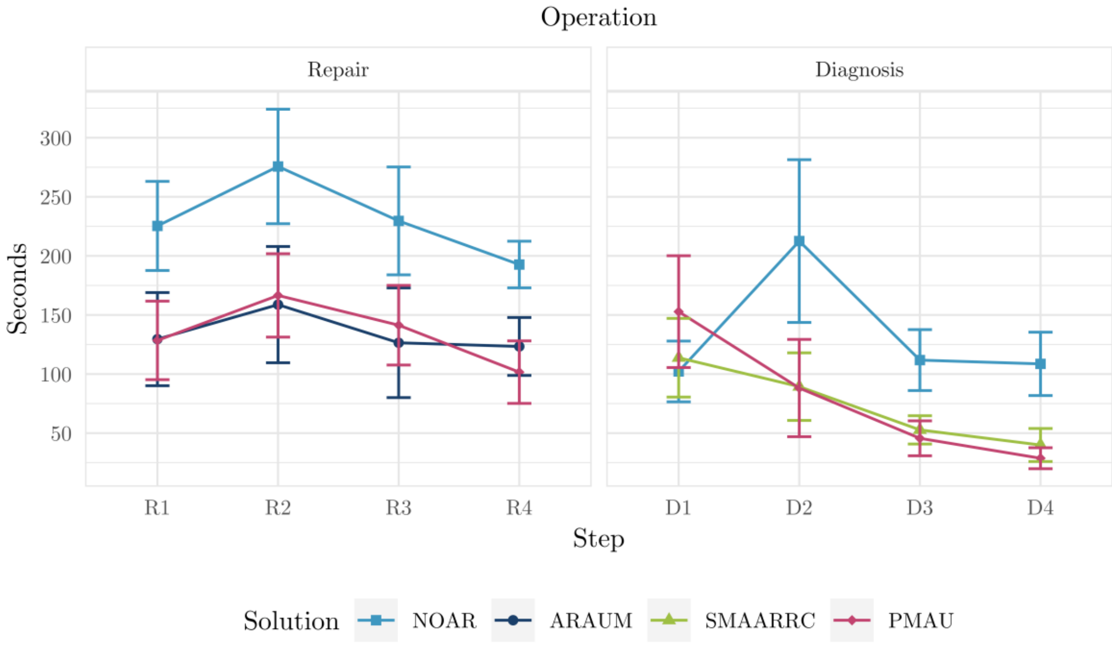


Figure 4-13. Average experimental operation's step completion time in seconds for NOAR, ARAUM, SMAARRC and PMAU solutions per case study operation, with the standard deviation represented by error bars.

Table 4-9 presents means and std. deviations for completion times grouped by solution and operation. These range from 134 to 231 seconds in repair and from 74 to 134 seconds in diagnosis. These results show a difference between repair and diagnosis operations, indicating that the assumption for separate experiment analyses was valid. For repair, means show a considerable difference in completion times among NOAR and AR (PMAU and ARAUM) solutions. For diagnosis, means also show a substantial difference in completion times between NOAR and AR (PMAU and SMAARRC), although there is also a smaller difference between SMAARRC and PMAU.

Further analyses can determine the significance of comparisons discussed above. Table 4-10 and Table 4-11 present the two-way ANOVA between-subjects tests conducted to analyse time variance according to step and solution factors for each operation. According to repair results (Table 4-10), it can be said with a confidence interval of 95% (p-value < 0.05) that both factors (Step and Solution) have a significant effect on completion times, but not their interaction. Hence, it can be said that for repair operations, the support AR provides does not depend on the type of step being conducted. For diagnosis experiments (Table 4-11), ANOVA results also indicate a significant effect of step and solution factors as well as their interaction. These confirm the results presented in (Fernández del Amo *et al.*, 2019), where AR support was found more effective for higher complexities of steps being conducted.

Table 4-9. Means and std. deviations on completion time per operation and solution factors.

Operation	Solution	Testers	Mean	Std. deviation
Repair	PMAU	15	134.48	39.37
	ARAUM	15	134.52	42.43
	NOAR	15	230.82	48.69
Diagnosis	PMAU	15	78.82	57.76
	SMAARRC	15	73.95	37.55
	NOAR	15	133.78	61.03

Table 4-10. Two-way ANOVA between-subjects test results on completion time for step and solution factors in Repair operation.

Operation: Repair						
Factor	Df	Sum Sq	Mean	F value	Pr (>F)	Significant (95% ci)
Step	3	086309	028770	020.12	3.42e-11	Yes
Solution	2	371076	185538	129.79	2.00e-16	Yes
Step:Solution	6	011061	001843	001.29	2.65e-01	No
Residuals	168	240168	001430	-----	-----	-----

ANOVA tests results help to corroborate the second hypothesis presented in Subsection 4.5.2. Based on these, it can be said that task completion times are dependent on the solution being used. Moreover, completion times group means (Table 4-9) indicate that these times decrease with the use of authoring (PMAU, ARAUM and SMAARRC) compared to NOAR solutions for each operation.

Table 4-11. Two-way ANOVA between-subjects test results on completion time for step and solution factors in Diagnosis operation.

Operation: Diagnosis						
Factor	Df	Sum Sq	Mean	F value	Pr (>F)	Significant (95% ci)
Step	3	176247	58749	53.08	2.00e-16	Yes
Solution	2	132501	66250	59.86	2.00e-16	Yes
Step:Solution	6	137561	22927	20.71	2.00e-16	Yes
Residuals	168	185940	00117	-----	-----	-----

Post hoc comparisons from Tukey HSD tests (Table 4-12 and Table 4-13) can compare differences between factors groups on time means for each operation. Although ANOVA results suggest the solution factor is a significant effect, solutions' time means differences between authoring solutions are low compared to differences with non-AR solutions. Moreover, post-hoc comparisons for repair and diagnosis operations show that the mean differences for same-step groups of PMAU and alternative authoring solutions (ARAUM and SMAARRC) are not significantly different (p-values < 0.05). So, it can be said that the main effect is driven by the difference between AR and NOAR solutions rather than among AR solutions. This indicates the third hypothesis' validity, which enounced that time does not vary significantly among authoring solutions for each operation.

Table 4-12. Significance (p-value) results on post hoc comparisons (Tukey HSD) in Repair operation.

Operation: Repair | Legend: R = Step, P = PMAU, A = ARAUM, N = NOAR												
	R1:P	R2:P	R3:P	R4:P	R1:A	R2:A	R3:A	R4:A	R1:N	R2:N	R3:N	R4:N
R1:P	---	0.2085	0.9987	0.7285	1.0000	0.5576	1.0000	1.0000	0.0000	0.0000	0.0000	0.0004
R2:P	0.2085	---	0.8018	0.0003	0.2456	1.0000	0.1498	0.0842	0.0020	0.0000	0.0006	0.7620
R3:P	0.9987	0.8018	---	0.1586	0.9994	0.9828	0.9953	0.9776	0.0000	0.0000	0.0000	0.0140
R4:P	0.7285	0.0003	0.1586	---	0.6770	0.0031	0.8152	0.9160	0.0000	0.0000	0.0000	0.0000
R1:A	1.0000	0.2456	0.9994	0.6770	---	0.6128	1.0000	1.0000	0.0000	0.0000	0.0000	0.0006
R2:A	0.5576	1.0000	0.9828	0.0031	0.6128	---	0.4548	0.3088	0.0002	0.0000	0.0000	0.3740
R3:A	1.0000	0.1498	0.9953	0.8152	1.0000	0.4548	---	1.0000	0.0000	0.0000	0.0000	0.0002
R4:A	1.0000	0.0842	0.9776	0.9160	1.0000	0.3088	1.0000	---	0.0000	0.0000	0.0000	0.0001
R1:N	0.0000	0.0020	0.0000	0.0000	0.0000	0.0002	0.0000	0.0000	---	0.0179	1.0000	0.4349
R2:N	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0179	---	0.0469	0.0000
R3:N	0.0000	0.0006	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	1.0000	0.0469	---	0.2481
R4:N	0.0004	0.7620	0.0140	0.0000	0.0006	0.3740	0.0002	0.0001	0.4349	0.0000	0.2481	---

Table 4-13. Significance (p-value) results on post hoc comparisons (Tukey HSD) in Diagnosis operation.

Operation: Diagnosis | Legend: D = Step, P = PMAU, S = SMAARRC, N = NOAR												
	D1:P	D2:P	D3:P	D4:P	D1:S	D2:S	D3:S	D4:S	D1:N	D2:N	D3:N	D4:N
D1:P	---	0.0000	0.0000	0.0000	0.0674	0.0000	0.0000	0.0000	0.0028	0.0001	0.0418	0.0184
D2:P	0.0000	---	0.0284	0.0001	0.6144	1.0000	0.1455	0.0059	0.9913	0.0000	0.7271	0.8727
D3:P	0.0000	0.0284	---	0.9644	0.0001	0.0207	1.0000	1.0000	0.0004	0.0000	0.0000	0.0000
D4:P	0.0000	0.0001	0.9644	---	0.0000	0.0001	0.7091	0.9988	0.0001	0.0000	0.0000	0.0000
D1:S	0.0674	0.6144	0.0001	0.0000	---	0.6833	0.0001	0.0001	0.9984	0.0000	1.0000	1.0000
D2:S	0.0000	1.0000	0.0207	0.0001	0.6833	---	0.1140	0.0041	0.9960	0.0000	0.7880	0.9119
D3:S	0.0000	0.1455	1.0000	0.7091	0.0001	0.1140	---	0.9961	0.0041	0.0000	0.0002	0.0005
D4:S	0.0000	0.0059	1.0000	0.9988	0.0001	0.0041	0.9961	---	0.0001	0.0000	0.0000	0.0000
D1:N	0.0028	0.9913	0.0004	0.0001	0.9984	0.9960	0.0041	0.0001	---	0.0000	0.9997	1.0000
D2:N	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	---	0.0000	0.0000
D3:N	0.0418	0.7271	0.0000	0.0000	1.0000	0.7880	0.0002	0.0000	0.9997	0.0000	---	1.0000
D4:N	0.0184	0.8727	0.0000	0.0000	1.0000	0.9119	0.0005	0.0000	1.0000	0.0000	1.0000	---

Overall, previous discussions support the validity of the following considerations regarding the effect on completion time of authoring and NOAR solutions:

- For repair operations, completion times for PMAU and ARAUM authoring solutions are 42% faster than NOAR solutions.
- For diagnosis operations, completion times for PMAU and SMAARRC are respectively 41% and 45% faster than NOAR solutions.
- Differences in completion times between authoring and NOAR solutions can be considered significant for each maintenance operation.
- Differences in completion times among authoring solutions in each maintenance operation cannot be considered significant for each operation's step.
- Effect of authoring solutions is dependent on steps conducted for diagnosis operations but not for repair operations.

These results support the validity of this research's hypotheses regarding the positive effect on efficiency of the proposed authoring solution for multiple maintenance operations. Such effect is assumed to be achieved by the proposed authoring's ability to automatically produce content that is adaptive for enhancing semantic understanding of maintenance operations. A relevant method to further evaluate content adaptiveness is measuring its usability. Hence, the following subsection analyses the usability surveys that testers completed after experiments.

4.6.3 Usability study

Usability surveys study aims to validate the following hypothesis: *"Content usability does not vary significantly among authoring solutions for each maintenance operation"*. Usability can be described as a qualitative measure of the degree to which augmented content achieves user's semantic understanding of maintenance operations (Nielsen, 1993; Dey *et al.*, 2018). Table 4-6 presented a set of usability criterions for evaluating usability along with content's aspects regarding which 30 testers completed their usability surveys (Appendix F).

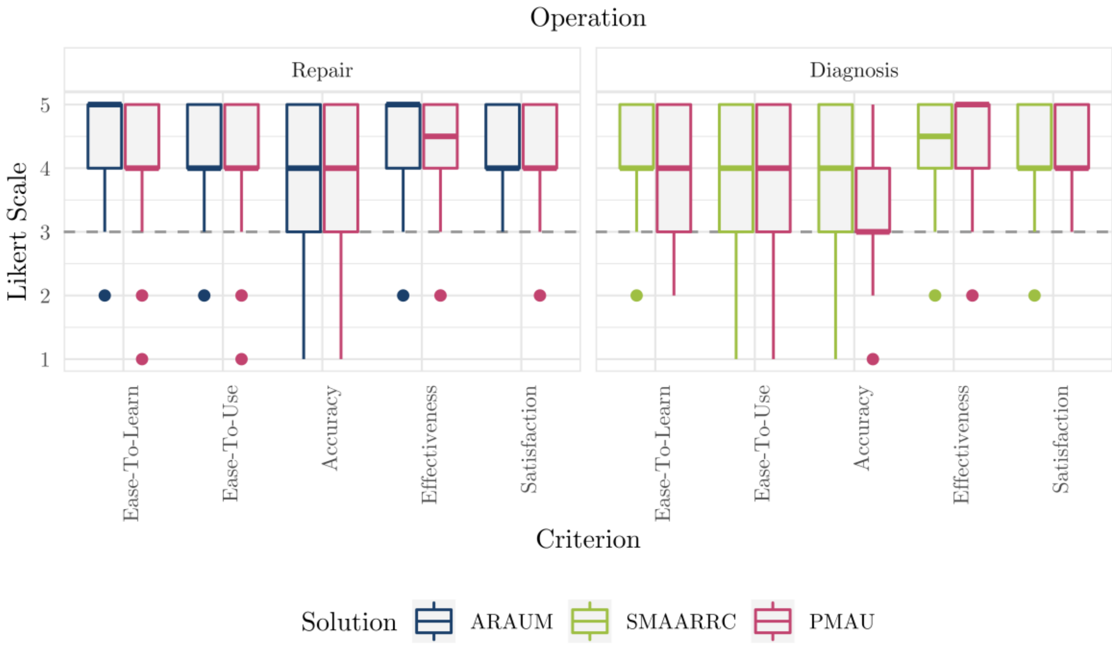


Figure 4-14. Distribution (maximum, upper quartile, median, lower quartile, minimum and outliers) of usability criterion's survey responses for each solution and operation.

Figure 4-14 illustrates a box and whiskers plot to summarise testers' responses for each usability criterion according to operation and solution experimental factors.

It shows that usability criterions do not differ considerably among PMAU and ad-hoc solutions (ARAUM and SMAARRC) in repair and diagnosis operations usability. Most criterions, medians scored between 4 and 5 with the exception of PMAU's accuracy in diagnosis that goes down to 3. Differences among medians and variabilities in each operation can indicate which AR solution was perceived by testers as more usable. In repair, ARAUM can be considered relatively more usable because variabilities are similar to PMAU for all criterions, but medians are higher in some of them. In diagnosis, SMAARRC and PMAU have similar medians for all criterions but for accuracy and effectiveness. While SMAARRC is considered a bit more accurate, PMAU is considered a bit more effective. Although distribution variabilities in these criterions show that these differences are not very significant. Overall, these numbers suggest that PMAU's content achieves is perceived by testers as usable as that from other authoring solutions. The only exception is PMAU's accuracy in diagnosis operation. A reason for this might be related to an event occurred during experiments that was connected to the HoloLens camera behaviour: tracking was being lost when testers were asked to get closer for inspecting the equipment.

The total number of testers' responses per criterion (60) provided sufficient data to analyse each of them separately. Independent box and whiskers plots (Figure 4-15 to Figure 4-19) for each criterion showing response averages per aspect can provide additional insights regarding further improvements on PMAU's usability.

Figure 4-15 displays testers' responses regarding Ease-To-Learn aspects compared by authoring solutions and operations. It suggests that PMAU's content was slightly more difficult to learn compared to that of other authoring solutions. ARAUM (tablet-based) had almost no differences between ease-to-use at start and at finish, while SMAARRC's had a slightly smaller difference between start and finish compared to PMAU. In terms of intuitiveness, only ARAUM's results indicate a better performance.

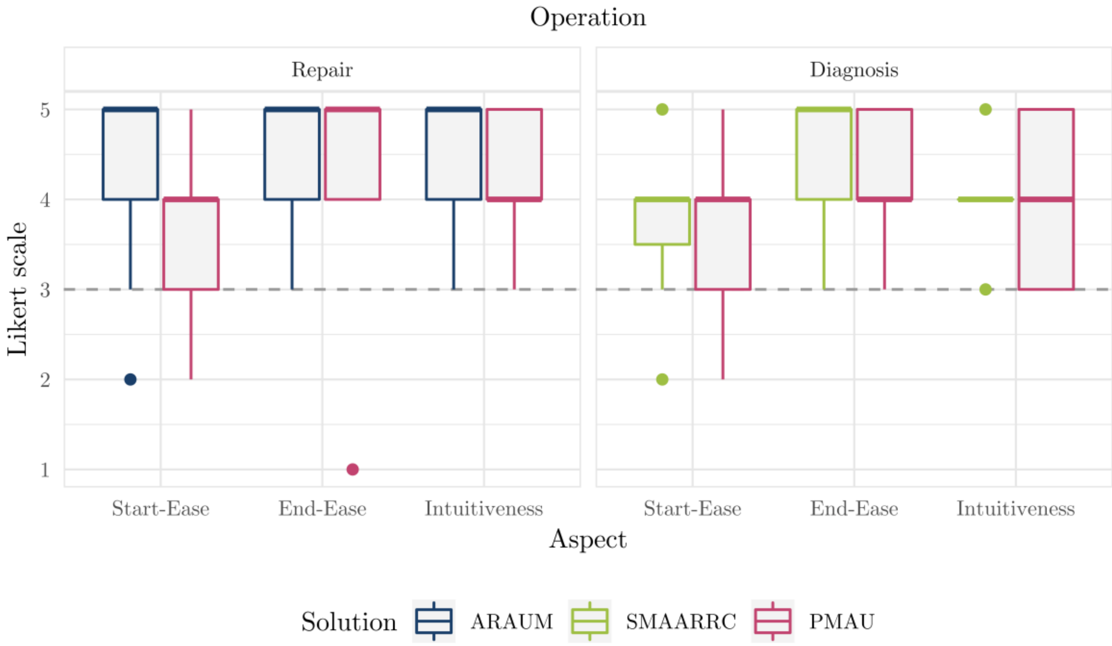


Figure 4-15. Distribution (maximum, upper quartile, median, lower quartile, minimum and outliers) of Ease-To-Learn aspects' responses for each solution and operation.

Figure 4-16 presents average testers' responses on Ease-To-Use aspects. These do not show interesting differences between authoring solutions in terms of content formats. Tablet-based solutions (ARAUM) showed better responses for text and buttons, while SMAARRC showed the worst results for 3D models.

Figure 4-17 describes testers' responses regarding Accuracy aspects of authoring solutions. These indicate that PMAU had a slightly worse performance in terms of latency. That could be explained due to the real-time PMAU's requirements regarding content generation. For other aspects, responses are quite similar for all three authoring solutions except for occlusion, where SMAARRC received a great variability on its responses.

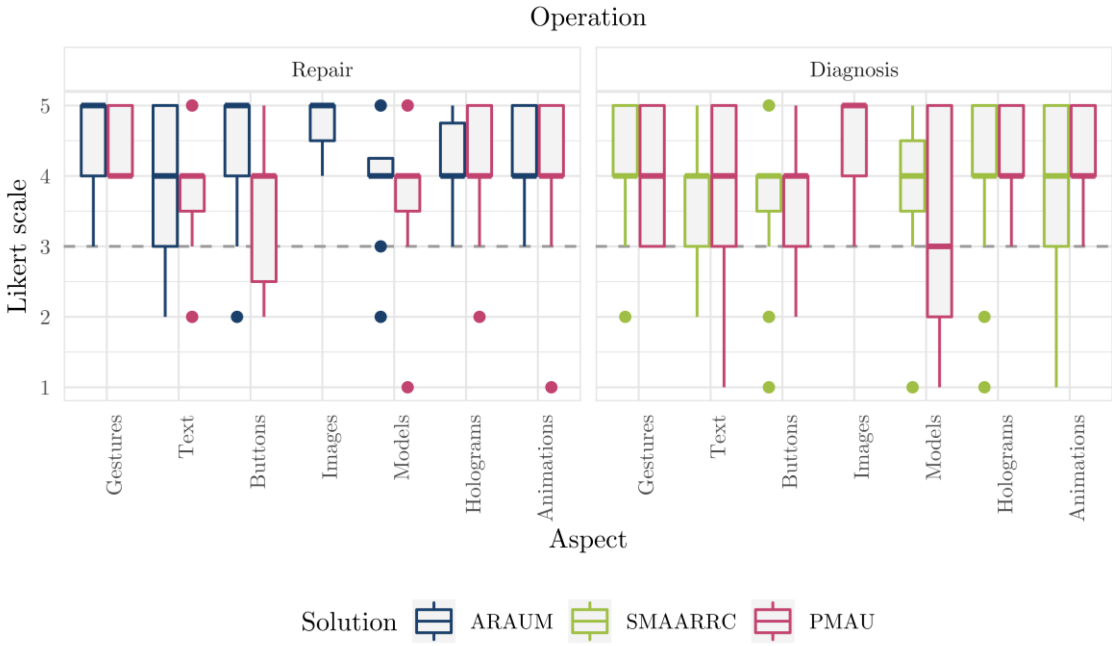


Figure 4-16. Distribution (maximum, upper quartile, median, lower quartile, minimum and outliers) of Ease-To-Use aspects' responses for each solution and operation.

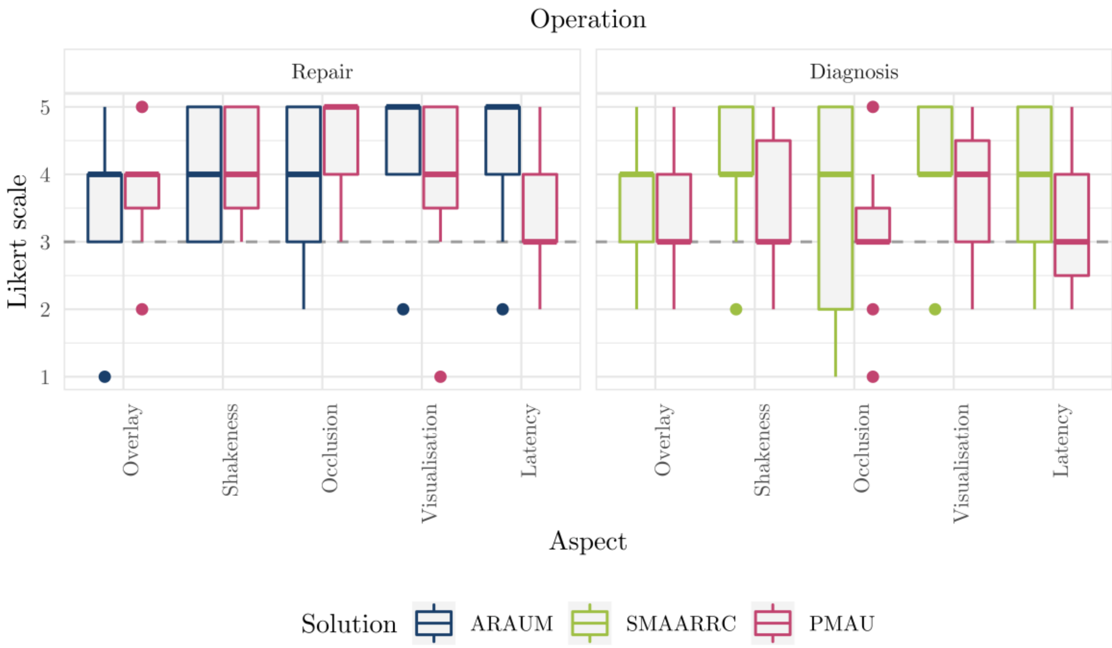


Figure 4-17. Distribution (maximum, upper quartile, median, lower quartile, minimum and outliers) of Accuracy aspects' responses for each solution and operation.

Figure 4-18 presents average testers' responses regarding Effectiveness aspects. These results indicate that all authoring solutions were considered similarly in terms of their abilities to reduce errors and missed instructions, and improve efficiency and confidence. One exception is PMAU's variability in ease-to-understand for diagnosis operations. Few testers noted during experiments that ontological naming conventions were sometimes difficult to understand. Thus, it seems important to adapt ontological wording for improved usability.

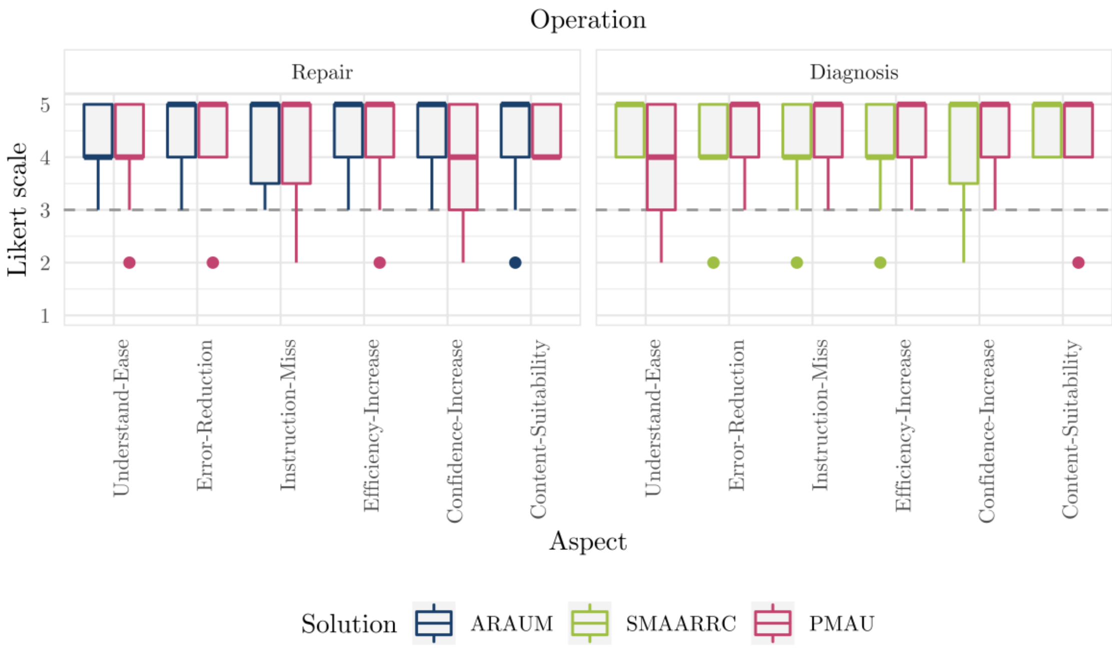


Figure 4-18. Distribution (maximum, upper quartile, median, lower quartile, minimum and outliers) of Effectiveness aspects' responses for each solution and operation.

Finally, Figure 4-19 summarises testers' responses regarding Satisfaction aspects. Satisfaction results were slightly higher for PMAU compared to other solutions. A reason for this can be the potential improvements testers identified in PMAU's ontological approach. Some of them noted the ability of PMAU's approach to track user's performance through content tracing (Figure 4-5). Because content is generated in real-time, content visualisation times can be easily traced to analyse content usage times and so, content adaptation effectiveness. Although it may require additional user-tracking techniques to ensure accurate measures.

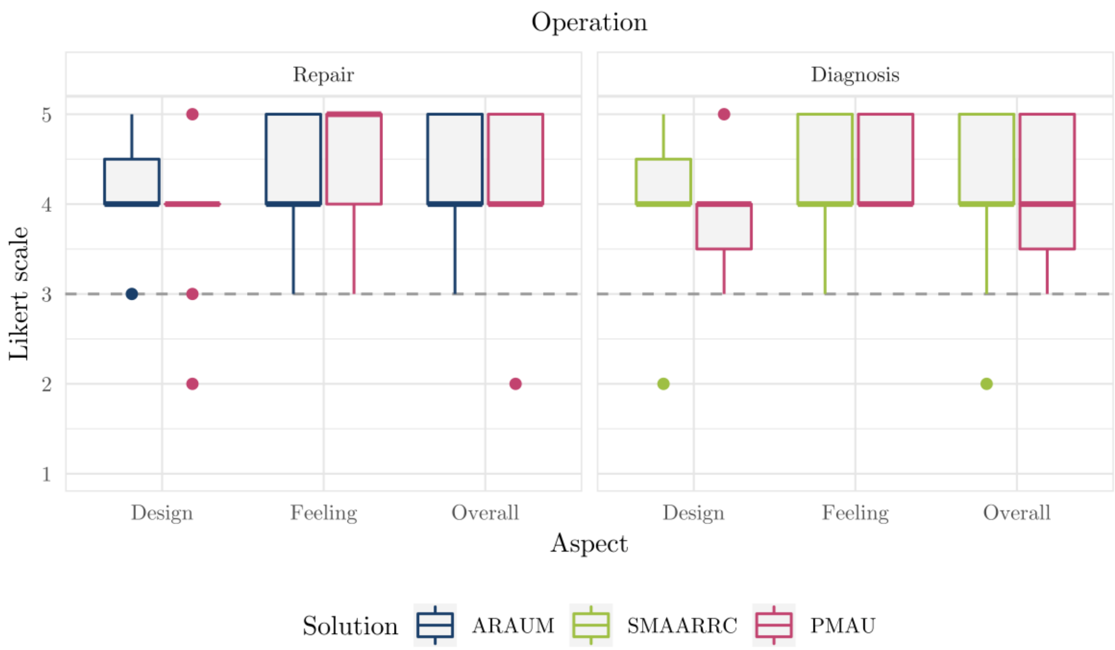


Figure 4-19. Distribution (maximum, upper quartile, median, lower quartile, minimum and outliers) of Satisfaction aspects' responses for each solution and operation.

Overall, testers survey results did not suggest a significant difference on content usability among authoring solutions. PMAU scored relatively lower in accuracy and text understanding, which are areas for further improvements. Moreover, PMAU's ability to track user's performance through content monitoring was also perceived as a good solution to further adapt content according to user's expertise. Hence, it can be said that these results indicate validity of the last research's hypothesis regarding insignificant content usability variance among authoring solutions for each maintenance operation.

4.6.4 Discussion

Previous analyses aimed to evaluate the research hypotheses described in Subsection 4.5.2, which intended to demonstrate the validity of this research contributions (Section 4.4).

The first validation hypothesis stated that "completion errors do not vary significantly among authoring and no-AR solutions for maintenance operations". Errors effect study (Subsection 4.6.1) analysed the correlation of operation and solution factors with experimental errors. It showed that number of errors per tester could be considered low, with an average of 0.422 errors. Results of ANOVA and t-tests did not indicate a significant variance on error results per solution and per operation. Therefore, the first hypothesis can be considered valid within the context of the experiments conducted. Thus, pondering completion times as a direct measure of maintenance efficiency. Nevertheless, number of errors were counted per test and not per test's step. So, it could not be studied the effect on completion errors of each experimental step. Though this may be an interesting element to evaluate, it was out of this study's scope because steps were predetermined to ensure maintenance quality's consistency among experiments. Future studies could investigate such effect by experimenting with the proposed authoring solution in real-life maintenance operations.

The second validation hypothesis assumed that "completion time decreases with authoring solutions compared to no-AR solutions for each maintenance operation". Instead, the third one stated that "completion time does not vary significantly between authoring solutions for each maintenance operation". Time effect study (Subsection 4.6.2) analysed the effect on time of solution and operation factors grouped by case of study. In repair operations (Erkoyuncu *et al.*, 2017), completion times for PMAU and ARAUM authoring solutions were found 42% faster than NOAR solutions. In diagnosis operations (Fernández del Amo *et al.*, 2019), completion times for PMAU and SMAARRC are respectively 41% and 45% faster than NOAR solutions. Besides, two-way ANOVA results indicated a significant difference on time between authoring and NOAR solutions but not among authoring solutions (PMAU, ARAUM and SMAARRC). Also, the effect of authoring solutions in diagnosis operations was found dependent on the maintenance step. Thus, confirming the results presented by Fernández del Amo, et al. (2019), which analysed that correlation. The student considered that analysis out of this research's scope because it aimed at proving the similarity between the effects of different authoring solutions. Nevertheless, future works can investigate the relation between augmented content usability and maintenance complexity to further improve content adaptiveness and relevant discard rules for content formats pairing (Subsection 4.4.2). Overall, these results prove valid the second and third hypothesis in the contexts of this research laboratory experiments. Therefore, it can be said that the proposed authoring's content can achieve similar effects on maintenance efficiency than other operation-specific authoring solutions.

The final validation hypothesis indicated that "content usability does not vary significantly among authoring solutions for each maintenance operation". Thus, aiming to evaluate whether the automatically generated content was usable from a tester's perspective for gaining semantic understanding of maintenance operations. Usability surveys results (Subsection 4.6.3) evaluated usability criteria according to different augmented content aspects. These results did not show significant differences on testers' responses about content usability between authoring solutions. Thus, confirming the assumptions of the abovementioned hypothesis. Nevertheless, PMAU scored relatively lower in accuracy and text understanding. But testers also noted its improved ability to track user's performance through more accurate content monitoring and further adapt content according to user's expertise. These are areas where future works can focus their efforts to achieve better effects of AR solutions in maintenance operations.

The analyses results discussed above aimed to validate this research contributions for their ability to automatically create adaptive content for multiple maintenance operations. Although this validation's hypotheses can be considered proven in the context of this research's cases of study, the following paragraphs consider some relevant aspects to discuss.

The first contribution described a method to declare programmable formats that semantically describe their data, user and environmental requirements for producing augmented content. Its aim is to create templates with certain augmentation behaviours that can later be matched with maintenance datasets. These formats and their behaviours comprise different combinations of visualisation and interaction modes. Thus, enabling AR developers to create content for all kinds of maintenance tasks, scenarios and expertise levels. Most content formats implemented in this research replicate those presented by Erkoyuncu *et al.* (2017) for repair and by Fernández del Amo *et al.* (2019) for remote diagnosis operations. This research also developed some more generic formats to ensure augmentation of all individual properties of datatype and object types. Moreover, there exist two on-going researches where this research's authoring proposal is being utilised to develop new formats for thermographic assessment (Wali *et al.*, 2020) and diagnosis reporting (Chapter 5) operations. The reason to implement formats from different researches was to demonstrate that the proposed authoring method can create content for multiple maintenance operations. This can be further corroborated in future works by using adaptive formats already researched such as those from Chang, Nee and Ong (2020) and Wang, Ong and Nee (2016a) for assembly. Future works can also develop more adaptive formats for less researched operations such as monitoring. Besides, future works can also focus on more basic research for improving content adaptiveness. These can investigate the relation among visualisation and interaction methods (e.g. animations) with human performance (e.g. sight). Thus, designing more accurate descriptors for content formats to enable automatic adaptation to user (e.g. expertise) and environmental (e.g. light) conditions.

The second contribution proposed a real-time, ontology-based, pattern-matching algorithm to pair content formats with ontology individuals for automatically creating, adapting and locating augmented content. The algorithm comprises different assignation and discard rules (Subsection 4.4.2) to match individual's properties with formats' data, user and environment facets. Although these rules have proven sufficient to match ontology individuals with specific content formats for repair and diagnosis applications, they still depend on formats' declarations made by AR developers. Future works can research more advanced methods to declare content formats and rules to pattern-match them. These may include techniques like natural language processing, environment and user tracking (e.g. light conditions or user attention) and so forth.

The proposed algorithm also parts content creation and information management authoring processes to automate the first one. For this reason, the student considered that authoring efficiency experiments were out of this research's scope. Nevertheless, maintenance experts still need to perform information management processes. In this research, the web-based ontology reporting tool presented in (Chapter 3) has been used for this purpose. Future works can further improve this process. They can analyse the effect of ontology wording and its impact on AR semantic understanding and design tools for declaring user-adaptive ontologies. Besides, they can also further evaluate the impact of different authoring solutions in AR deployment costs. Thus, easing the implementation of AR technologies in maintenance organisations.

A relevant feature of the proposed algorithm relates to its ability for generating augmented content in real-time. This allows not only to enable AR applications such as remote collaborative diagnosis (Fernández del Amo *et al.*, 2019) but also to perform tracking of content being used (Figure 4-5). This research's system implementation enabled reporting capabilities to trace individuals augmented and their content creation dates, although its benefits have not been explored. Future works can further study this ontology-based content-tracking feature and its impact on content adaptiveness as well as maintenance performance evaluation. Moreover, they can also improve its accuracy with more advanced user-tracking techniques like eye-tracking and other biometric technologies.

Another relevant algorithm's feature involves its use of ontologies. Unlike other ontology-based authoring techniques (Flotyński and Walczak, 2015; Zhu, Ong and Nee, 2015; X Wang, Ong and Nee, 2016a; Walczak and Flotyński, 2019), this algorithm does not use ontologies for inferencing purposes but to enable information standardisation for semantic analysis. Thus, allowing individuals to be augmented adaptively through different formats according to user and environment facets. Hence, it is feasible to consider that this algorithm could also be used with other information management methods (e.g. SQL or graphical databases), if those were to meet these standardisation requirements (Subsection 4.4.1). Future works could further investigate on those requirements for extending this algorithm's applicability to other information management methods. Besides, this algorithm aims to augment existing ontology individuals but does not consider the creation of new instances. Future works can extend this research by including new algorithm features as well as content formats to enable for ontology individuals to be created using AR applications. Thus, enabling AR technologies not only to transfer but also to capture knowledge. Moreover, ontology individual's instantiation may require defining links to existing individuals. The number of individuals to select from in new instantiations can be high and so, their content can overload the augmented scene. Therefore, future works should also investigate recommendation techniques for improved information filtering on AR-based knowledge capture applications. These can consider the relation between recommender systems and the algorithm's content-tracing capabilities to enable maintenance context-aware recommendations, as studied in Chapter 5.

This research's contributions aim at automatically producing adaptive content for multiple maintenance operations. Although the described experiments proved so for repair and remote diagnosis tasks, these were conducted in laboratory setups. The reason to do so was to maintain consistency with previous researches. Nevertheless, future works described above can further corroborate this research results with experiments in real-life conditions, including evaluation of other relevant factors in AR usability like ergonomics. Besides, the proposed contributions focus on AR-maintenance applications and so, they include some assumptions regarding the use of certain AR methods like tracking and registration (Subsection 4.4.2). Therefore, future works can study the applicability of these contributions to other AR fields of application such as medicine, tourism and so forth. Thus, aiming to develop a framework for automatic authoring in AR that could ease its implementation in commercial and industrial markets.

4.7 Conclusions and future works

4.7.1 Conclusions

This chapter proposed (1) a method to declare programmable content formats according to its data, user and environmental augmentation requisites and (2) a real-time, ontology-based pattern-matching algorithm to couple content formats with ontology individuals. Their aim was to prove that adaptive content can be automatically created for multiple maintenance operations independently of the information system's structured utilised. These methods were implemented in a cloud-based AR application prototype for validation with efficiency experiments and usability surveys in comparison against two manual-authoring solutions specific for repair and remote diagnosis operations. Experimental results indicated that augmented content from the proposed solution (PMAU) provided similar support to that of alternative authoring solutions (ARAUM and SMAARRC) in terms of maintenance efficiency improvements (42% time reduction). Besides, survey results suggested no significant differences on testers' opinions regarding content's usability from experimented authoring solutions. Thus, proving that the proposed authoring solution can automatically create content of similar quality and effectiveness than content produced by manual, operation-specific authoring alternatives for multiple maintenance operations.

The proposed methods for automatic adaptive authoring contribute to fill an important research gap towards the understanding of multiple maintenance operations and the ease of AR deployment in industrial contexts. For maintenance organisations, deploying AR applications that can provide adaptive content to support maintainers for sustainably achieving efficiency improvements can be costly. That is because creating adaptive content requires either manual authoring or specific information systems that duplicate the existing ones for automatic authoring. The proposed method for declaring programmable content formats can standardise them in terms of data, user and environmental requisites. This allows automatic authoring approaches to re-use adaptive content formats for enhancing user's understanding of multiple maintenance operations. Besides, the proposed pattern-matching algorithm automates the authoring process while using existing ontology-based information systems. This allows to further reduce the associated costs of alternative automatic authoring solutions. Hence, this research's proposal can simplify the deployment of AR applications to adaptively support multiple maintenance operations and facilitate their integration with existing ontology-based maintenance information systems.

4.7.2 Future works

Future works will explore further applications and enhancements of the proposed methods to: (1) improve content adaptiveness according to user performance and expertise, environment conditions and other maintenance operations, (2) ease their integration with non-ontology-based maintenance information systems and (3) extend the proposed automatic authoring framework to cover other AR application areas. The following list summarises future works identified within this chapter's discussion that cover the areas mentioned above:

- Investigate the relation among augmented content usability and maintenance complexity regarding efficiency improvements to further improve content formats adaptiveness and associated discard rules.
- Study the effect of ontology wording on augmented content's semantic understanding and develop methods to enhance it through user-adaptive ontologies.
- Research in-depth the relation between visualisation/interaction methods and maintainer's performance and design content formats for automatic adaptation to maintainer's conditions like expertise, ergonomics, cognition and so forth.
- Research in-depth the relation between visualisation/interaction methods and environment conditions and design content formats that use sensor data (e.g. light) for automatic adaption.
- Besides on-going studies for diagnosis reporting and thermographic assessment, research the relation between visualisation/interaction methods and less researched maintenance tasks (e.g. structural monitoring) to develop suitable content formats for achieving all-in-one AR maintenance applications.
- Improve format's user and environment facets and algorithm's discard rules for pattern-matching more adaptive formats as those above. That can include advanced methods like eye- or light-tracking to automate user and environment characterisation.

- Improve format's data facets and algorithm's assignation rules for pattern-matching more adaptive formats as those above. That can include advanced methods like content-tracing (Subsection 4.4.2), ontology inferencing or natural language processing to automate maintenance context's description.
- Extend content formats and facets and algorithm's rules for enabling ontology individuals to be instantiated through augmented content. Thus, enabling AR technologies for knowledge capture applications.
- Research recommendation techniques for improved information filtering on AR knowledge capture and transfer applications. That can include the relation with the proposed algorithm's content-tracing capabilities to enable maintenance context-aware recommendations.
- Study alternative maintenance information standardisation methods for pattern-matching formats to extend the proposed algorithm's applicability to other information management techniques like SQL or graphical databases.
- Corroborate this research's results and future works described above with experiments in real-life maintenance conditions. These can involve evaluating additional factors regarding AR usability like ergonomics or environment.
- Study these research's contributions applicability to other fields of AR application such as medicine or tourism. That can involve the algorithm's assumptions on used AR techniques like object tracking and registration, content elements allocation, and so forth.
- Evaluate the proposed authoring solution's impact in AR deployment costs in comparison with existing alternatives.

Augmented Reality technologies are just information visualisation and interaction tools. As such, they should be as easy to deploy and integrate as similar technologies like web pages. These future works aim to find the necessary research towards a framework for adaptive and automatic authoring that is applicable to any existing application and information system. Thus, envisioning a future where AR applications can be easily deployed by any business organisation.

Chapter 5
Hybrid recommendations and dynamic authoring for Augmented Reality knowledge capture and re-use in maintenance diagnosis applications

5.1 Introduction

Industry 4.0 is bringing digitalisation of complex equipment to improve availability and sustainment for achieving more intelligent and adaptive industrial operations (Cimino, Negri and Fumagalli, 2019). Technologies like Artificial Intelligence (AI) and Digital Twins (DT) are enabling real-time and automatic data capture and analysis, monitoring and optimisation for faster and improved decision-making (Angelopoulos *et al.*, 2020). These require integrated data management that is still challenging due to heterogeneity of data formats (e.g. audio or video) and lack of structure of numerous data sources (e.g. manuals or reports) (Khan and Yairi, 2018). In this context, human knowledge capture is fostering research attention due to its potential to create valuable data streams (Longo, Nicoletti and Padovano, 2019). A relevant area for knowledge capture is maintenance diagnosis. Digital diagnosis systems aim to automate fault-finding tasks through sensor data capture and analysis to identify failures' root causes (Angelopoulos *et al.*, 2020). But when these systems fail to determine the cause of an identified failure like in No-Fault-Found (NFF) events, experts are still required to do so (Wan *et al.*, 2019). When maintainers conduct diagnosis procedures, they are usually asked to report them along with failure modes and conditions identified (Li, Fast-Berglund and Paulin, 2019). These reports contain valuable knowledge to be re-used not only for enhancing diagnosis systems (Pérez-Salazar *et al.*, 2019) but also consequent reporting tasks where those modes and conditions reappear (Pistofidis *et al.*, 2016). Although capturing such knowledge seems relevant to enhance diagnosis systems, that should be done in a structured manner to enable knowledge to be re-used (Liu *et al.*, 2019).

Augmented Reality (AR) technologies can embed digital information in human operations in co-existence with real-world objects (Palmarini *et al.*, 2018). They have been widely applied to support complex human tasks (e.g. maintenance repair or medical surgeries) because of their abilities to transfer knowledge (Bottani and Vignali, 2019). AR has also been applied for knowledge capture and re-use in manual maintenance operations like repair and assembly (Bhattacharya and Winer, 2019). But as this thesis identified in Chapter 2, there is very little academic evidence of AR knowledge capture and re-use applications in abstract operations like diagnosis reporting. A possible reason relates to the complexities associated with representing abstract knowledge and data input forms through augmented content, which can result in incorrect reports (Ramirez-Amaro, Beetz and Cheng, 2017). Another reason involves diagnosis reporting tasks that require selection of items from extensive datasets like fault conditions (Liu *et al.*, 2019). These can overload the augmented scene and the user's field of view if large selection lists are augmented without prior filtering or recommendation of relevant items (Gattullo *et al.*, 2019).

In order to enhance knowledge capture and re-use in diagnosis reporting tasks, this research proposes hybrid recommendations and dynamic authoring for AR maintenance applications. This proposal includes the following contributions:

1. A method to provide context-aware and ontology-based AR recommendations to reduce extensive selection lists in reporting applications through knowledge re-use for decreasing reporting time.

2. A method to dynamically create and allocate content in augmented scenes to avoid augmented content overload and ease data input tasks for reducing reporting time and errors.

This research aims to demonstrate that AR knowledge capture and re-use applications can achieve sufficient reporting effectiveness and efficiency to enable structured knowledge capture. Thus, allowing to re-use it in consequent diagnosis reporting activities. Therefore, this research validates the proposed contributions against conventional alternative reporting solutions for diagnosis reporting tasks.

The rest of this chapter's structure is as follows. Section 5.2 presents a literature review on diagnosis applications and recommendation techniques in AR research to detect current research gaps. Section 5.3 describes the methodology employed to determine, develop and validate this research's contributions, which are declared in Section 5.4. Section 5.5 presents this research's validation protocol along with experimental and survey methods and cases of study. This validation's results are analysed and discussed in Section 5.6. Finally, Section 5.7 presents this research's conclusions and advances future works to enable human knowledge integration in extended Industry 4.0 contexts.

5.2 Literature review

5.2.1 Diagnosis applications in Augmented Reality

Diagnosis can be described as the series of human activities that determine the causes of asset's failures or abnormal behaviours (Zhou, Yu and Zhang, 2015). Such activities include different tasks such as manual's consultation, fault-finding investigation and reporting (Medina-Oliva *et al.*, 2014). AR technologies have the ability to contextualise virtual information and so can improve the efficiency and effectiveness of diagnosis tasks (Wójcicki, 2014). For example, Longo, Nicoletti and Padovano (2019) described an AR-Digital Twin integration model to augment sensor data and service records for enhancing fault-finding tasks in shop-floors. A similar approach was utilised by Avalle *et al.* (2019) to support fault detection on industrial robots. Ghimire, Pattipati and Luh (2016) and Khalil *et* *al.* (2019) use integrative models of equipment to understand relations between sensor data from different components and visualise remaining useful life indicators (e.g. fatigue). Others like Priya and Vasudevan (2018), Wang *et al.* (2018) or Das, Dong and Scherer (2018) made use of extended AR tracking capabilities to detect missing or mis-shaped components for enhancing visual inspection tasks. These studies took advantage of AR technological capabilities (e.g. tracking) to develop advanced data-driven prognostics for enhancing real-time sensor-data visualisation that supports diagnosis tasks.

Besides data-driven diagnostics, AR research have also focused on knowledge-based methods. Due to its knowledge transfer abilities, academics identified AR as a suitable technology to provide remote expert support for diagnosis tasks. Authors like Masoni *et al.* (2017), Rambach *et al.* (2018) or Zenati-Henda *et al.* (2014) focused their efforts on user interfaces that allowed remote experts to seamlessly integrate indications for on-site technicians to conduct fault-finding tasks. Others like Oyekan *et al.* (2017), Hadar *et al.* (2017) and Mourtzis, Vlachou and Zogopoulos (2017a) provided intelligent fault-tree algorithms to support early diagnosis stages and reduce remote expert's workload. These studies aimed to ease the knowledge transfer from remote experts to on-site technicians, although few papers analysed the effects of models utilised to send such knowledge through AR. One of them, proposed by the student (Fernández del Amo *et al.*, 2019), described a message structure and associated automatic content-creation rules to enhance knowledge sharing in remote diagnosis scenarios.

Overall, AR research in diagnosis applications has focused on improving sensor-data visualisation and remote expert knowledge transfer for several diagnosis tasks such as fault-finding or visual inspection. However, there is little academic evidence on AR applications to support diagnosis reporting tasks. Although, few papers made use of AR abilities for knowledge capture (Zenati-Henda *et al.*, 2014; Fernández del Amo *et al.*, 2019; Longo, Nicoletti and Padovano, 2019) or demonstrated its potential benefits in terms of knowledge re-use (Fernández Del Amo *et al.*, 2018).

5.2.2 Recommendation techniques in Augmented Reality

Knowledge re-use can be described as the series of processes to capture, package and share knowledge (Markus, 2001). Hence, knowledge can be seen as a set of items (e.g. documents or ontology's instances) to be classified, evaluated and suggested for a given purpose in a given context. That is the purpose of knowledge-based recommendation systems (Aggarwal, 2016). Recommender techniques have been widely applied in AR research to filter and contextualise augmented content for reducing augmented scene's overload and improving task's support (Schaeffer *et al.*, 2018). For example, Schaeffer *et al.* (2018) proposed a content-based recommender algorithm to inform consumers on product's ecological impact at the point of sale. Collaborative-filtering (user-based) recommenders have also been used in AR, Lin and Chen (2020) described a deep-learning algorithm to advise learnings tasks and methods through AR according to their learning stage. The same applies to knowledge-based recommenders, Torres-Ruiz *et al.* (2020) described a pattern-matching algorithm to suggest museum itineraries. AR studies have evaluated all kinds of recommendation techniques (e.g. context-aware, content-based, ontology-based, etc.), although most of them were applied for commercial and educational purposes. While there exists evidence of recommendation techniques for knowledge reuse in diagnosis applications (Wang, Tang and Wu, 2010; Dendani, Khadir and Guessoum, 2012; Medina-Oliva *et al.*, 2014; Zhou, Yu and Zhang, 2015; Renu *et al.*, 2016), the student could not find any that evaluated the potential benefits of AR technologies like context-awareness for enhancing knowledge re-use.

5.2.3 Research gaps

Latest research reviewed on AR for diagnosis applications makes use of its knowledge delivery abilities to support human-related tasks like fault-finding or visual inspection with enhanced sensor-data visualisation. Although the student found some academic literature on AR diagnosis applications for supporting knowledge capture tasks like reporting (Chapter 2), it presented little evidence of the effects and benefits of knowledge re-use. However, AR knowledge re-use abilities have been widely applied in other applications like education or tourism in the form of recommendation methods (Schaeffer *et al.*, 2018). Various recommender techniques (e.g. knowledge-based, collaborative filtering or context-aware) have been used mainly either to reduce user's field-of-view overload or to improve AR support of human-related tasks. Moreover, research on diagnosis have also analysed the benefits of knowledge re-use recommendation methods, although the student could not find any that evaluated AR benefits like context-awareness.

Based on the discussions above, the student identified several research gaps on AR research for diagnosis applications. First, there is lack of evidence on AR diagnosis applications for reporting tasks. Second, there is also little evidence of AR research that evaluates the effect of knowledge re-use in diagnosis applications. If both gaps were to be fulfilled, such research would be able to evaluate the effects of AR technologies for diagnosis reporting applications and also the benefits of context-aware knowledge re-use techniques. This will help to further improve the efficiency of diagnosis tasks and to ease the integration of expert knowledge as a data source for diagnosis systems or Digital Twins in the context of Industry 4.0. In order to contribute to the fulfilment of both research gaps, this chapter proposes (1) an AR-based hybrid recommender method for improving diagnosis reporting tasks through knowledge re-use and (2) a method for dynamic reporting-content creation. The former takes advantages of both, knowledge-based and context-aware recommendations, through AR authoring and tracking techniques. This research aims to validate the proposal's ability to enhance diagnosis knowledge capture and re-use while sustaining the efficiency of fault-finding tasks. Thus, improving the integration and applicability of expert knowledge in the context of Industry 4.0.

5.3 Methodology

This research aims to prove that AR applications can achieve sufficient reporting efficiency and effectiveness for enabling knowledge capture and re-use in diagnosis reporting tasks. Inspired by similar researches (Chi *et al.*, 2012; Gimeno *et al.*, 2013; X Wang, Ong and Nee, 2016a; Longo, Nicoletti and Padovano, 2019) and well-established frameworks in design research (Peffers *et al.*, 2008), this research employed the following methodological steps:

1. Objectives identification: "define specific opportunities and justify the value of a solution". Section 5.2 presented a literature review to identify academic research gaps that followed the protocol presented by Booth, Papaioannou and Sutton (2012).

2. Solution design: "create a solution to satisfy the research opportunities". Section 5.4 presents the proposed recommendation and authoring methods designed to satisfy AR-maintenance research contributions.

3. Solution demonstration: "prove the solution's use to satisfy the research opportunities". Section 5.5 presents this research application of feature-driven software development (Nawaz, Aftab and Anwer, 2017) to implement the proposed solution in an AR prototype for two experimental cases of study.

4. Research validation: "measure the solution's impact on research opportunities". Efficiency experiments and usability surveys evaluated to impact of this research's proposals on two reporting tasks through comparison against alternative reporting tools. Section 5.5 describes these methods protocols, while Section 5.6 analyses and discusses their results. Finally, Section 5.7 summarises this research's conclusions and suggests future works.

5.4 Hybrid recommendations and dynamic authoring methods

In Chapter 4, the student proposed a framework for automatic authoring in AR knowledge transfer applications (PMAU). It included (1) a method to declare programmable content formats and (2) a real-time, ontology-based, pattern-matching algorithm to pair formats and maintenance datasets. This proposal (Figure 5-1-a) had several limitations for knowledge capture and re-use applications. First, it proposed static content allocation that resulted on occlusion and overload issues for assets bigger than 1.5 meters in width. Second, it proposed adaptive content formats for knowledge transfer applications but did not consider formats for knowledge capture ones that ensure correctness of data inputs. Third, it did enable to adapt maintenance information to different content formats but did not consider adapting information itself to the maintenance context.

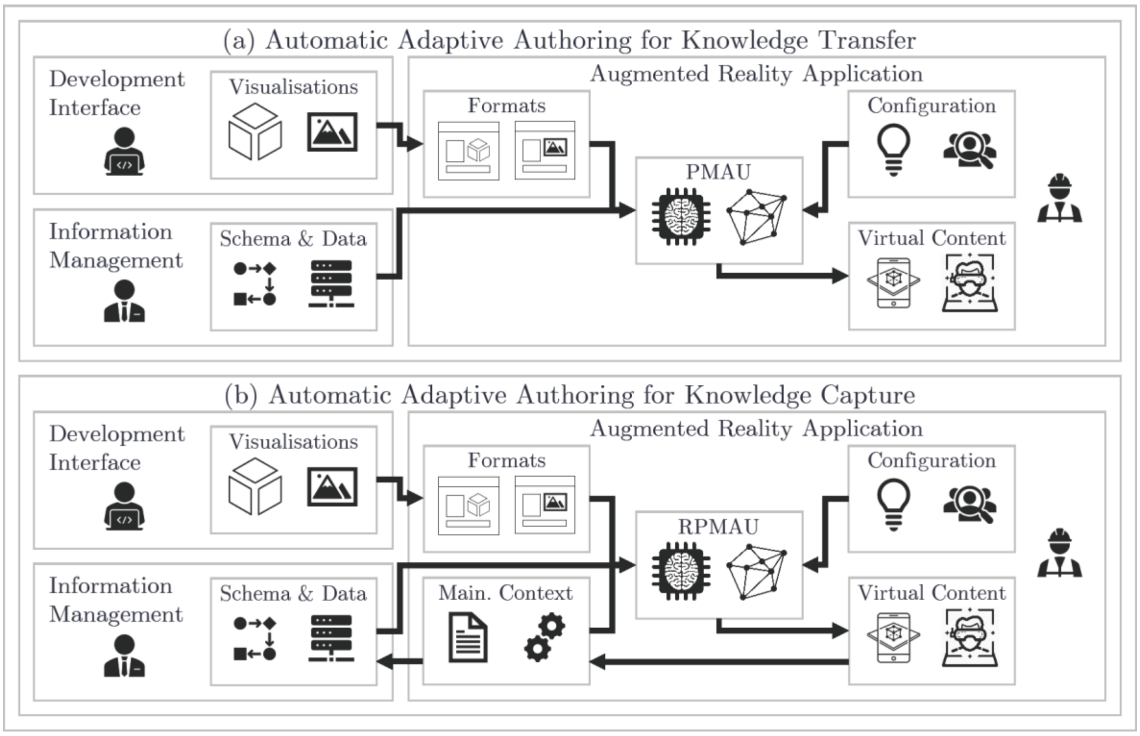


Figure 5-1. Overview of automatic authoring for knowledge (a) transfer and (b) capture applications – It compares software modules (development interface, information management and AR application) in authoring approaches to identify new components needed for automatic authoring in knowledge capture applications.

For solving the issues above and fulfil the research gaps identified in Section 5.2, this chapter proposes hybrid recommendation and dynamic authoring techniques (RPMAU – Figure 5-1-b) to extend the previous proposal for knowledge capture and re-use applications. This chapter's contributions consist of the following:

1. A method to provide context-aware and ontology-based AR recommendations to improve efficiency of diagnosis reporting tasks through knowledge re-use.
2. A method to dynamically create and allocate content in augmented scenes to avoid their overload and improve efficiency and effectiveness of reporting tasks.

Figure 5-2 presents some examples of augmented content for knowledge capture and re-use applications, which have been generated using the contributions above. These include recommendations to enhance knowledge re-use and models and text-to-speech to enhance knowledge capture effectiveness by avoiding mistyping of data inputted.

Figure 5-2. Examples of automatically created augmented content of different visualisation and interaction modes for knowledge capture applications.

Figure 5-3 presents an overview of the extended authoring method (RPMAU) including the contributions proposed in this research. Similarly to PMAU, RPMAU starts when the user 'selects an ontological class'[3] that represents the information that the user wishes to report. Then, it analyses user-configured environment and user features to discard content formats that are not suitable for the augmented scene according to real-world conditions like lack of light, noise,

[3] Words between quotation marks in this paragraph refer to Figure 5-3.

etc. After that, RPMAU also conducts semantic analysis of content formats' data facets and ontological asserted properties to match which content formats (e.g. models or text) are suitable to augment properties' information. It is at this point where RPMAU starts differing from PMAU. RPMAU proposes new content formats that include checking features to ensure correctness of data being reported. Matching of content formats and ontological properties is referred in Figure 5-3 as 'assign fabrications' because fabrications are defined as pieces of augmented content that implement formats visualisation and interaction modes to augment a set of individual properties. Once fabrications are assigned, RPMAU also differs from PMAU as its 'applies recommendation facets'. At this point, RPMAU uses the proposed recommender technique to filter augmented information prior to augmentation. Recommendation rules are declared as facets so they can be implemented as an extension for programmable content formats.

Once recommendations are applied, fabrications are ready for augmentation in the scene in the form of elements. Elements are sets of fabrications that comprise all asserted properties of the ontological class being reported. Again, RPMAU differs from PMAU as so does the method to 'allocate a spot' (Figure 5-3) for the element in the augmented scene. In RPMAU, the asset being maintained is analysed in terms of its size to determine what type of allocation is best to avoid content overload, either around the user or around the asset. Once the content is augmented in the scene, the user can interact with it for reporting purposes. When the user is finished, RPMAU, differently to PMAU, 'submits the resultant ontology individual' to the knowledge base and keeps 'trace' of it for analysing the maintenance context in further recommendations.

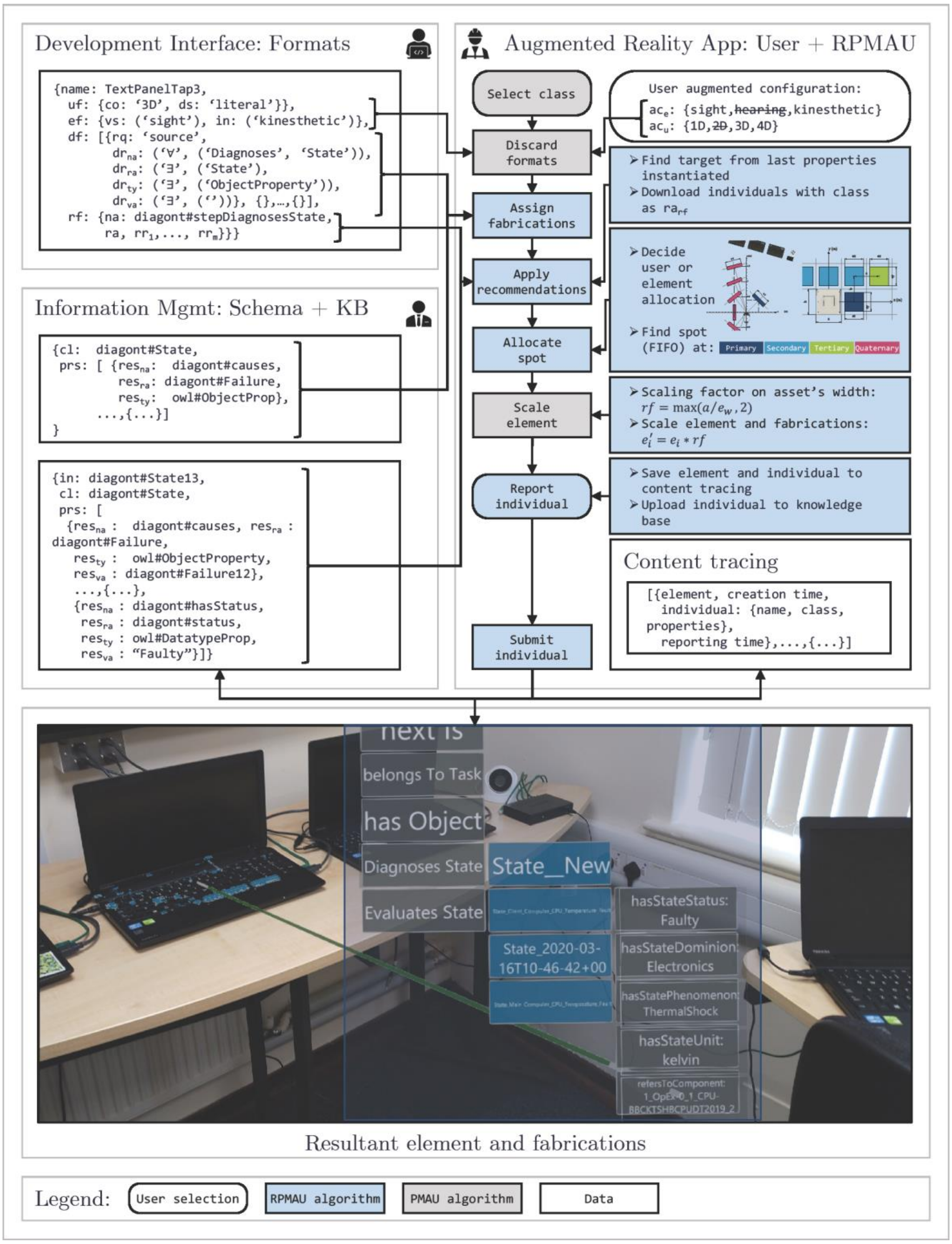


Figure 5-3. Dynamic authoring algorithm including hybrid recommendations to enable knowledge capture and re-use – The figure details the overview presented in Figure 5-1-b. It exemplifies how the RPMAU algorithm uses a class (cl(prm,dcp,dop) – (5-1)), formats with recommendation facets (ft(uf,ef,dfn,rfl) – (5-4)) and augmentation configuration (ac(ace(seni),acu(dmj,dsk) – (4-13)) to automatically generate augmented content that provides context-aware, ontology-based

recommendations of individuals (in(prm,dcp,dop) – (4-1)). The figure displays classes (cl), individuals (in), formats (ft) and augmentation configuration (ac) as JSON objects to show their data. The figure also presents the different steps of the RPMAU algorithm, including their use of the abovementioned data along with the steps inherited from PMAU. New RPMAU and PMAU's modified steps (blue boxes) are later explained in Subsections 5.4.1.2 and 5.4.2: assign fabrications, apply recommendations, allocate spot, report individual and submit individual. The resultant exemplary element and its fabrications are displayed at the bottom of the figure. Each box containing data (cl, in, ft, ac), RPMAU steps and user selections have a different shape and colour as described in the legend.

The previous description of RPMAU in comparison with PMAU includes the contributions proposed by this research, which are explained in-depth in following subsections. First, Subsection 5.4.1 presents the dynamic authoring method as an extension for knowledge capture applications of the previous automatic authoring method for knowledge transfer applications (Chapter 4). This includes RPMAU's content formats and the context-aware content allocation method and. Then, Subsection 5.4.2 describes the context-aware, ontology-based AR recommender technique that extends the dynamic authoring proposal to enable knowledge re-use, which comprises RPMAU's recommendation facets and their application. Finally, Subsection 5.4.3 presents these contributions implementation to diagnosis reporting tasks.

5.4.1 Dynamic authoring for Augmented Reality knowledge capture applications

The dynamic authoring proposal for AR knowledge capture applications can be considered an extension of a previous automatic authoring method for knowledge transfer applications presented in Chapter 4. This extension can be considered a relevant contribution since it proposes two new algorithmic modifications to the previous method for enabling AR knowledge capture applications. These modifications are (1) content formats for knowledge capture with data correctness checking features and (2) the context-aware element allocation procedure. These are explained in the following subsections.

5.4.1.1 Content formats for structured and correct knowledge capture

PMAU's pattern-matching algorithm uses data, user and environment facets to assign and discard suitable content formats for augmenting a given ontology's individual for knowledge transfer. PMAU's proposal assumes that ontology individuals have properties (attributes and relationships) asserted and so, these can be assigned to formats for augmentation. Nevertheless, AR-based individual instantiation (knowledge capture) would require the AR user to make those assertions. In order for RPMAU to augment elements for individual instantiation, it would require some modifications to PMAU. These are the following:

1. It would require allowing the user to select for both individual visualisation and instantiation procedures.

2. It would require determining ontology classes' "exemplary" individuals to accept possible data facets' rules of value.

3. It would require submitting instantiated individuals to the ontology server and so, ensuring instantiation correctness.

4. It would require programmable content formats capable of capturing data correctly and efficiency through user interaction.

RPMAU's proposal extends PMAU with the modifications above to enable automatic creation of augmented content for knowledge capture applications.

The first requirement is solved by embedding a procedure selection option as part of AR application's initialisation behaviour. Thus, users can choose whether use the AR application to visualise or instantiate ontologies' individuals, which will affect further AR application's behaviour (e.g. formats colour coding).

The second requirement has been solved by automatically generating "ideal" individuals for each ontological class. A knowledge base call has been programmed so that it identifies an instantiation of each property (attribute or relationship) asserted to the class the user aims to instantiate. The subsequent http request retrieves the class example(cl) including asserted properties to the class (pr) and inferred "component" and "operation" distances (d_{cp}, d_{op}) utilised for content allocation, just as an individual in PMAU.

$$cl = (pr_1, pr_2, \cdots, pr_m, d_{cp}, d_{op}) \tag{5-1}$$
$$m = number\ of\ asserted\ properties\ to\ individual$$

$$pr_{in} = (res_{in,m,na}, res_{in,m,ra}, res_{in,m,ty}, res_{in,m,va}) \tag{5-2}$$
$$res_{in,m} \in URI \cup S \cup \mathbb{R} = uri, string\ or\ number\ that\ a\ property\ instantiates$$

$$d_{cl,c}: C_{in} \times C_c \rightarrow [0, \infty) \tag{5-3}$$
$$c \in \{component, operation\} = class\ to\ which\ individual's\ distance\ is\ measured\ to$$

If a property assertion is not found in the server's knowledge base, then the resultant json object will include no value for that property. Therefore, any "reporting" content format that has a non-null data value rule won't be pairable to that property. This behaviour allows RPMAU algorithm to treat content formats for knowledge transfer and capture similarly. Thus, identifying what file formats are required to record in case a content format allows to do so (e.g. pictures or audio).

The third requirement involves additional server calls to ensure submitted individuals are correct instances of ontology classes including correct assertion in terms of domain and range of attributes and relationships instantiated. The reported individual is sent to the server in the form of a json object. A http post request makes the necessary checks to ensure that the individual only includes properties whose domains are the instantiated class. It ensures that properties' ranges are those declared on the server's ontology schema. If a property fails to comply with these, the individual won't be added to the knowledge base.

The final requirement involves creating programmable content formats whose user interaction gives as result data inputs. Figure 5-4 presents those developed for this research classified by visualisation modes. Their programmed behaviours and facets can be seen at 10.17862/cranfield.rd.12382604. These formats aim to reduce reporting time by easing input visualisation and filtering input information. They use a colour code to show interactive states: blue for selectable, green for selected.

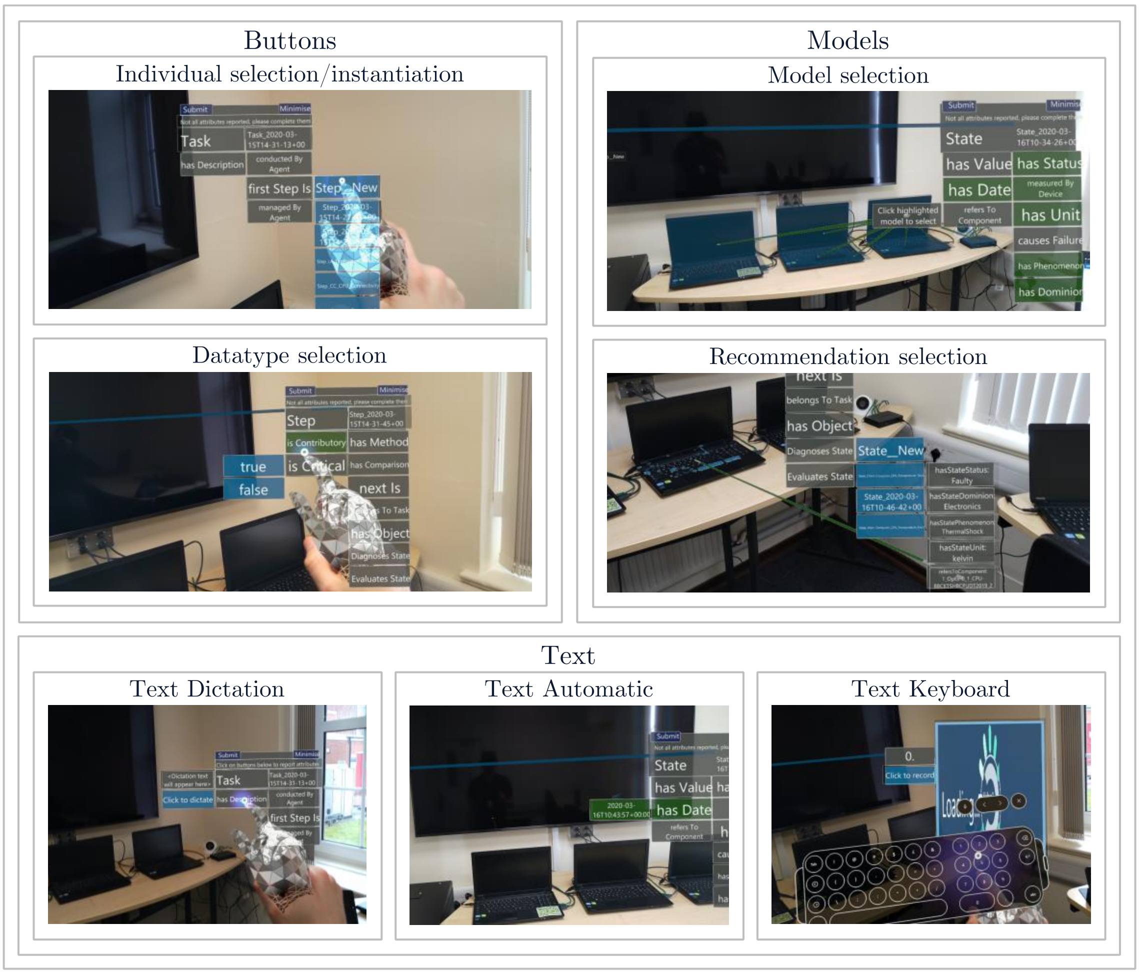


Figure 5-4. Examples of fabrications comprising all formats developed in this research categorised by visualisation mode – These are shown in screenshots from the AR application's resultant prototype. They include a colour code regarding user interaction: blue for selectable, green for selected and grey for menu activation.

Different to PMAU's content formats, RPMAU's content formats require extra features to ensure data inputted through user's interaction is correct according to formats data facets. For programming content formats in RPMAU, developers need to implement data checking rules that comply with assigned properties instantiation rules for their ranges and types. For validation purposes, this research implements content formats that comply with expected properties types and ranges: individuals, numerical values, textual values, dataset values and components. There are other datatypes like images or audio files which could also be considered for creating new content formats. Future works can implement these to extend heterogenous datatypes that can be captured through RPMAU.

Table 5-1 describes RPMAU's content formats and their behaviours implemented in this research for validation purposes. For example, 'Text keyboard' is meant for inputting numerical values. So, when the user introduces a non-numerical character, the format will return an error asking the user to input numerical values only. 'Individual selection' allows to select from all individuals of a certain class by name as well as to instantiate new individuals for continue with the reporting procedure. 'Model selection' allows to select an ontology individual based on model files asserted to it as a datatype property. Its aim is to reduce input time by using a different visualisation mode (model) than just referring to the individual by its name (text). 'Recommendation selection' uses hybrid recommendation methods to pre-filter selectable individuals. Thus, decreasing input time by reducing user's selection options. Moreover, this format also implements diverse visualisation modes (e.g. models) to simplify the selection process. The recommendation framework for improving AR knowledge re-use applications utilised in this content format is presented in Section 5.4.2.

Table 5-1. Description of content formats classified by visualisation and interaction.

Format	Visualise	Interact	Behaviour
Text dictation	Text	Dictate	Utilises MRTK text-to-speech module to capture strings
Text automatic	Text	None	Automatically records data value if property's range is of datetime type
Text keyboard	Text	Tap	Utilises HoloLens 2 default keyboard to input numerical values
Individual selection	Text	Tap	Utilises user's tap gesture to select individual from knowledge base. If new individual selected, then activates new individual instantiation.
Datatype selection	Text	Tap	Utilises user's tap to select a value from a datatype set (e.g. boolean).
Model selection	Model	Manipulate	Utilises user's manipulation to select an individual based on its referred model (obj file).
Recommend selection	Model + Text	Tap	Utilises individuals' properties to filter them according to recommendation algorithm and allow user to select from filtered selection.

5.4.1.2 Context-aware element allocation procedure

PMAU's content allocation assumed that augmented content needs to be overlaid around the equipment. So, maintainers have non-occlusive information visuals from any perspective while operating on the equipment. Nevertheless, this assumption has limitations when equipment's size is bigger than 1.5 meters according to student's testing. In this case, maintainers are required to highly deviate from their focus for consulting augmented information. This can become even more problematic when users are required to move around the asset for reporting purposes. To alleviate this issue, RPMAU proposes to analyse equipment size and dynamically adapt content allocation accordingly. When the equipment is too big ($>$ 1.5 meters), it maintains augmented content at specific distances from the user. Otherwise, it enables for augmented content to still be allocated around the equipment. Thus, enabling users to have good visuals of augmented content at any given situation.

Figure 5-5 describes the proposed dynamic content allocation method. Firstly, it analyses equipment's size measures using equipment's CAD model boundaries. If any of these are bigger than 1.5 meters, then it enables the user-based allocation algorithm. Otherwise, it activates the asset-based allocation algorithm. Secondly, this method allocates newly created elements according to their assigned location (primary, secondary, tertiary and quaternary) calculated from ontological distances similarly to PMAU. In asset-based allocation (Figure 5-5-b), elements are placed at the first-free spot in its location (primary (1), secondary (4) or tertiary (6)). If there is not any free spot, then the allocation algorithm eliminates the earliest element allocated to that location and replaces it with the new one. In user-based allocation (Figure 5-5-a), newest element created is assigned to the primary location (1 spot). Any previous element in such location is moved to the quaternary (6 spots) location and placed in the first free spot available. If there is not any free spot, then the element replaces the earliest one moved to the quaternary location. Overall, both algorithms use a FIFO approach (First-In-First-Out) to allocate elements to spots at any location.

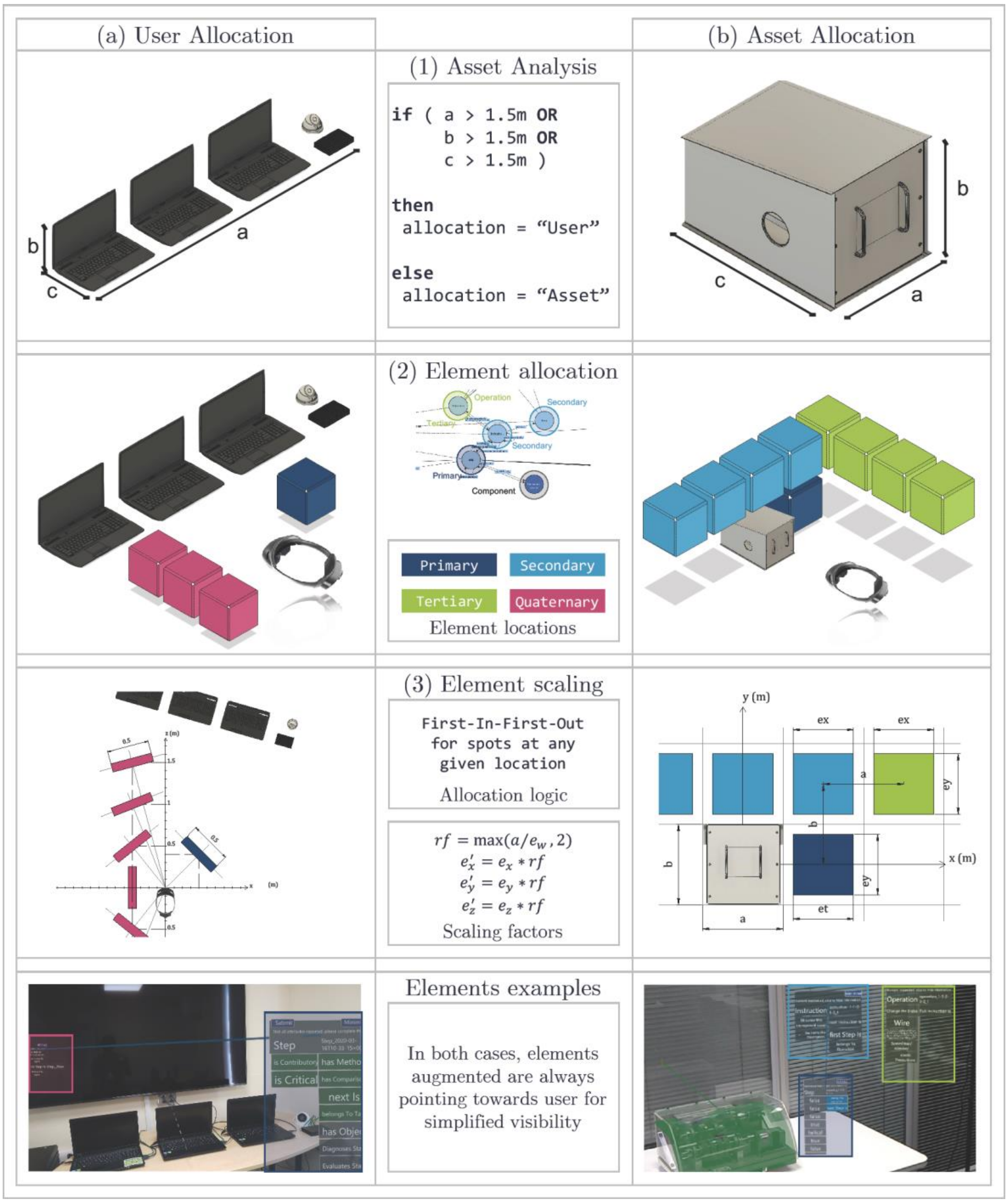


Figure 5-5. Description of dynamic content allocation procedure based on maintained asset’s size analysis – Procedural steps are numbered (1 to 3) to describe the content allocation sequence for user-based (a) and asset-based (b) approaches. Additional formulae is presented at each explain to detail the procedural’s rationale. Examples of resultant allocated elements are displayed at the bottom of the figure.

Because elements are placed at different distances from the user, these need to be re-scaled in some cases. The re-scaling process occurs considering element’s width as re-scaling factor. Other measures (length and height) are re-scaled accordingly. Element's width (e_w) is initially set at 0.5 meters, which is sufficient for non-occlusive visualisation in user-allocation cases (Figure 5-5-a). Therefore, elements' re-scaling only occurs on asset-based allocation cases and it is dependent on equipment's width. If asset's width is bigger than 0.5 meters, then element is re-scaled according to the maximum between 2 and the rate between asset's and element's width ($rf = \max\ (a/e_w\,, 2)$). Besides, elements are embedded additional behaviour for always pointing in user's direction. Thus, allowing for easier visualisation and avoiding occlusive scenarios.

For this method to work, AR requires asset tracking to be enabled. This research's implementation uses Vuforia's model and image target registration and tracking features to do so. Similarly to other methods in this proposal, it uses ontology-related files to load in real-time model or image target tracking and related CAD files (".obj") to allow AR users to select the asset or equipment to be maintained. This also enables other asset-based features of this proposal like component spatial proximity filtering for content-based recommendations.

5.4.2 Hybrid recommendations for Augmented Reality knowledge re-use applications

An important limitation in AR-based knowledge transfer and capture relates to the size that different visualisation formats take to augment maintenance information. AR user's field of view is limited and so is the space available to augment content. In AR knowledge transfer applications, this issue can be solved by accurate information management. For example, if there is a maintenance task that comprises various steps, this issue can be solved by linking steps in order rather than connecting all to the main task. Nevertheless, it is necessary to use more advanced information filtering methods to solve this in AR knowledge capture applications. For example, if the AR application required to report the step that was done last, then it wouldn't be necessary to augment those steps that were conducted previously. Besides, such filtering techniques can help to further improve efficiency of AR knowledge capture applications. For example, if the AR application was able to determine that the report was describing a repair task, then it could further filter "selectable" steps discarding those that are not repair-related. Overall, context-aware recommendations seem like reasonable methods to improve content adaptiveness in AR knowledge transfer and capture applications.

For realising automatic adaptive authoring in AR knowledge re-use applications, this chapter proposes a recommendation framework to enable content formats to filter the information they augment. This framework extends the concept of facets presented in PMAU (Subsection 4.4) to include recommendation algorithms as part of content formats in a generic manner. So, AR developers can create formats which can filter the information they augment. The framework aims to describe hybrid content-based recommendation methods: context-aware and ontology-based. So, content formats can apply them automatically without re-programming coded visualisation and interaction behaviours. Recommendation facets (rf) are the features declared by content formats that enable them to filter augmented information (selectable ontology individuals). These facets declare the instantiable property $(na_{pr,ft}, ra_{pr,ft})$ subject to recommendation and the rules (rr) to assess similarity between target and cases. The instantiable property subject to recommendation is declared by its name $(na_{pr,ft})$ and range $(ra_{pr,ft})$ for enabling the algorithm to identify recommendation targets. This instantiable property must be present in a source data facet of the format that declares the recommendation facet.

$$rf_{ft,n} = (na_{pr,ft}, ra_{pr,ft}, rr_{ft,n,1}, \cdots, rr_{ft,n,l}) \quad (5\text{-}4)$$

$t = property\ attributes\ of\ target\ case$

$l = number\ of\ rules\ that\ comprises\ the\ recommendation\ algorithm\ of\ a\ given\ format$

Recommendation rules (rr) comprise the methods implemented in the recommendation facet to assess similarity between individual target and potential cases. These are defined by similarity method's type (rt) implemented, name $(na_{pr,rr})$ and range $(ra_{pr,rr})$ of property evaluated and the semantical weighted features (x_i, s_i) the implemented method utilises. Property name and range declare the property being assessed, which can be different to the property (attribute or relationship) being recommended. This research proposes four different similarity functions: binary, symmetric, discrete and spatial.

$$rr_{rf,l} = (na_{pr,rr}, ra_{pr,rr}, rt_{rr,l}, rs_{rr,l}) \quad (5\text{-}5)$$
$$rt_{rr,l} \in \{binary, symmetric, discrete, spatial\}$$
$$rs_{rr,l} = ((x_1, s_1), \cdots, (x_m, s_m))$$
$$x_i \in \mathbb{R} = real\ numbers;\ \ s_i \in S = strings$$

Binary similarity assessment ($rt = binary$) implements Kronecker's delta function to compare target and case. It applies to any datatype and object property assessed.

$$sim_{rt_{rr,i}=binary}(t,c) = \begin{cases} 1, & va_{pr_c} = va_{pr_t} \\ 0, & va_{pr_c} \neq va_{pr_t} \end{cases} \quad (5\text{-}6)$$
$$sim_{rt_{rr,i}=binary}(t,c) \in [0,1]$$

Symmetric similarity assessment ($rt = symmetric$) implements a triangular function (Smyth, 2007) to numerical-range properties (e.g. xsd:double).

$$sim_{rt_{rr,i}=symmetric}(t,c) = 1 - \frac{va_{pr_t} - va_{pr_c}}{max\ (va_{pr_t}, va_{pr_c})} \quad (5\text{-}7)$$
$$sim_{rt_{rr,i}=symmetric}(t,c) \in [0,1]$$

Discrete similarity assessment ($rt = discrete$) implements an analogous function to symmetric assessment but applicable to string-range and object type properties for semantic evaluation. To do so, it uses weighted semantical features declared by recommendation rules.

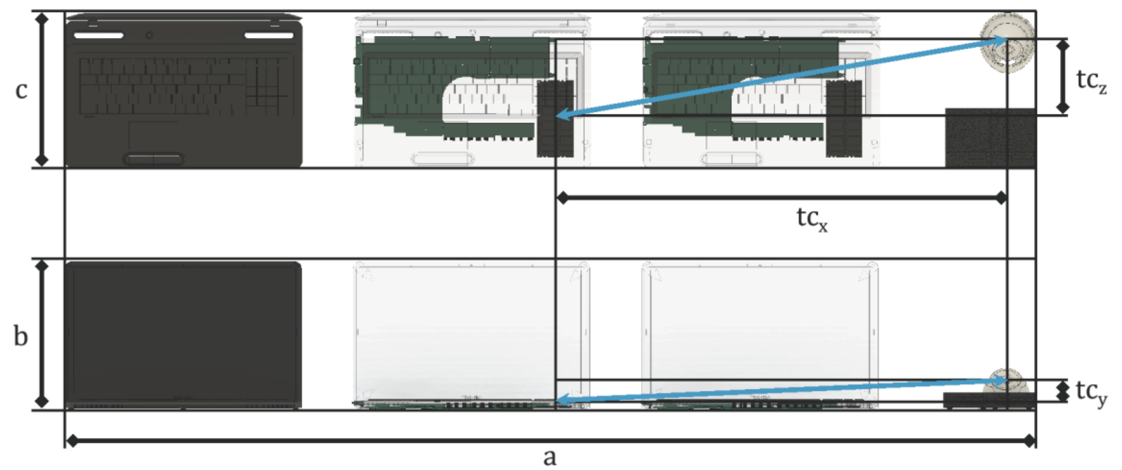
$$sim_{rt_{rr,i}=discrete}(t,c) = \begin{cases} x_i, & s_i \supset va_{pr_c} = va_{pr_t} \\ 0, & otherwise \end{cases} \quad (5\text{-}8)$$
$$\Sigma_m x_i = 1$$
$$sim_{rt_{rr,i}=discrete}(t,c) \in [0,1]$$

Spatial similarity assessment ($rt = spatial$) implements an analogous function to symmetric assessment that aims to evaluate the distance between real-world objects in the augmented space. To do so, it uses context-aware features from AR tracking and registration techniques to measure distances between objects.

$$sim_{rt_{rr,i}=spatial}(t,c) = \begin{cases} 1 - \frac{\overrightarrow{tc}}{|\overrightarrow{av}|}, & va_{pr_c} \in Asset \wedge va_{pr_t} \in Asset \\ 0, & otherwise \end{cases} \quad (5\text{-}9)$$

$$sim_{rt_{rr,i}=spatial}(t,c) \in [0,1]$$

Figure 5-6 presents the formulae declared to implement the spatial similarity assessment function. When the asset to maintain is selected through RPMAU, AR registration can identify its size, position and rotation in real-life. Besides, RPMAU does AR tracking by overlaying the asset's CAD model and scaling to match the registered real-world counterpart. Because the asset's CAD model includes its components, then these can be traced through their names in recommendation cases to measure the distances between them according to the formulae in Figure 5-6. Figure 5-6 also presents an example of a distance calculation between two components of this research's case study.



$$Maximum\ distance\ (Asset\ Volume)\ \equiv |\overrightarrow{av}| = \sqrt[2]{a^2 + b^2 + c^2}$$

$$Target\ case\ distance\ \equiv |\overrightarrow{tc}| = \sqrt[2]{tc_x^2 + tc_y^2 + tc_z^2}$$

Figure 5-6. Formulae for spatial similarity assessment and examples – The figure presents an example to describe formulae used in the spatial similarity assessment function (5-9). It shows how to calculate the distance between target and case components (tc) and the maximum distance (av) considered as the asset volume.

Recommendation rules are assessed through a top-k nearest neighbour strategy ($similarity\ t, c$) by adding similarities calculated from each rule. Based on fabrication's size, number of similar neighbours was set to three ($k = 3$) to avoid augmented content overload in user-allocated scenes. Nevertheless, future research works should study the optimal number of similar neighbours according to recommendations' accuracy.

$$similarity\ t, c\ = \frac{\sum_{i=1..l} w_i * sim_{rt_{rr,i}}(t,c)}{\sum_{i=1..l} w_i}; w_i = 1/l \tag{5-10}$$

Recommendation facets aim to filter object type properties being augmented for instantiation ($na_{pr,rf}, ra_{pr,ft}$) and recommendation rules apply to target (t) and cases (c) individuals. Cases are therefore all individuals in knowledge bases whose asserted class coincides with the range ($ra_{pr,ft}$) of the recommendation facet. Target is the last individual instantiated within the AR knowledge capture application whose asserted class also coincides with the range ($ra_{pr,ft}$) of the recommendation facet. Recommendation facets download cases from knowledge bases and use the proposed algorithm's content tracing feature to identify targets. In RPMAU, recommendation facets are applied once fabrications have been assigned (Figure 5-3). So, they can identify if the properties assigned to their formats match their name and range.

Recommendation similarity strategy and methods aim to evaluate the impact of content-based recommendation techniques in AR knowledge re-use applications for reducing content overload and improving knowledge-capture efficiency. Moreover, semantic and spatial similarity assessment methods have been proposed to take full advantage of ontology-based AR context-awareness features. Nevertheless, this framework proposes a method to generically declare recommendation rules, so other approaches can be implemented to take advantage of other AR proposed features (e.g. content tracing, user performance, etc.). Future works will study alternative recommendation approaches (e.g. content-based diversity, collaborative-filtering) and implement advanced techniques for automatic data collection (content-tracing, eye-tracking) to enhance them.

5.4.3 Application to failure diagnosis reporting

Recommendation facets and rules (Section 5.4.2) have been implemented in different content formats (Section 5.4.1.1) for an AR-maintenance knowledge capture and re-use application in diagnosis reporting. This maintenance operation has been studied previously in Chapter 3 regarding ontology-based reporting of unstructured expert knowledge in failure diagnosis. Figure 5-7 presents the resultant ontology (diagont) developed as part of that research. It describes an asset's *'failure*[4]*'* as a series of connected component's faults from its initial symptom to its root cause, which *'causes'* the *'failure'*. The reporting procedure is therefore seen as a series of *'steps'* that *'evaluate'* *'states'* or conditions of *'components'* in comparison with *'diagnosed'* *'states'*. *'Evaluated'* *'states'* are those that reflect the current conditions of '*components'*. Instead, *'diagnosed'* *'states'* describe conditions of *'components'* that are known to be healthy or

[4] Words between quotation marks refer to diagont's ontology in Figure 5-7

faulty. These comparisons, which can be of different types (e.g. *'equalTo'*, *'greaterThan'*, etc.), allow reporters to determine if *'evaluated' 'states'* are also faults and whether these faults apply to the *'failure'* being reported. To produce these comparisons, 'states' report different quantitative and qualitative measures regarding the component's condition being evaluated. Table 5-2 and Table 3-2 describe diagont's datatype properties (attributes) and datatypes values developed to improve structure and accuracy of reported *'components conditions'* and diagnosis *'steps'*. These include failure diagnosis descriptors like *'components status'* and *'failure domains'* and aim to unify the vocabulary used by maintainers to report diagnosis-related concepts. Additional details on these properties and datatypes can be consulted in Chapter 3. Future works in Chapter 3 included the study of AR and recommendation methods to enhance failure diagnosis reporting operations. This research aims to fulfil that through proposed content formats (Section 5.4.1.1) and recommendation facets (Section 5.4.2).

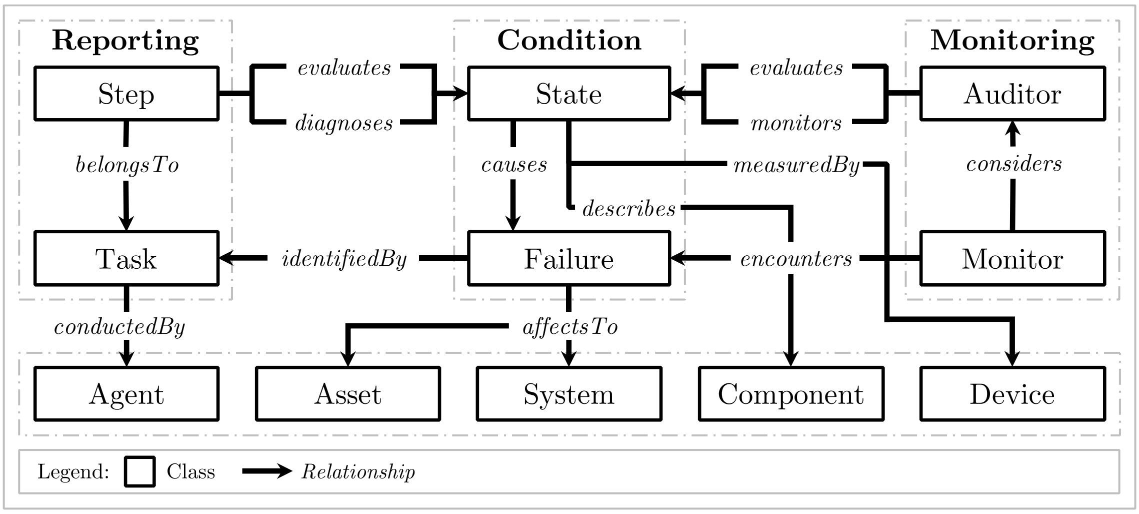


Figure 5-7. Depiction of classes and relationships of diagont's ontology schema – This schema, already presented in Subsection 3.4.1, is used to apply the proposed recommendation formats to failure diagnosis reporting scenarios.

Table 5-2. Depiction of classes and attributes of diagont's ontology schema.

Task		Failure		Monitor	
hasDescription	*string*	hasDescription	*string*	hasDescription	*string*
		hasImpact	*impact*		
		hasDomain	*domain*		
		hasPhenomenon	*phenomenon*		
		hasImage	*anyUri*		
		hasAudio	*anyUri*		
Step		State		Auditor	
isCritical	*boolean*	hasStatus	*status*	isValidated	*boolean*
isContributory	*boolean*	hasDomain	*domain*	hasComparison	*comparison*
hasObject	*object*	hasPhenomenon	*phenomenon*		
hasMethod	*method*	hasMeasureValue	*double*		
hasComparison	*comparison*	hasMeasureUnit	*unit*		
		hasMeasureDate	*date*		
Legend: Class | Attribute | *Datatype*					

Table 5-3. Summary of diagont's proprietary attributes datatypes.

impact	Status	domain	phenomenon		
local	*Normal*	*mechanics*	*Fracture*	*thermal shock*	*signal error*
global	*safely degraded*	*electrics*	*fatigue*	*thermal runaway*	*error*
	unsafely degraded	*electronics*	*corrosion*	*short circuit*	*material*
	Faulty	*hydraulics*	*impact*	*open circuit*	*process*
		pneumatics	*blockage*	*electric loss*	
		humanics			
object	Method	comparison	unit		
symptom	*Inspect*	*equal to*	*metre*	*pascal*	*hertz*
trace	*Measure*	*not equal to*	*degree*	*joule*	*watt*
cause	*Repair*	*greater than*	*kilogram*	*mol*	*ampere*
	Replace	*less than*	*second*	*kelvin*	*volt*
		less than or equal to	*newton*		*ohm*
		greater than or equal to			
Legend: datatype | *datavalue*					

Figure 5-8 presents a relevant fabrication example of a content format with recommendation facets (ModelPanelTap3). According to diagont's reporting proposal, there is one step at which recommendations would improve reporting efficiency. That is the cases of *'diagnosed' 'States'*. The number of *'diagnosed' 'States'* in the ontology-based cloud server is quite big due to the number of states declared by monitoring sensors. Therefore, they required previous filtering to avoid augmented scene's overload. Moreover, recommendations of *'diagnosed' 'States'* would allow to re-use already known components conditions to diagnose the next fault in the failure tree. Figure 5-8 presents the recommendation rules implemented to assess similarity of existing states, which include their *'domain'* (binary), *'phenomenon'* (binary), *'unit'* (binary), *'status'* (subset) and *'component'* (spatial). Besides, Figure 5-8 describes additional inferencing rules applied on the server side to discard all 'states' that do not represent a relevant condition. These discard all 'states' that have not been used in previous reports or monitoring rules as described in Chapter 3.

Figure 5-8 also displays examples of fabrications from formats that implement the recommendation facet described above. They use text and models to identify the component and its condition. Besides this format, there are other that also implement recommendation facets specific for diagont. These can be consulted at 10.17862/cranfield.rd.12382604.

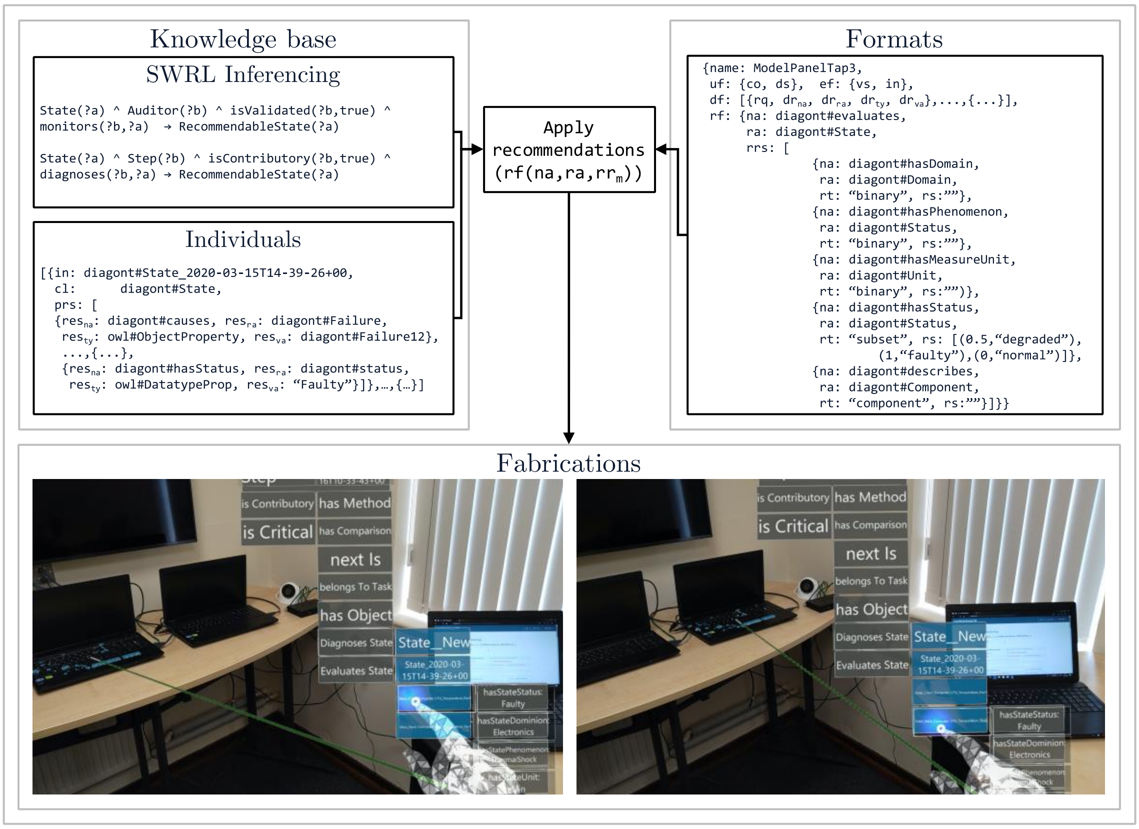


Figure 5-8. Examples of fabrications employing ontology-based context-aware recommendations in fault diagnosis reporting scenarios – The figure exemplifies how automatically-created fabrications produce recommendations using recommendation facets as described above. The inferencing rules proposed are detailed using SWRL notation. The figure also displays a format (ModelPaneltTap3) including recommendation facets (rf – (5-4)) and rules (rr – (5-5)) and a recommendable individual. These are shown as JSON objects to show their data for exemplifying a similarity assessment (5-10). Some resultant fabrications, within the same automatically-created element, are displayed at the bottom of the figure.

5.5 Validation protocol

5.5.1 System implementation

The proposed solutions were implemented within a prototype system for experimentation. This prototype consists of two subsystems: (1) a cloud server for maintenance ontologies storage and (2) a HoloLens 2 AR application. Figure 5-9 presents the languages and platforms utilised to code each subsystem. The cloud server storage (Subsection 3.5.1) uses the graphical database Neo4j (Zhu, Zhou and Shao, 2019) to store maintenance ontologies and Cypher (Panzarino, 2014) and neosemantics (Barrasa, 2019) to support data transfer through OWL and RDFS. Besides, the server incorporates a web-based application coded in EJS (Eernisse, 2015) for maintenance experts to input maintenance data, which has already been described in Chapter 3. And also a service provider built in NodeJS (Surhone, Tennoe and Henssonow, 2010) to transfer ontology data (e.g. classes, individuals, etc.) and related files (e.g. obj, png, etc.) using HTTP requests and JSON objects to the AR application. The HoloLens-based AR application has been coded and deployed using Unity Game Engine (Unity Technologies, 2019) and Visual Studio (Microsoft Corporation, 2019). Content formats and recommendation facets and RPMAU's algorithm have been coded in C# (Hejlsberg, 2011). HoloLens interaction has been enabled with Mixed Reality Toolkit (Microsoft Corporation, 2020). Besides, Vuforia (PTC Corporation, 2020) has been used to enable registration and tracking capabilities in the AR application and has been coded to use the same JSON-based API to transfer from the cloud server necessary ontology-related files like Vuforia's model targets.

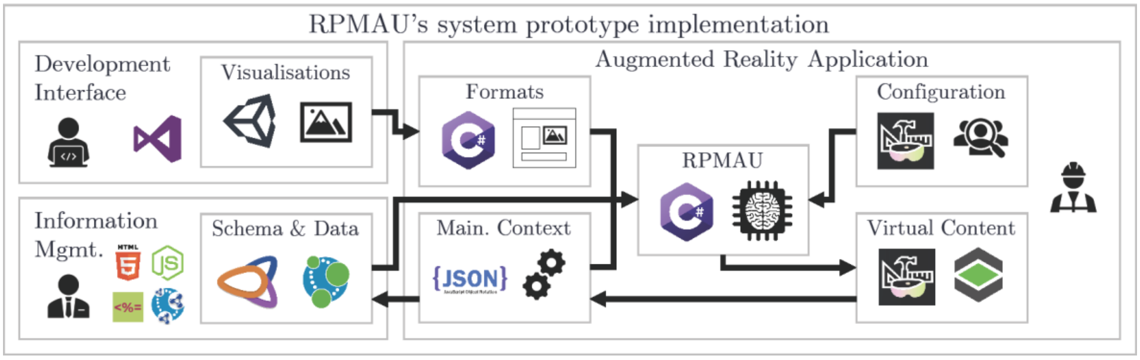


Figure 5-9. Description of the proposal's implementation as a software system prototype – It replicates the structure of Figure 5-1-b to present the logos of each tool/language utilised to develop the system's prototype.

5.5.2 Experiment design

This chapter proposes recommendable content formats and dynamic authoring in for knowledge capture and re-use in AR maintenance applications. These aim to improve efficiency of maintenance reporting applications by (1) enhancing data input through instantiable augmented content and (2) data selection through hybrid (context-aware and ontology-based) recommender facets. These have been implemented for a case of study in diagnosis reporting. Hence, this research's validation aims to evaluate the impact dynamically-authored, recommendable content in the efficiency of diagnosis reporting operations.

Since reporting operations are mostly human tasks, validation of human-computer interaction technologies should focus on their impact on human performance. Common methods in academia to evaluate the impact on human performance measure quantitative and qualitative criteria regarding efficiency (time), effectiveness (errors) and usability (Gimeno *et al.*, 2013; Dey *et al.*, 2018). Workload assessment has also been found useful when human tasks do not solely depend on manual performance, but also in other elements like temporal or mental demand (Chi *et al.*, 2012). Qualitative workload evaluation helps to identify the nature of tasks being performed to contextually analyse efficiency and usability measurements. Besides, this chapter proposes recommendations that can affect human performance in reporting operations and so, their accuracy should also be analysed (Gunawardana and Shani, 2015; Wang *et al.*, 2018).

This research aims to analyse the validity of this research's contributions:

1. Recommendation facets (Subsection 5.4.2) can improve descriptiveness (recommendations accuracy and workload) of fault conditions in diagnosis scenarios for improving reporting efficiency (time).

2. Dynamic authoring and associated content formats (Subsection 5.4.1) can improve human performance (time, errors and usability) in diagnosis reporting operations.

Inspired by similar research (Chi *et al.*, 2012; X Wang, Ong and Nee, 2016a; Wang *et al.*, 2018; Longo, Nicoletti and Padovano, 2019), this research's validation proposes experimental and survey methods to evaluate the abovementioned research contributions. Table 5-4 presents these methods, the criterions they aim to analyse and the objectives for doing so.

Table 5-4. Overview of validation methods, criterions and objectives.

Method	Quantitative	Qualitative	Objective
Experiments	Accuracy		Evaluate the proposal's ability to produce recommendations for identifying faulty conditions in diagnosis scenarios
	Time		Evaluate the proposal's ability to reduce errors for improving effectiveness of diagnosis reporting operations
	Errors		Evaluate the proposal's ability to reduce time for improving efficiency of diagnosis reporting operations
Surveys		Usability	Evaluate the proposal's perceived usability to enhance semantic understanding of diagnosis reporting operations
		Workload	Evaluate perceived workload of diagnosis reporting operations

For these methods and criterions to appropriately evaluate diagnosis reporting effects, the following assumptions must hold true:

- Implemented recommendation facets select component conditions that can be considered faults of the contextual failure. So, accurate recommendations and their selection can affect diagnosis reporting performance through their ability to simplify reporting tasks.
- Reporting operations consist of data input steps. So, errors or incorrectly inputted values can be considered a measure of reporting effectiveness. If the above is true, then time and errors results should not be correlated.
- Reporting operations are human tasks. If errors can be considered a measure of reporting effectiveness, reporting time can be considered a measure of reporting efficiency.
- Reporting tools usability can affect human performance if it is not compatible with reporting tasks requisites like temporal, mental or physical demand.

Validation experiments require additional reporting tools to which compare this research's proposal. These tools or solutions should have different attributes regarding this research's contributions. These attributes refer to recommendations (Subsection 5.4.2) and dynamic augmented content (Subsection 5.4.1). As part of this validation, the student developed the following tools for comparison:

- RPMAU (ARR): this chapter's proposal includes both, augmented content and recommendations, for knowledge capture applications (Section 5.4).
- PMAU (ARN): an alternative AR solution that includes augmented content for knowledge capture applications but not recommendations (Chapter 4).
- Web-based recommendable reporting (TBR): an alternative non-AR solution that only includes recommendations (Figure 5-8) but not augmented content (Chapter 3).
- Web-based reporting (TBN): an alternative non-AR solution that neither includes augmented content nor recommendations (Chapter 3). This was used as baseline for comparative analysis against the other three.

The following subsections describe these methods and their hypotheses as well as the case of study, experimental scenarios and the experimental testers sample.

5.5.2.1 Stopwatch time, errors and accuracy studies

Stopwatch time, errors and accuracy studies aim to analyse the effect of the proposed authoring solution (ARR) on reporting effectiveness and efficiency compared to alternative solutions (ARN, TBR, TBN) in different failure conditions (CNN and TEM). Stopwatch studies consist of testers performing diagnosis reporting procedures regarding different failures, which collect quantitative data regarding the abovementioned criterions. In order to validate this research's contributions (augmented content and recommendations), these studies evaluate the following hypotheses:

1. "Recommendations accuracy improves with AR-based recommendations compared to non-AR recommendations". It means that ARR should be better than TBR in terms of recommendations accuracy.

2. "Reporting errors reduce with the use of augmented content compared to non-AR reporting solutions". It means that ARR and ARN should reduce reporting errors compared to TBR and TBN.

3. "Reporting errors do not vary with the use of recommendations compared to non-recommender reporting solutions". It means that ARR should be similar to ARN and TBR to TBN in terms of reporting errors.

4. "Reporting time decreases with the use of AR content compared to non-AR reporting solutions". It means that ARR and ARN should reduce reporting time compared to TBR and TBN.

5. "Reporting time decreases with the use of recommender methods compared to non-recommender solutions". It means that ARR should reduce reporting time compared to ARN and TBR should do similarly compared to TBN.

Table 5-5 defines the quantitative variables relevant in these studies, while Table 5-6 declares factor variable levels to evaluate validity of these studies' hypotheses. These stopwatch studies have shared features with those from previous Chapters (Subsection 3.5.2.5 and 4.5.2.1). Firstly, they aim to evaluate the impact of one reporting solution (ARR) in comparison to others (ARN, TBR and TBN), similarly to previous experiments. Secondly, they employ the case studies from Chapter 3 because they are focused in diagnosis-related tasks. These case studies refer to the failures that testers need to report as part of the experiments and their reporting attributes are later described in Subsection 5.5.3. Finally, they also do not consider testers' IT expertise level as a relevant factor, similarly to the experiments presented in Chapter 4. The decision was related to the scope of the experiments. They aimed to evaluate the impact of content formats and recommendations in reporting (data input) tasks in terms of time and errors. As part of the experimental protocol (Subsection 5.5.4) testers were briefed on the reporting tasks to conduct (what data to input about the failure occurred). This was to avoid potential reporting errors due to lack of testers' knowledge that would have impacted the results. Thus, testers' expertise level seemed irrelevant as they were to have enough background on tasks to perform.

Table 5-5. Description of measured response and factor variables in reporting stopwatch studies.

Variable	Type	Definition
Time	Response	Number of seconds taken by a tester to complete a diagnosis reporting task of a given failure
Errors	Response	Number of mistakes when inputting failure data made by testers when conducting diagnosis reporting steps
Accuracy	Response	Number of times reporting fault condition is recommended and selected by a tester in a diagnosis reporting step
Failure	Factor	Incorrect asset behaviour that triggers a diagnosis reporting task
Solution	Factor	Reporting tool used by a tester to conduct a diagnosis reporting step

Table 5-6. Description of relevant factors' levels in reporting stopwatch studies.

Factor	Level	Definition
Failure	CNN	Electrical failure to which the experimented asset is setup to for testers to conduct diagnosis reporting procedures
	TEM	Electronic failure to which the experimented asset is setup to for testers to conduct diagnosis reporting procedures
Solution	ARR	Proposed AR solution that includes content formats and recommendations for knowledge capture applications
	ARN	Alternative AR solution that includes content formats for knowledge capture applications but not recommendations
	TBR	Alternative non-AR solution that includes recommendations for knowledge capture applications
	TBN	Alternative non-AR solution that does not include recommendations for knowledge capture applications

Stopwatch studies aimed to test the proposed reporting solution (ARR) against alternative solutions (ARN, TBR and TBN) in diagnosis reporting tasks about two different failures. The experimental procedure, described in Appendix G, consists of three reporting tasks, each one comprising an ontology individual to instantiate (*'Step'*, *'evaluatesState'*, and *'diagnosisState'* – Figure 5-7). Figure 5-10 describes the experimental setup, including the hardware devices embedded with the reporting solutions and the case study asset (Subsection 5.5.3). Figure 5-11 presents examples of reporting procedures using AR and non-AR solutions.

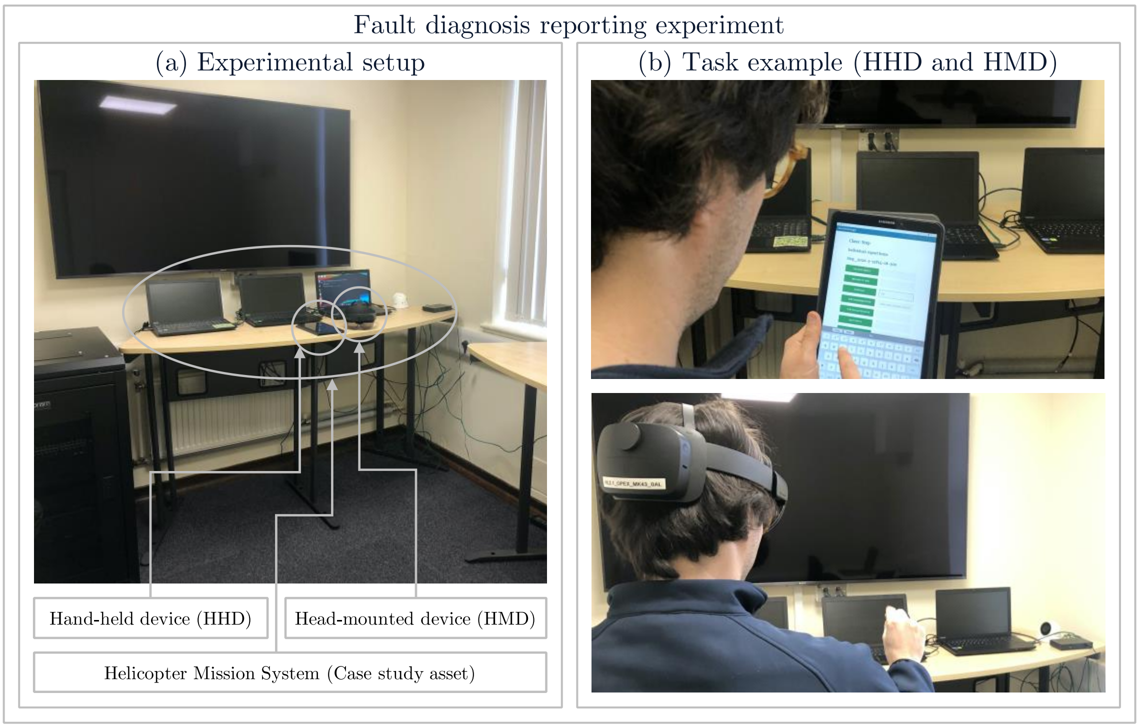


Figure 5-10. Description of diagnosis reporting experiment: (a) experimental setup and (b) task examples with hand-held (HHD) and head-mounted (HMD) devices.

Each experiment consisted of a tester reporting (Figure 5-11) one failure, using one reporting solution. Twenty-eight testers completed two experiments with different solutions, one for each failure. This was for testers to be enabled to compare the usability of different solutions, as they had none or very little with AR and ontology-based reporting (Subsection 5.5.3.3). Testers were randomly allocated in one of four different groups according to experimental factors. Table 5-7 presents these groups according to their failure and solution. These failures (Section 5.5.3) and solutions can be considered sufficiently different in nature to not expect carry-over effects between experiments. So, the experimental design can be considered between-subjects for further statistical analysis.

Table 5-7. Overview of stopwatch experimental groups.

	ARR	ARN	TBR	TBN
CNN	A	C	D	B
TEM	B	D	C	A

Figure 5-11. Failure diagnosis reporting experimental procedure – The figure displays screenshots of (a) AR and (b) non-AR solutions being used at each experimental stage. Each stage aims to report an individual of each class involved in reporting a fault: (1) 'Step', (2) 'evaluatedState' and (3) 'diagnosedState'. Recommendations, different in TBR and ARR solutions, are used in stage 1. If the right 'diagnosedState' is found within recommendations in stage 1, then stage 3 is automatically completed.

5.5.2.2 Workload surveys

Workload surveys aim to evaluate testers perceived performance requisites regarding reporting experiments to contextually analyse experimental time and errors results. In order to evaluate perceived reporting performance, the student employed the NASA Task Load indeX (NASA-TLX) surveys. NASA-TLX is a standard questionnaire developed by NASA Ames Research (Hart, 2006) for collecting workload self-evaluation results from experimental testers. It is a testers' self-rating procedure that provides an overall workload score based on six weighted aspects. Table 5-8 defines these workload factors.

Table 5-8. Description of workload factors employed in NASA-TLX surveys.

Workload factor	Definition
Mental Demand	How much mental and perceptual activity is required (e.g. thinking. deciding, calculating, remembering. looking, searching, etc)? Is the task easy or demanding, simple or complex, exacting or forgiving?
Physical Demand	How much physical activity is required (e.g. pushing, pulling, turning. controlling, activating, etc.)? Is the task easy or demanding, slow or brisk, slack or strenuous restful or laborious?
Temporal Demand	How much time pressure do you feel due to the rate or pace at which the tasks or task elements occurr? Is the pace slow and leisurely or rapid and frantic?
Performance	How successful do you think you are in accomplishing the goals of the task set by the experimenter? How satisfied are you with your performance in accomplishing these goals?
Effort	How hard do you have to work (mentally and physically) to accomplish your level of performance?
Frustration	How insecure, discouraged, irritated. stressed and annoyed versus secure, gratified, content, relaxed and complacent do you feel during the task?

Workload surveys consist of two, self-rating questionnaires, which are presented in Appendix H. Prior to experiments, testers were asked to evaluate the relative importance of each workload aspect given an understanding of the tasks to perform. After the experiments, testers were asked to complete a second questionnaire to evaluate the importance of each aspect independently. These help to provide weighted scores for each workload aspect for contextually analysing experimental results regarding performance effectiveness and efficiency.

5.5.2.3 Usability surveys

After conducting stopwatch experiments, testers were asked to complete a survey regarding the usability of the reporting solutions utilised. Usability surveys aim to evaluate the perceived validity of the proposed tool (ARR) to report diagnosis information compared to alternative solutions (ARN, TBR and TBN). Usability refers to a reporting tool's ability to submit correct information regarding the diagnosis operation being conducted. Usability is a feature perceived by users and subject to opinion and so, its evaluation requires qualitative criteria. As with previous surveys in this PhD thesis (Subsection 4.5.2.2), testers spent considerable time trying reporting solutions. Therefore, it was relevant to study in detail their attributes with impact in usability (Nielsen, 1993). In order to enhance survey's replicability, the student employed the same attributes as previous researches (Gimeno *et al.*, 2013; Lee and Lee, 2016; Dey *et al.*, 2018) concerning well-established usability criterions proposed by Nielsen (1993). Table 5-9 defines these criterions and Table 5-10 the solutions' aspects they comprise.

Table 5-9. Description of usability criterions employed to evaluate reporting solutions inspired by Nielsen (1993) and Lee and Lee (2016).

Criterion	Definition
Ease-to-learn	Ability of the reporting solution to show its functionality by itself
Ease-to-use	Ability of the reporting solution to be self-understandable
Effectiveness	Ability of the reporting solution to support a maintenance operation
Satisfaction	Tester's overall impression of the reporting solution after using it

Usability surveys consisted of a separate section for each criterion including for each aspect regarding the reporting solutions tested in stopwatch experiments. The survey's questionnaire can be found in Appendix I. Testers were asked to declare their agreement with these statements in a Likert Scale from 1 to 5. The reason to select such scale is based on the results presented by Weijters, Cabooter and Schillewaert (2010), who suggested that it maximises potential information transmission when surveying non-expert populations (Subsection 5.5.3.3). Results collected served to evaluate the proposed contributions' usability compared to other specific approaches.

Table 5-10. Description of AR solutions' aspects considered for each usability criterion based on those presented by Gimeno et al. (2013) and Dey et al. (2018).

Criterion	Aspect	Scale
Ease-to-learn	Start, Finish, Intuitiveness	Likert 1-5
Ease-to-use	Buttons or Gestures, Keyboard or Dictation, Text	Likert 1-5
Effectiveness	Efficiency, Confidence	Likert 1-5
Satisfaction	Design, Feeling, Overall	Likert 1-5

Protocols to collect and analyse experimental and survey data are described in Section 5.5.4. Instead, the following section presents the experimental cases of study along with the testing population's sample.

5.5.3 Cases of study

This case of study is based on reporting tests discussed previously in this thesis (Section 3.6.3). It comprises two diagnosis reporting tasks (electric and electronic) of no-fault-found scenarios in a complex-engineering asset. The reason to select no-fault-found scenarios was to reflect in validation experiments the complexity of reporting tasks. Figure 5-12 presents a picture of the case study's asset. The asset, named Helicopter Mission System (HMS), is a replica of an electronic system whose aim is to control the navigation mission of a helicopter. This replica was built with the same specifications as the original in order to enable laboratory experimentation. This system comprises three computers, one camera and an ethernet switch to connect them altogether. The first computer, called 'main mission computer', is used as controller for the rest of the elements and also controls the navigational parameters of the helicopter. The second computer, or 'client mission computer', is that used by helicopter pilots set the navigation mission. The third computer, which acts merely as a 'monitor', is that from which pilots control the ‘client mission computer’. The ‘camera’ is there to provide pilots with a visual of the terrain while handling the helicopter. Finally, the ‘ethernet switch’ aims to connect the main mission computer to the client mission computer, the ‘monitor’ and the ‘camera’ for further control. The system comes with an integrated control monitoring system that evaluates electronic performance parameters from its different elements. Due to its criticality for piloting the helicopter, real-life maintainers are very careful when reporting diagnosis procedures on it. Besides, the control monitoring has some limitations regarding the electronic parameters it can control due to system's configuration.

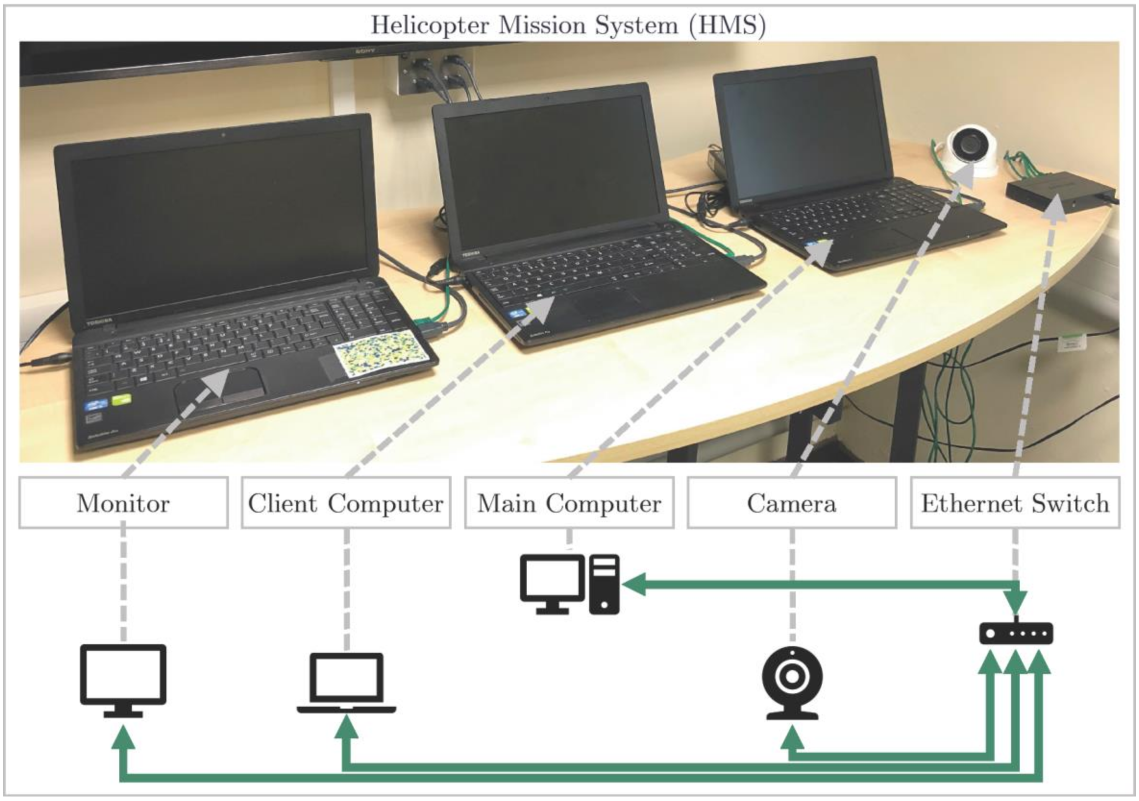


Figure 5-12. Overview of diagnosis reporting case of study: Helicopter Mission System (HMS) – It displays a picture of the HMS prototype at the top along with a schema of its components at the bottom.

These system's complexities make them suitable for this research's experimental methods as they make this systems prone to incur in no-fault-found scenarios, which are difficult to diagnose and report according to expert testers (Subsection 3.6.3). Moreover, these allowed to identify two common no-fault-found scenarios to be used as experimental procedures (Appendix G). The following subsections explains these failures and their reporting procedures.

5.5.3.1 Electric failure reporting scenario (CNN)

Figure 5-13 describes the most mentioned electronic failure in reporting tests (Subsection 3.6.3) regarding the HMS. The failure is caused by a fault in the cable that connects the "main computer" with the "ethernet switch". This failure provokes a no-fault-found condition because that cable cannot be monitored by the system's control module. The control module is managed by the "main computer" and one of its limitations is that it cannot evaluate its own connectivity to the "ethernet switch". When this disconnection occurs, the control module shows the rest of the system's components ("client computer", "monitor" and "camera") as disconnected even though the connectors are in good condition. As shown in Figure 5-13, connectivity is measured through time, which is identified as zero by the control module when disconnected. The experimental reporting tasks are also shown in Figure 5-13. Testers were asked to report the failure's root cause (blue), including a 'Step' and its 'evaluated' and 'diagnosed' 'States'.

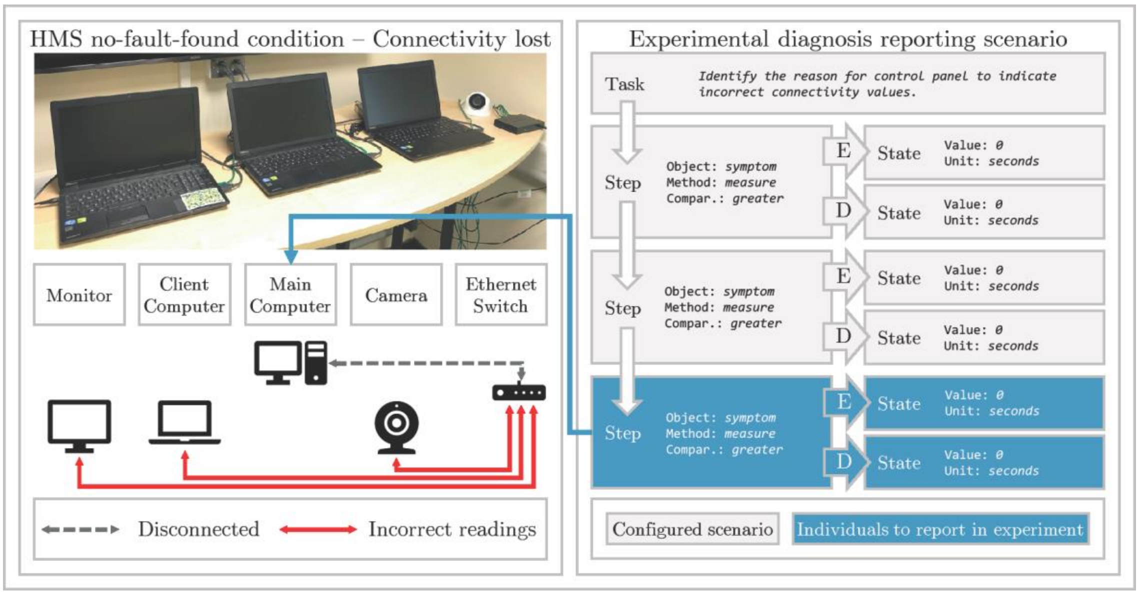


Figure 5-13. First case of study diagnosis reporting experiment: computers' connectivity (CNN) – The figure includes a graphical representation of the system components (left-side) as well as diagont's classes and object and datatype properties

assertions pre-configured (grey – used by the recommender method) and to be reported (blue – by the tester) within the experiment as described in Appendix G.

5.5.3.2 Electronic failure reporting scenario (TEM)

Figure 5-14 describes the most mentioned electric failure in expert interviews regarding the HMS. It is another example of no-fault-found because the thresholds established by the system's control module are higher than the values that cause the failure. The failure consists of a hardware overload caused by too many software applications being run in the HMS simultaneously. When this occurs, both "main computer" and "client computer" reach CPU temperatures higher than 60° Celsius (~333 Kelvin). However, the system's control module cannot detect this issue because the CPU temperatures monitoring thresholds are set for each CPU independently at a temperature of 90° Celsius. Similarly to previous failures, Figure 5-14 shows a simplified version of the failure report including the steps to be reported by experimental testers (blue).

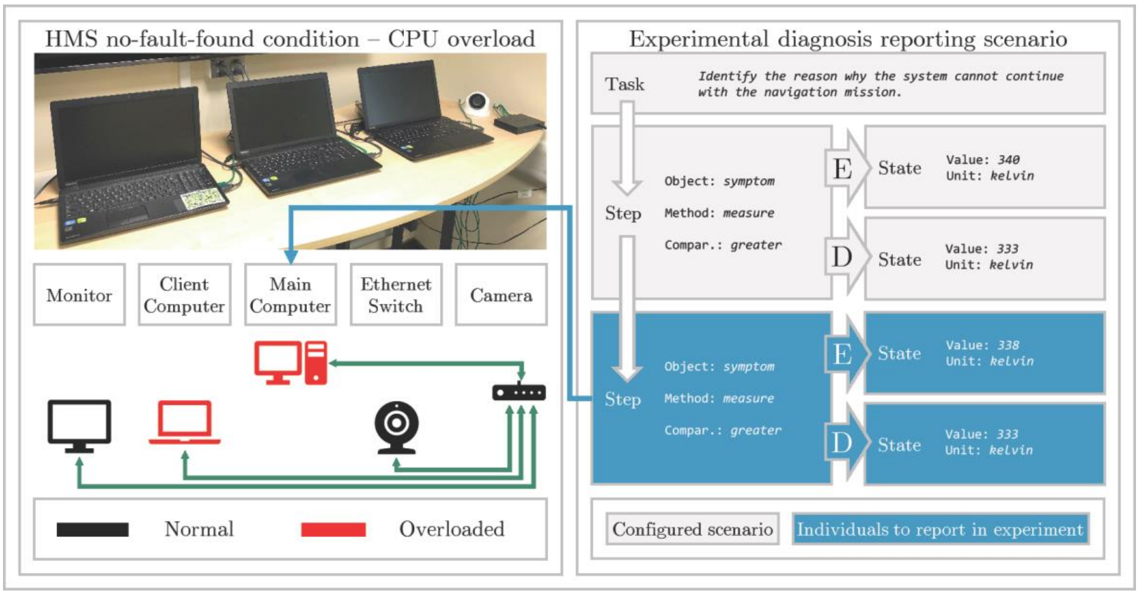


Figure 5-14. Second case of study diagnosis reporting experiment: CPUs temperature overload (TEM) – The figure includes a graphical representation of the system components (left-side) as well as diagont's classes and object and datatype properties assertions pre-configured (grey – used by the recommender method) and to be reported (blue – by the tester) within the experiment as described in Appendix G.

5.5.3.3 Experimental sample

A total of 28 MSc students (19 males and 9 females) participated as testers in laboratory experiments. Their ages range from 21 to 30 years and they are all enrolled in engineering-related MSc degrees. Although they have some basic knowledge in AR, ontology-based methods and maintenance due to their courses, they have no previous hands-on experience in any of them. So, they were given a short training on AR devices right before experimentation to avoid the presence of any learning curves. Testers were randomly allocated to one of the four groups (A (7), B (7), C (7) or D (7) – Table 5-7) to avoid "carry-over" effects between failures while using two different reporting solutions.

5.5.4 Experimental protocol

The protocol comprises the steps to collect and analyse experimental and survey data for validating this research proposal against its expected contributions. The abovementioned validation methods in the case study contexts described above. The following list summarises this protocol:

1. Data collection (28 testers per experiment):

a. AR-maintenance introduction: to briefly train testers on the purpose of experiments, the use of reporting solutions and the experimental failures to report.

b. Stopwatch time, errors and accuracy experiments: to capture quantitative data on the effect on effectiveness and efficiency of reporting in different diagnosis reporting operations.

c. Usability and Workload surveys: to capture qualitative data on tester's opinions regarding usability of reporting tools and workload of reporting operations.

2. Data analysis (30 testers per analysis):

a. Recommendations accuracy study: to evaluate the correctness of recommendations for describing fault conditions in reporting diagnosis and its impact in diagnosis reporting. Results should reflect that there is a significant difference in recommendations accuracy between hybrid (ARR) and ontology-based (TBR) recommender methods. Graphical analysis and t-tests (between-subjects) will be used for this matter.

b. Errors effect study: to evaluate the effect on reporting effectiveness of recommender and AR methods in failure reporting procedures. Results should reflect a significant difference in reporting errors between AR (ARN, ARR) and non-AR (TBN, TBR) reporting tools. They should also reflect a significant difference between recommendable (ARR,TBR) and non-recommendable (ARN,TBN) solutions. Due to the number of experimental factors (Failure, Solution), a two-way ANOVA between-subjects test will be used to test these hypotheses. Additional post hoc comparisons (TukeyHSD test) will help to further analyse interactions between factors.

c. Time effect study: to evaluate the effect on reporting efficiency of recommender and AR methods in failure reporting procedures. Results should reflect that AR (ARR,ARN) solutions' results are significantly different to non-AR (TBN,TBR) results. There should also be significant differences in AR and non-AR solutions between those that implement (ARR,TBR) and do not implement (ARN,TBN) recommendations. Similarly to errors study, two-way ANOVA between-subjects and TukeyHSD tests will be used to validate these hypotheses. Besides, Pearson's coefficient will be used to evaluate the correlation between time and errors results for testing the assumption that errors measure effectiveness and time measures efficiency.

d. Workload study: to quantitatively evaluate the relevancy of workload requisites in diagnosis reporting procedures. Results should help to contextualise previous experimental results analyses and refute the differences between time and errors. Basic statistics and graphical analyses will be used to analyse workload results.

e. Usability study: to quantitatively evaluate testers' opinions on reporting tools' usability. Results should reflect improved usability for those tools that implement recommendations and AR content (ARR,ARN) to indicate validity of effectiveness and efficiency improvements previously tested. Basic statistics and graphical analyses will be used for this matter.

This experimental protocol aims to validate this research's proposed methods against its expected contributions. For this validation to be coherent, there are few assumptions to consider:

- In order to keep consistency within experiments, these were conducted in a laboratory environment to maintain constant other factors (e.g. ergonomics or lighting conditions) that may affect the results. Hence, these factors were considered out of these experiments' scope.
- Experimental sample size for the abovementioned statistical tests can be estimated "a priori". Such estimation can be done using a F test for the most requiring analytical test (two-way ANOVA between-subjects). With 4 factor groups (failure and solution), a variance of 0.3 (partial eta squared), a type-I error of 0.1 (alpha) and a power of 0.9 (1 – beta), the resultant sample size is 31 people. That is quite close to the 28-sample size used in these experiments. Besides, these numbers are similar to those achieved by similar researches (Gimeno *et al.*, 2013; Longo, Nicoletti and Padovano, 2019) (30-sample size).
- As described above, testers are MSc students with none or very little experience in AR or maintenance. Although this ensures a baseline for measuring reporting effectiveness and efficiency, further experiments should be required to corroborate laboratory results in real-life working conditions with real maintainers to ensure validity of these hypotheses.

This protocol's results are discussed in the following section.

5.6 Results

This section aims to discuss experimental results regarding the validity of research hypotheses presented in Section 5.5.2. Its results are analysed and discussed in the following subsections to evaluate this research's hypothesis validity. The complete datasets and analyses can be found at 10.17862/cranfield.rd.12382604.

5.6.1 Recommendations accuracy study

The aim of this study was to validate the following hypothesis: *"Recommendations accuracy improves with AR-based recommendations compared to non-AR recommendations"*. Recommendations accuracy is defined as the number of times recommender methods suggested failures' root causes and these were reported by testers. Stopwatch experiments consisted of testers completing diagnosis reporting procedures about two different failures (CNN and TEM – Subsection 5.5.3) using different reporting tools. Since stopwatch experiments only offered one recommendation per experiment (Figure 5-10), recommendation accuracy values can range between 0 and 1. Two of the experimented reporting tools were AR (ARR) and non-AR (TBR) recommender solutions. According to this study's aim, this chapter's proposal (ARR) was hypothesised to provide higher accuracy than TBR.

Figure 5-15 presents average recommendation accuracy results group by solution and failure factors. These results suggest a considerable difference in accuracy between ARR and TBR for both, electric (CNN) and electronic (TEM) failures. In electric failure experiments, recommendations accuracy was almost double for ARR (0.857) compared to TBR (0.429). In electronic failure experiments, accuracy was 2.4 times better for ARR (0.714) compared to TBR (0.286). These differences can be considered statistically significant (p-value = 0.022) with a confidence interval of 95% (p-value < 0.05) according to the results of t-test (between-subjects) presented in Table 5-11.

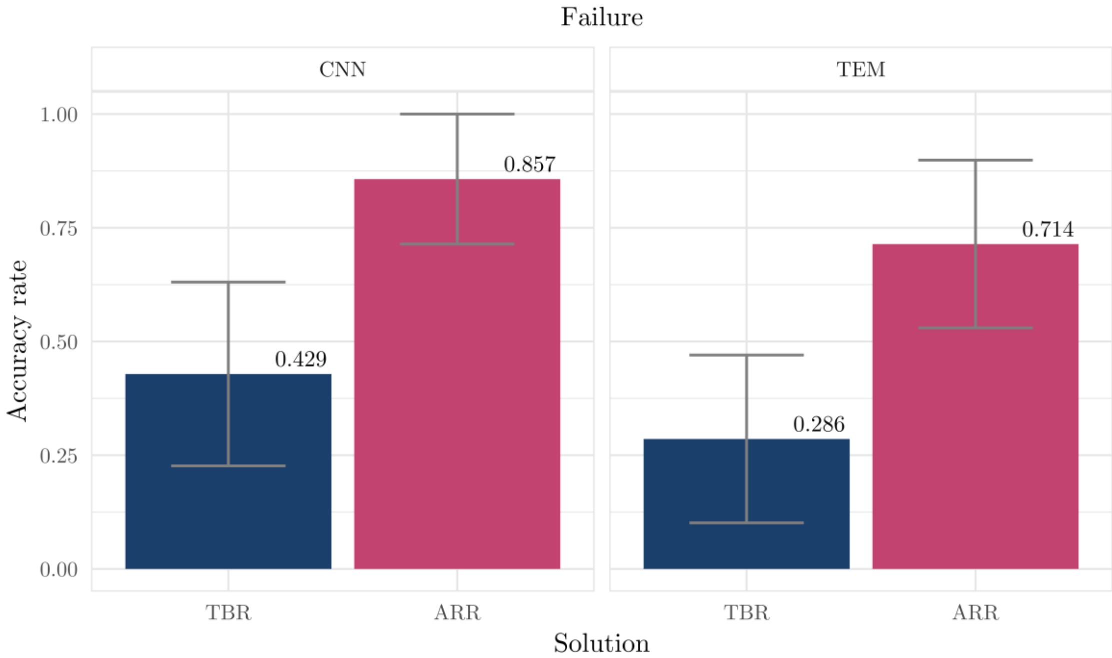


Figure 5-15. Average recommendations accuracy for TBR and ARR solutions per experimental failure, with the standard deviation represented by error bars..

Table 5-11. Between-subjects t-test results on recommendations accuracy variance per solution.

Solution	Testers	Mean	Std. dev.	t	df	p-value	Significant (95%)
ARR	14	0.786	0.426	2.449	25.39	0.0215	Yes
TBR	14	0.357	0.497				

Overall, recommendations accuracy analyses indicate that the proposed AR-based content-aware recommender (ARR) has improved accuracy compared to a more conventional ontology-based (TBR) recommender. Thus, demonstrating the validity of this study's hypothesis. Based on the definition of recommendations accuracy, there are two complementary explanations for these results. First, AR-based hybrid recommendations (ARR) are more precise and provide correct suggestions more often than conventional ontology-based tools (TBR). Second, augmented content formats provide easier visualisation of recommended fault conditions allowing the tester to choose correctly more often. Future works can investigate the independent effect of each cause through experimentation in real-life conditions.

5.6.2 Errors effect study

The aim of the errors effect study was to validate the two following hypotheses:

- *"Reporting errors reduce with the use of augmented content compared to non-AR reporting solutions".*
- *"Reporting errors do not vary with the use of recommendations compared to non-recommender solutions".*

Stopwatch experiments also counted reporting errors, which aimed to measure reporting effectiveness as the number of testers' mistakes in data input tasks. The number of data input tasks varies with the ontology classes being instantiated at each experiment (Appendix G). Hence, errors rates are the percentage of errors by total number of data input tasks per reporting experiment. According to the hypotheses above, errors rates are expected to decrease with AR (ARR, ARN) compared to non-AR (TBR, TBN) reporting solutions but be similar among recommender and non-recommender (ARR vs ARN and TBR vs TBN) solutions.

Figure 5-16 displays average errors rates per experimental solution and failure. These results indicate a considerable difference between AR (ARN, ARR) and non-AR (TBN, TBR) solutions but no relevant effect of recommendations (ARR vs ARN, and TBR vs TBN) in terms of reporting errors. For both experimental failures, average errors rates vary similarly for each solution. Non-AR reporting tools had errors rates ranging between 19%-21%, while AR-based reporting methods achieved smaller errors rates ranging between 10%-12%. Besides, recommender solutions (ARR, TBR) had slightly higher errors rates compared to their non-recommender counterparts (ARN, TBN). This can be caused due to the impact of recommendations on the number data input tasks, which decrease with the use of recommendations.

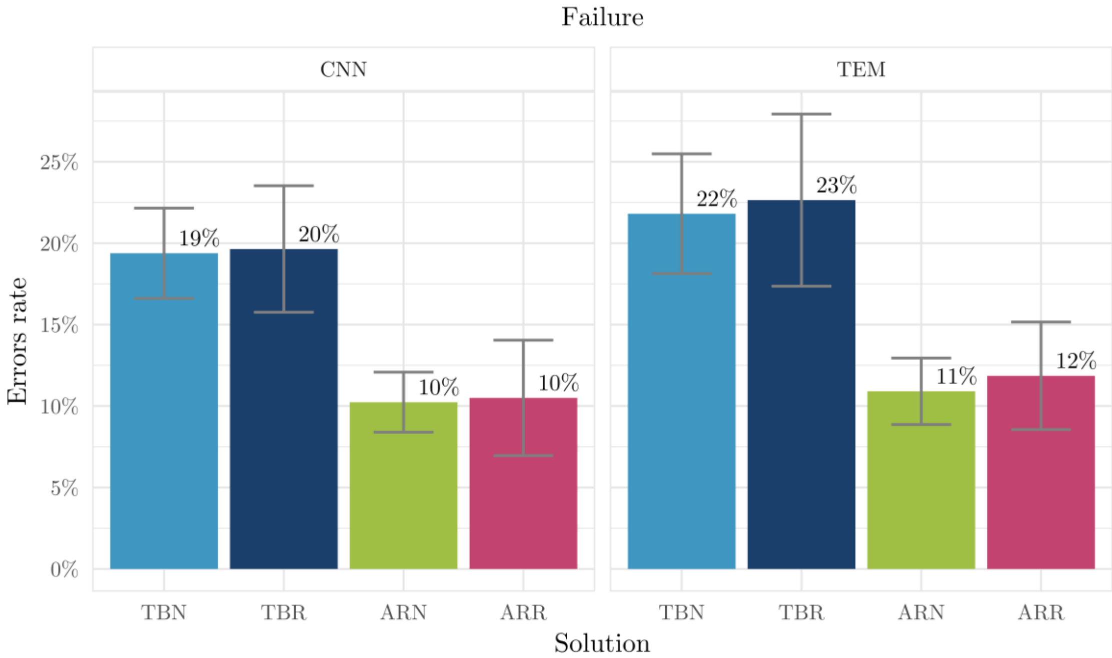


Figure 5-16. Average reporting errors rates for TBN, TRB, ARN and ARR solutions per experimental failure, with the standard deviation represented by error bars.

Further analyses can identify the significance of the differences discussed above. Table 5-12 presents a two-way ANOVA between-subjects test conducted to analyse errors rates variance with failure and solution experimental factors. According to its results, it can be said with a confidence interval of 95% (p-value < 0.05), that 'solution' is a significant factor (p-value $= 0.0019$) while 'failure' and their interaction (solution-failure) are not. Moreover, post-hoc comparisons results from the Tukey HSD test (Table 5-13) can help to evaluate differences between solutions on errors rates results. These indicate that ARR and ARN errors rates are significantly different (p-value < 0.05) to TBR and TBN. They also show no significant differences between ARR and ARN, and TBR and TBN.

Table 5-12. Two-way ANOVA between-subjects test results on errors rates for failure and solution factors.

Factor	Df	Sum Sq	Mean	F Value	Pr (>F)	Significant (95% ci)
Failure	1	0.0020	0.00015	0.010	0.9215	No
Solution	3	0.2439	0.08131	5.263	0.0019	Yes
Failure:Solution	3	0.0010	0.00035	0.022	0.9954	No
Residuals	122	1.8848	0.01545			

Table 5-13. Significance (p-value) results on post hoc comparisons (Tukey HSD) between solution factors in errors rates results.

	TBN	TBR	ARN	ARR
TBN	----	0.99996	0.00779	0.05976
TBR	0.99996	----	0.01886	0.09085
ARN	0.00779	0.01886	----	0.98113
ARR	0.05976	0.09085	0.98113	----

This studies' hypotheses assume reporting errors and time to measure respectively effectiveness and efficiency. For this assumption to be true, these variables should not be correlated. Figure 5-17 presents Pearson's correlation test results. With a confidence interval of 95% (p-value < 0.05), it can be said that the correlation is not significant (p-value $= 0.53$). So, the assumption regarding errors as effectiveness measure and time as efficiency measure can be considered valid.

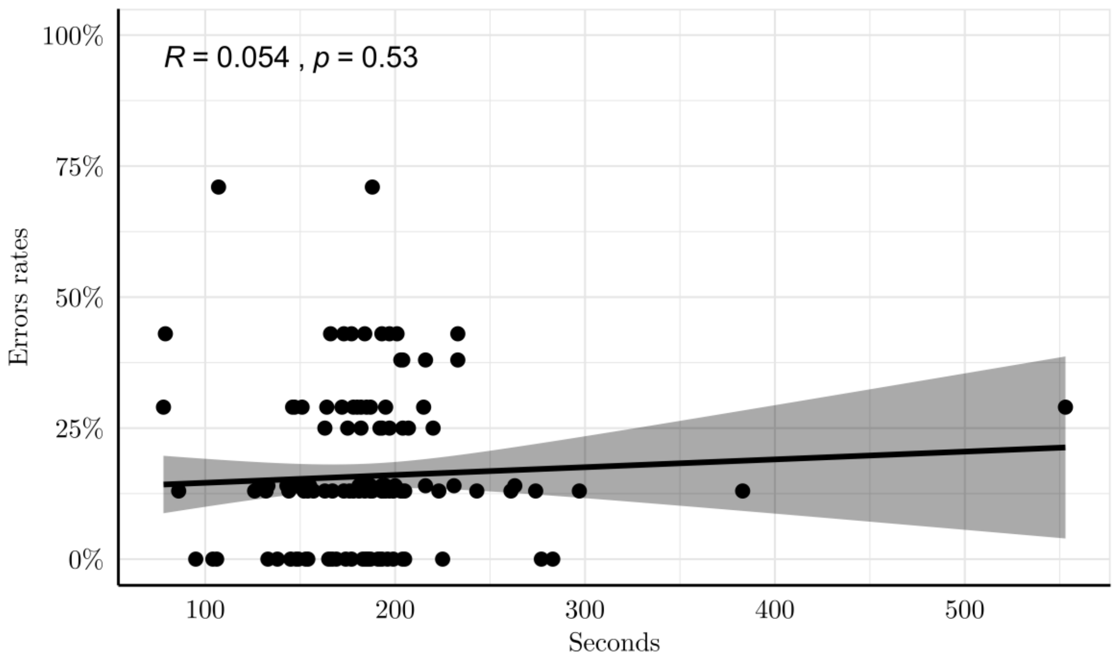


Figure 5-17. Scatter plot displaying error rates versus completion time in seconds to show their correlation, with Pearson's correlation test results at the top left.

Overall, previous discussions support validity of the following considerations regarding the effect on completion errors of AR-based and recommendable reporting solutions:

- Correlation between reporting errors and times results cannot be considered significant.
- Errors can be considered a measure of reporting effectiveness and time a measure of reporting efficiency.
- Differences in errors rates among experimental failures cannot be considered statistically significant.
- Errors rates for AR reporting tools (ARR, ARN) are half of those from non-AR reporting solutions (TBR, TBN).
- Differences in errors rates between AR and non-AR reporting solutions can be considered statistically significant.
- Errors rates for recommender tools (ARR, TBR) are 10% higher than those from non-recommender reporting solutions (ARN, TBN).
- Differences in errors rates between recommender and non-recommender solutions cannot be considered statistically significant.
- Increase on errors rates for recommender tools can be caused due to the impact of recommendations in the number of data input tasks.

These results indicate the validity of this research's hypotheses regarding the effect of AR content and recommenders in reporting effectiveness. AR content formats for knowledge capture enabled data input methods with embedded correctness-checking features, which can have a positive effect on data input mistakes. Instead, recommenders reduce the number of data input tasks and so, their effect cannot be considered significant in reducing their errors.

5.6.3 Time effect study

The aim of the time effect study was to validate the two following hypotheses:

- *"Reporting time decreases with the use of AR content compared to non-AR reporting solutions".*

- *"Reporting time decreases with the use of recommendations compared to non-recommender solutions".*

Stopwatch experiments measured reporting time to evaluate reporting efficiency. Time is defined as the number of seconds taken by a tester to complete reporting tasks regarding the diagnosis of a failure's root cause (CNN and TEM). Testers employed diverse reporting solutions (ARR, ARN, TBRN and TBN – Subsection 5.5.2) to accomplish these tasks. According to this study's hypothesis, this chapter's proposal (ARR) is expected to obtain lowest reporting times, TBN the highest and TBR and ARN in between the other two.

Figure 5-18 presents average reporting times per experimental solution and failure. These results show considerable differences between the proposed solution (ARR) and non-AR (TBR) and non-recommendable (ARN, TBN) alternatives in both failure experiments. In electric failure experiments (CNN), ARR testers were 45 seconds faster than TBN's, while ARN's and TBR's were 10 and 4 seconds faster than TBN testers. In electronic failure experiments (TEM), results also show similar differences. ARR testers are 61 seconds faster than TBN, while ARN's are 22 and TBR's are 11 seconds faster than TBN testers. These results indicate that both AR content and recommendations have a positive impact on reporting times, according to the comparison of ARN and TBR against TBN results. Moreover, these results suggest that the combined effect of both, AR content and recommendations, has a greater impact than each independently, according to the comparison of ARR against ARN and TBR results.

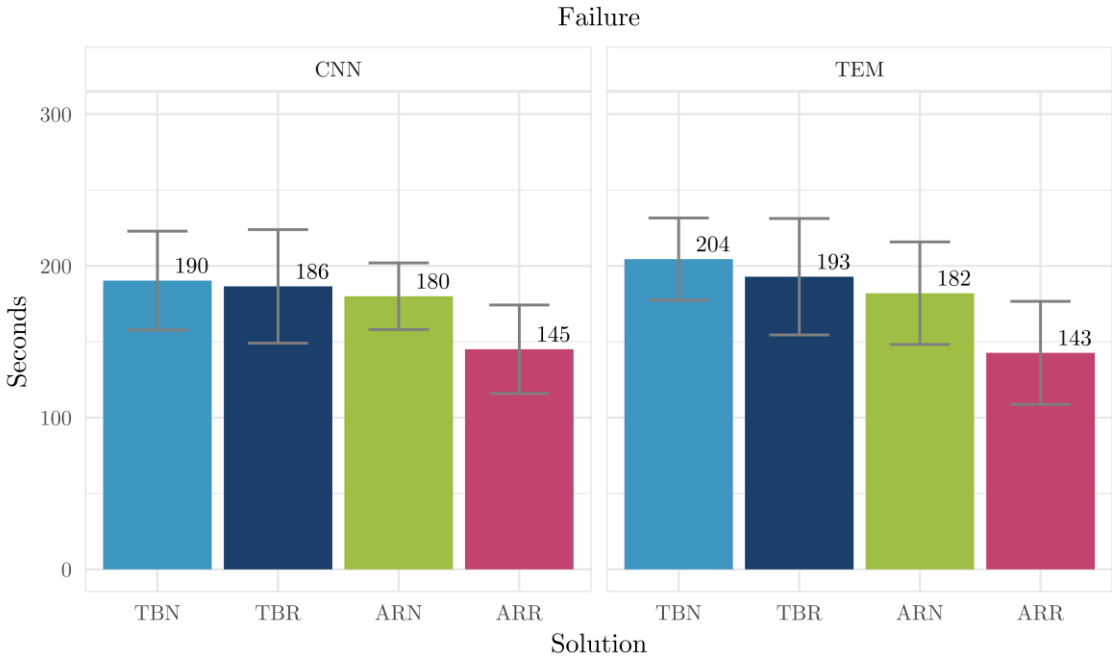


Figure 5-18. Average reporting times in seconds for TBN, TBR, ARN and ARR solutions per experimental failure, with the standard deviation represented by error bars.

Further analyses can identify the statistical significance of differences discussed above. Table 5-14 presents the results of a two-way ANOVA between-subjects test conducted to analyse reporting time variances with failure and solution experimental factors. According to its results, the solution factor effect is statistically significant (p-value = 4.54e-07) while the failure factor cannot be considered significant (p-value = 0.08) for a confidence interval of 95% (p-value < 0.05). Moreover, post hoc comparisons results from the Tukey HSD test (Table 5-15) can help to evaluate differences between solutions on reporting times per experimental failure. Comparison results indicate that ARR is significantly different to other alternatives in electric failure experiments, and significantly different to non-AR alternative solutions (TBN and TBR) in electronic failure experiments. Besides, the non-recommender AR solution (ARN) is not significantly different on reporting times achieved to non-AR alternatives in any experiment. Hence, it can be said that AR achieves improved reporting times when implementing both, AR content and context-aware recommendations.

Table 5-14. Two-way ANOVA between-subjects test results on times for failure and solution factors.

Factor	Df	Sum Sq	Mean	F Value	Pr (>F)	Significant (95% ci)
Failure	1	006072	06072	02.964	8.75e-02	Yes
Solution	3	074638	24879	12.145	4.54e-07	Yes
Failure:Solution	3	010500	03500	01.708	1.68e-01	No
Residuals	132	270415	02049			

Table 5-15. Significance (p-value) results on post hoc comparisons (Tukey HSD) between solution and failure factors groups in time results

	CNN:ARN	CNN:ARR	CNN:TBN	CNN:TBR	TEM:ARN	TEM:ARR	TEM:TBN	TEM:TBR
CNN:ARN	---	0.33708	0.99571	0.99990	1.00000	0.25393	0.01221	0.99151
CNN:ARR	0.33708	---	0.08181	0.23972	0.26759	1.00000	0.00001	0.10554
CNN:TBN	0.99571	0.08181	---	1.00000	0.99891	0.05421	0.10068	1.00000
CNN:TBR	0.99990	0.23972	1.00000	---	0.99999	0.17890	0.11494	0.99995
TEM:ARN	1.00000	0.26759	0.99891	0.99999	---	0.19622	0.01922	0.99700
TEM:ARR	0.25393	1.00000	0.05421	0.17890	0.19622	---	0.00000	0.07378
TEM:TBN	0.01221	0.00001	0.10068	0.11494	0.01922	0.00000	---	0.27459
TEM:TBR	0.99151	0.10554	1.00000	0.99995	0.99700	0.07378	0.27459	---

Overall, previous discussions support validity of the following considerations regarding the effect on reporting time of AR-based and recommendable reporting solutions:

- Differences in times between experimental failures cannot be considered significant.
- Differences among AR-recommender (ARR) solutions and alternative solutions (ARN, TBR, TBN) can be considered significant.
- Differences among AR-non-recommender (ARN) solutions and alternative solutions (TBR, TBN) cannot be considered significant.
- In electric failure experiments, average reporting time for ARR is 20%-23% faster than alternative solutions (ARN, TBR, TBN).
- In electronic failure experiments, average reporting time for ARR is 22%-38% faster than alternative solutions (ARN, TBR, TBN).

These considerations indicate the validity of this research's hypotheses regarding the effect of AR content and recommendations in reporting efficiency. AR content formats enabled significantly faster reporting times but only when implementing context-aware AR-based recommendations.

5.6.4 Workload study

Workload surveys aim to evaluate testers perceived requisites on diagnosis reporting tasks for validating the effectiveness and efficiency measures assumption. Prior to experimentation, testers were asked to complete a pair-wise comparison survey for weighting workload factors described by NASA-TLX methodology (Hart, 2006). After experiments, testers were questioned to rate each workload factor in the experiments conducted. According to NASA-TLX workload factors (Table 5-8), testers were hypothesised to perceived experimental workload with higher requisites on Mental Demand and Performance for providing accurate reports on experimental failures' root causes.

Figure 5-19 display testers' average responses regarding workload factors weighted ratings (Hart, 2006). These results show that 'Performance' and 'Effort' were the most relevant factors as perceived by testers, with Mental Demand and Frustration following closely. Overall, these results suggest the validity of this research's hypothesis that enounces diagnosis reporting tasks as mentally and performance demanding.

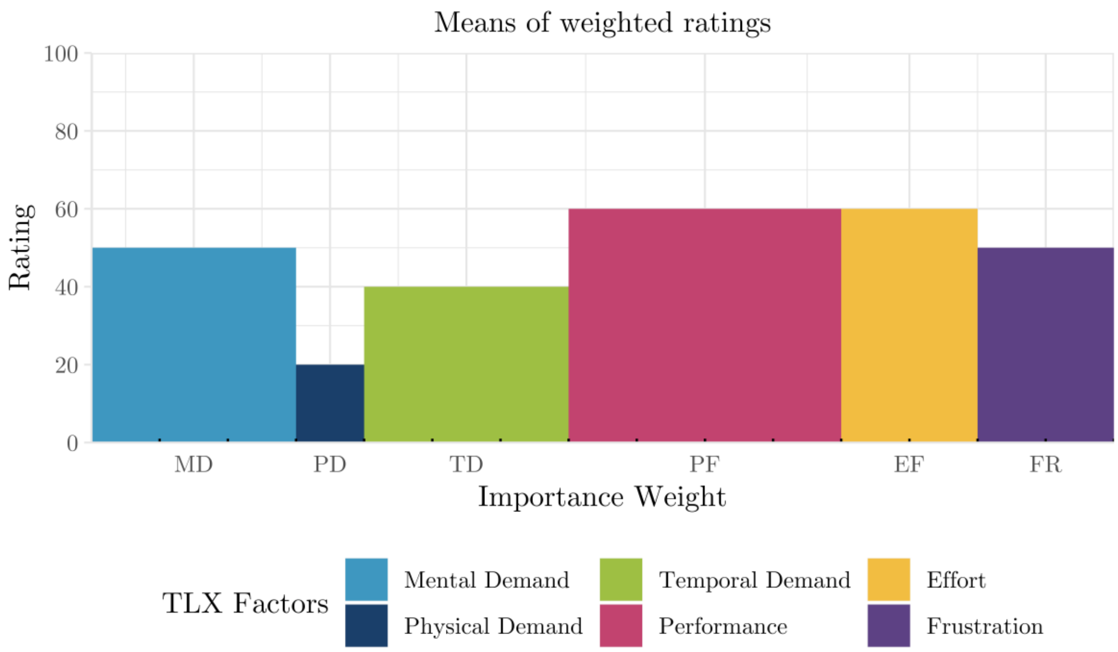


Figure 5-19. NASA-TLX weighted rates plot on workload factors showing average results of diagnosis reporting testers (Hart, 2006).

5.6.5 Usability study

Usability surveys' study aims to validate the following hypothesis: *"testers perceived usability improves for reporting tools implementing AR content and recommendations"*. Usability is defined as a qualitative measure regarding the degree to which reporting solutions (ARR, ARN, TBR, TBN) enhance completion of diagnosis reporting tasks (Nielsen, 1993). Usability surveys consisted of a separate section for each criterion including for each aspect regarding the reporting solutions tested in stopwatch experiments. The survey's questionnaire can be found in Appendix I. Testers were asked to declare their agreement with these statements in a Likert Scale from 1 to 5. The reason to select such scale is based on the results presented by Weijters, Cabooter and Schillewaert (2010), who suggested that it maximises potential information transmission when surveying non-expert populations (Subsection 5.5.3.3). Results collected served to evaluate the proposed contributions' usability compared to other specific approaches.

Table 5-10 described surveyed usability criterions and solutions' aspects against to which assess those. In order to confirm experimental results, 28 testers were hypothesised to perceive usability of AR reporting solutions (ARR, ARN) at least as good as non-AR solutions (TBR, TBN) with small variances between recommender and non-recommender ones.

Figure 5-20 illustrates a box and whiskers plot to summarise testers' responses for each usability criterion per reporting solution. Medians for criterions responses range between 3 and 4 in a Likert Scale 1-5 with higher variabilities for ARR and TBN solutions. Ease-To-Learn was the lowest scored criterion for all solutions with the lowest medians and reasonable distributions' variabilities. Ease-To-Use was the criterion with lowest variabilities and medians scoring around 4. Effectiveness and Satisfaction responses were higher for AR solutions in terms of their medians but with higher variability for ARR. Overall, TBN and TBR solutions were better perceived regarding Ease-To-Use, while ARR and ARN were better at Effectiveness and Satisfaction. Due to the number of statements included in surveys questionnaires per criterion (Appendix I), further analyses on these criterions can be done studying different solutions' aspects they are affected by.

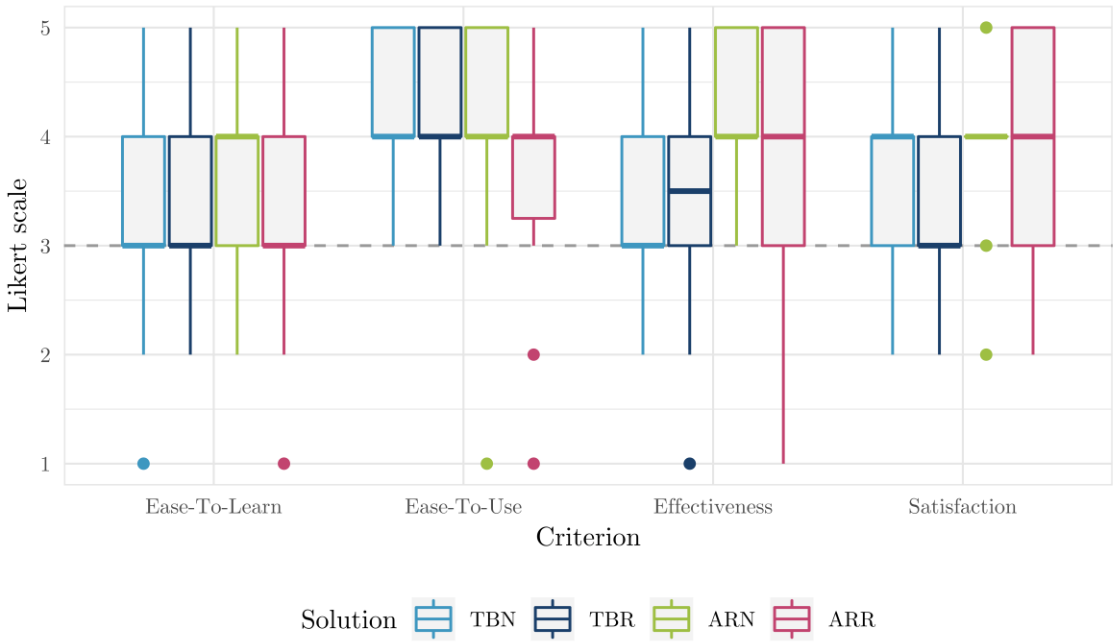


Figure 5-20. Distribution (maximum, upper quartile, median, lower quartile, minimum and outliers) of usability criterions' survey responses for TBN, TBR, ARN and ARR solutions.

Figure 5-21 presents average testers' responses regarding Ease-To-Learn aspects of alternative solutions. These results suggest that AR solutions (ARR, ARN) have higher learning curves than non-AR solutions (TBR, TBN) according to the differences between ease at start and at end. Also, AR solutions seem more intuitive than non-AR ones.

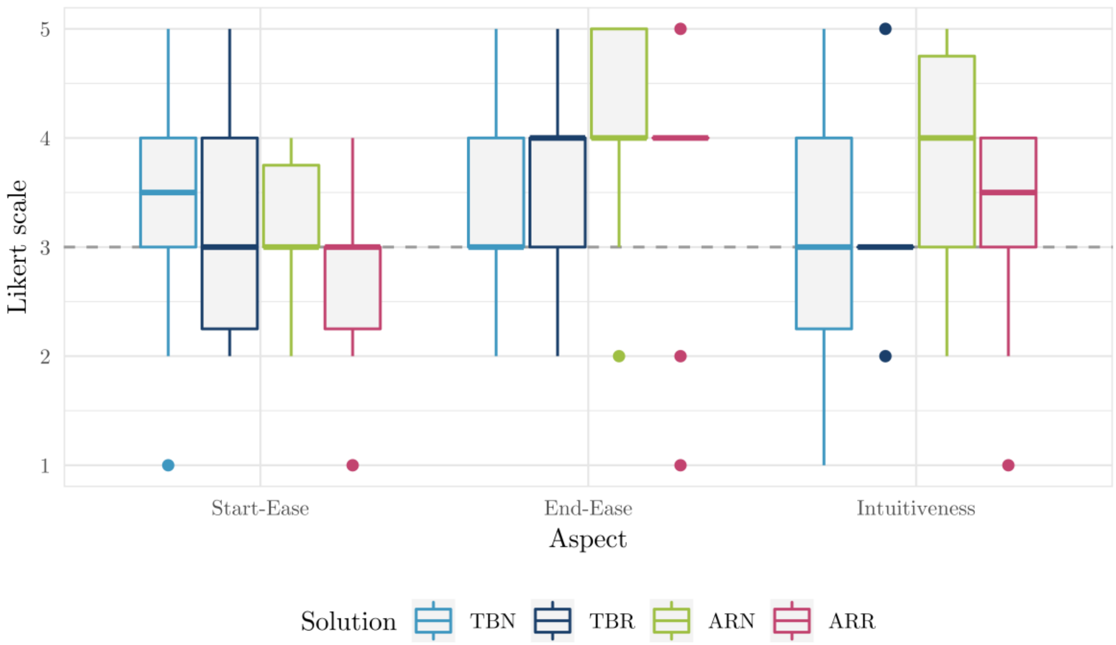


Figure 5-21. Distribution (maximum, upper quartile, median, lower quartile, minimum and outliers) of Ease-To-Learn aspects' survey responses for TBN, TBR, ARN and ARR solutions.

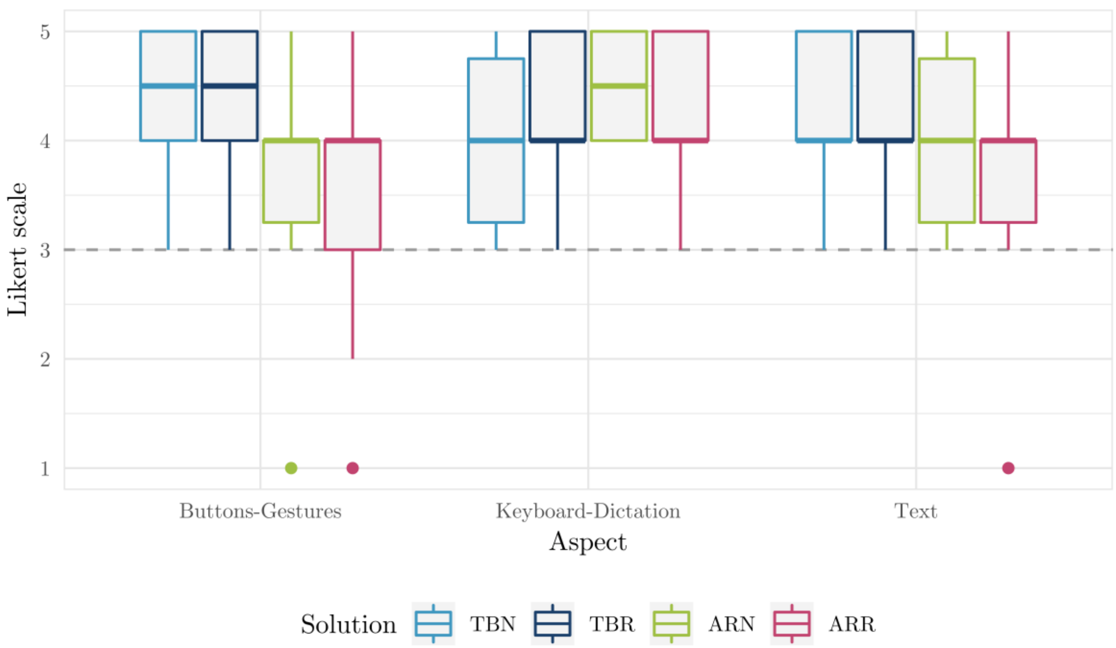


Figure 5-22. Distribution (maximum, upper quartile, median, lower quartile, minimum and outliers) of Ease-To-Use aspects' survey responses for TBN, TBR, ARN and ARR solutions.

Figure 5-22 presents average testers' responses regarding Ease-To-Use aspects of alternative solutions (TBN, TBR, ARN, ARR). Ease-To-Use aspects refer to solutions' user interface items. Gestures and text of AR solutions were perceived lower than buttons and text from non-AR tools. Instead, testers perceived AR dictation capabilities better for data input than normal tablet keyboard.

Figure 5-23 presents average testers' responses regarding Effectiveness aspects of alternative solutions (TBN, TBR, ARN, ARR). AR solutions scored higher in certain aspects such as ease to understand, efficiency and confidence increase and content suitability. Besides, AR and non-AR solutions had similar testers responses for error reduction and report accuracy. ARR is the solution with higher variabilities.

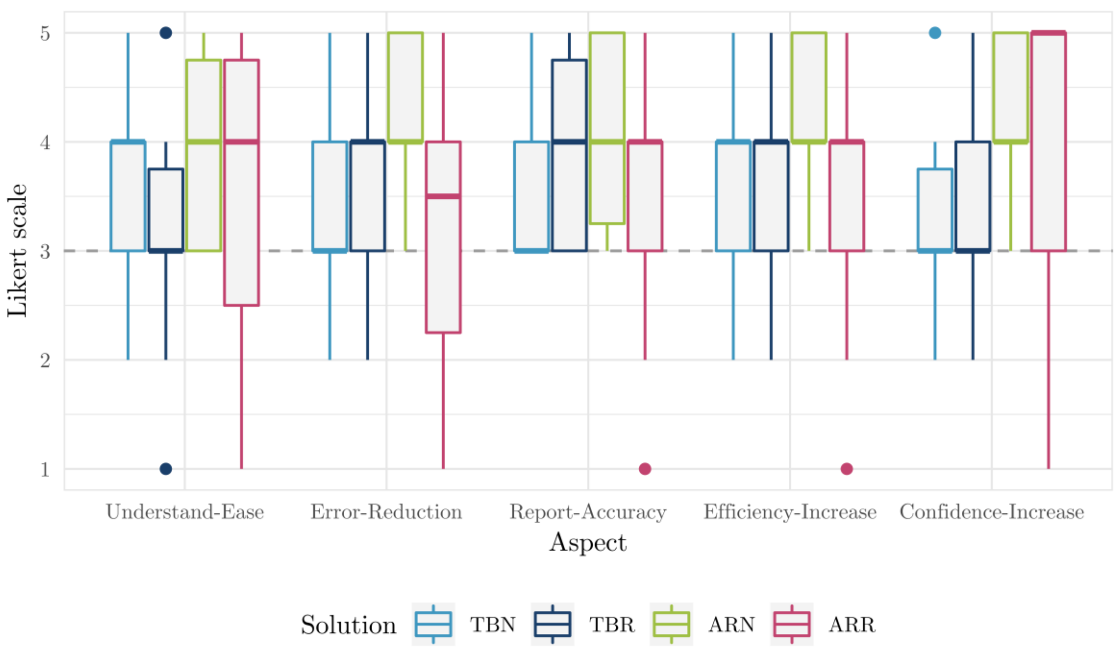


Figure 5-23. Distribution (maximum, upper quartile, median, lower quartile, minimum and outliers) of Effectiveness aspects' survey responses for TBN, TBR, ARN and ARR solutions.

Figure 5-24 presents average testers' responses regarding Satisfaction aspects of alternative solutions (TBN, TBR, ARN, ARR). Testers perceived design of different solutions very similarly. Instead, feeling and overall satisfaction of AR solutions (ARR, ARN) was better perceived by testers than non-AR ones.

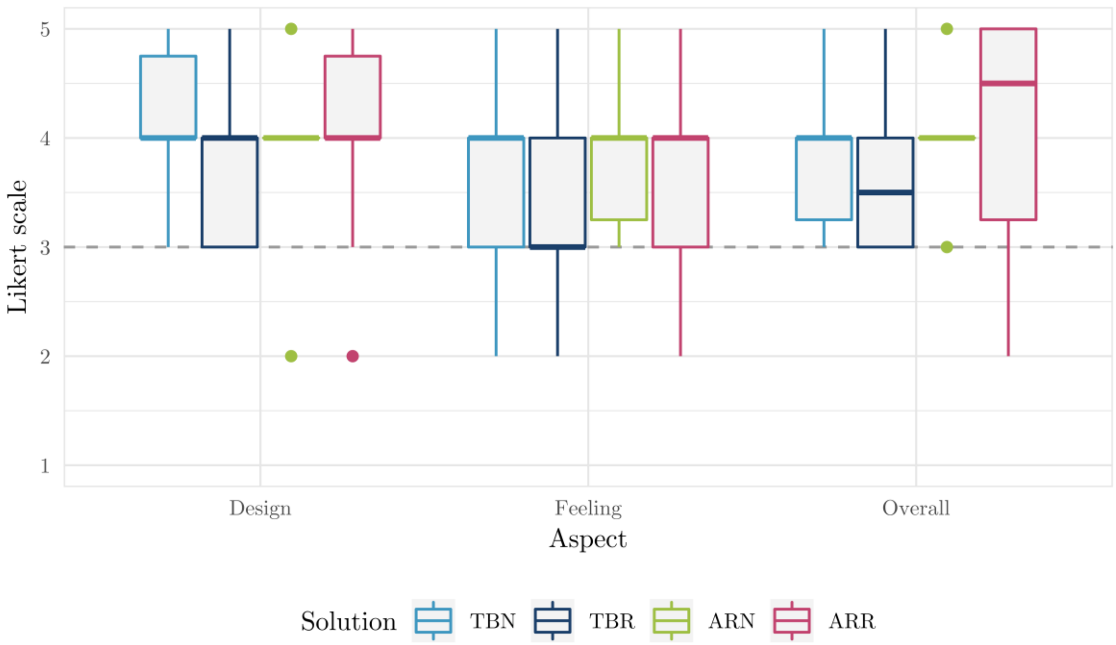


Figure 5-24. Distribution (maximum, upper quartile, median, lower quartile, minimum and outliers) of Satisfaction aspects' survey responses for TBN, TBR, ARN and ARR solutions.

Overall, usability surveys indicate that testers perceived AR solutions as more effective and satisfactory, while non-AR solutions were perceived as easier to learn and understand. According to these results, AR solutions showed higher learning curves than non-AR tools, although advanced data input methods (e.g. dictation) were also well considered by testers. Besides, AR solutions were perceived as more accurate, enhanced and assuring for diagnosis reporting operations. These results indicate validity of this research's hypothesis regarding improved usability of AR reporting tools (ARR and ARN), but do not suggest significant differences among recommender and non-recommender solutions (ARR vs ARN and TBR vs TBN).

5.6.6 Discussion

Previous analytical results aimed to evaluate this research's hypothesis (Section 5.5.2) for demonstrating the validity of this research's contributions (Section 5.4).

The first validation hypothesis stated that *"recommendations accuracy improves with the use of AR-based hybrid recommendations compared to conventional ontology-based recommendations"*. Accuracy effect study (Section 5.6.1) analysed the relation between recommendations proposed and chosen by testers and recommender methods in diagnosis reporting experiments. Its results showed that the proposed method's accuracy (ARR) was 2.2 times higher than conventional recommender's (TBR) on average for both experimental failures. T-tests results indicated a statistically significant difference on accuracy results per method. Therefore, this hypothesis can be considered valid within the context of the experiments conducted. These results imply that the proposed context-aware and ontology-based AR recommender method enhances knowledge re-use in diagnosis reporting tasks. Besides, identified accuracy improvements can be the result of: (1) more precise recommendations and (2) more correct testers' selections caused by contextualised visualisations. Future studies can investigate the independent effects of each cause in real-life experiments to quantitatively measure their independent impacts.

The second validation hypothesis assumed that *"errors are a measure of reporting effectiveness and time is a measure of reporting efficiency"*. Pearson's correlation test's results (Section 5.6.2) indicated that the correlation between these response variables could not be considered statistically significant. Moreover, Cohen's interpretation of correlation test's results suggested that even if correlation was significant, it was small. Hence, it can be said that within the context of this research's experiments the assumption above can be considered valid.

The third validation hypothesis enounced that *"errors reduce with AR reporting solutions compared to non-AR ones"*. And the fourth one stated that *"errors reduce with recommender reporting methods compared to non-recommender ones"*. Errors study (Section 5.6.2) analysed the effect on errors rate per reporting task of diverse reporting tools (ARR, ARN, TBR, TBN) in failure diagnosis experiments. Errors rates for AR-based tools (10%) reduced 50% compared tablet-based solutions' errors rates (20%) in both experimental failures. Besides, two-way ANOVA test results indicated statistical significance of these differences within the experiments' context. Therefore, the second hypothesis can be considered valid but not the third one. These results also align with this research's contributions that expected only the proposed dynamic authoring method to improve effectiveness (errors) of diagnosis reporting tasks. Due to the nature of experiments, errors were measured per reporting task. A possible explanation for these conclusions is that AR methods allow to contextualise complex data input tasks, while recommender methods reduce their total number. Future works can extend recommendation facets to different datasets for studying recommenders' effects in singular data-input tasks.

The fifth validation hypothesis stated that *"time decreases with the use of AR reporting solutions compared to non-AR ones"*. And the sixth one enounced that *"time decreases with the use of recommender reporting tools compared to non-recommender ones"*. Time study (Section 5.6.3) evaluated the effect on time of AR and recommender reporting and their counterparts in failure diagnosis experiments. In electric failure experiments (CNN), ARR testers were 45 seconds faster than TBN's, while ARN's and TBR's were 10 and 4 seconds faster than TBN testers. In electronic failure experiments (TEM), results also show similar differences. ARR testers are 61 seconds faster than TBN, while ARN's are 22 and TBR's are 11 seconds faster than TBN testers. Besides, two-way ANOVA test results indicated a significant effect of the solution factor on experimental time results. Post hoc comparisons from Tukey's HSD tests confirmed that resultant differences were mostly driven by the variances between the proposed reporting tool (ARR) and different alternatives. Overall, these results suggest the validity of both time-related hypothesis together as the proposed AR-recommender solution was found faster than its counterparts in diagnosis reporting experiments. These results imply that the proposed AR recommender and dynamic authoring methods improve efficiency of diagnosis reporting tasks. They also suggest the need to correlate AR content and recommendations to improve their efficiency. Thus, indicating that efficiency improvements in reporting tasks through AR are only achieved through knowledge capture and re-use combined. Future studies can investigate this correlation more in-depth to quantitatively measure their independent effects. They can also further corroborate this research's results applying the proposed methods to other AR-maintenance knowledge capture applications.

The seventh validation hypothesis stated that *"testers' should perceived experimental workload as mentally and performance demanding"*. Thus, aiming to corroborate previous hypothesis regarding improvements on reporting effectiveness and efficiency. Surveys results (Section 5.6.4) evaluate NASA-TLX criterions regarding tasks workload requisites. These suggested that testers perceived reporting tasks mostly as performance and effort demanding, with temporal demand and frustration factors following closely. Nevertheless, post-experimental surveys for scoring factor rates were done after testers completed both experiments with alternative AR and non-AR solutions. Also, testers were novices with little maintenance experience. Future works can investigate the difference in perceived workload with different solutions and with real-life maintainers to further clarify these tasks' requisites.

The final validation hypothesis enounced that *"perceived usability improves for reporting tools implementing AR content and recommendations"*. It helped to corroborate previous hypotheses regarding improvements on reporting effectiveness and efficiency. Surveys results (Section 5.6.5) evaluated usability criterions according to different reporting tools' aspects. These indicated that testers' perceived AR solutions as more effective and satisfactory, while non-AR solutions were perceived as easier to learn and understand.

5.7 Conclusions and future works

5.7.1 Conclusions

This chapter proposed (1) a method to provide context-aware and ontology-based AR recommendations to reduce extensive selection lists in diagnosis reporting tasks and (2) a method to dynamically create and allocate content in augmented scenes to avoid augmented content overload and ease data input tasks. Their aim was to prove that automatic recommendable authoring can improve efficiency and effectiveness of AR knowledge capture and re-use in diagnosis reporting applications. They were implemented in a cloud-based AR system prototype for validation with effectiveness and efficiency experiments and usability surveys in reporting diagnosis operations. Experimental results indicated that the proposed AR-recommender reporting method (ARR) reduces reporting errors (50%) and time (20%) compared to alternative non-AR (TBR) and non-recommender (TBN, ARN) solutions. These results also displayed that recommendations' accuracy doubles for the proposed AR-based hybrid techniques (ARR) compared to conventional ontology-based methods (TBR). Besides, surveys results suggested that testers perceived the proposed reporting solution as more effective and satisfactory than its non-AR and non-recommender counterparts. Thus, proving that the proposed methods can improve effectiveness and efficiency of diagnosis reporting applications.

The proposed methods for automatic recommender authoring contribute to fill an important research gap towards the integration of human operations in digital maintenance. Maintenance reporting operations are performance and mentally demanding and so, prone to errors in efficiency-challenging conditions. These often result in reports with decreased accuracy and unstructured knowledge difficult to re-use. The proposed hybrid recommender can increase the accuracy of recommendations, reducing the size of selection lists to decrease reporting time. The proposed dynamic authoring method can reduce augmented scene’s overload to further reduce reporting time. Besides, it can automatically produce augmented content that checks correctness and eases identification of reported data to reduce reporting errors. Hence, this research’s proposal can contextualise and structure diagnosis reporting tasks and increase the correctness of reported data. Thus, enhancing the digitalisation of diagnosis reporting operations and facilitating capture and re-use of human knowledge in digital maintenance.

5.7.2 Future works

Future works will explore further applications and enhancements of the proposed methods for pursuing human knowledge integration in digital maintenance. The following list extends the future works described within this chapter’s discussions:

- Dynamic content allocation:

- Investigate factors that can cause occlusion in AR maintenance applications and improve proposed allocation and scaling mechanisms to reduce it.
- Study dynamics of AR knowledge capture applications and improve proposed allocation mechanisms to enhance content navigation.

- Content formats:

    - Study dynamics of maintenance knowledge capture applications and improve content formats adaptability to enhance simultaneity of knowledge transfer and capture.
    - Develop advanced methods to determine input data correctness for further reducing reporting errors and improving effectiveness.
    - Develop advanced content formats to report heterogenous and unstructured data types (e.g. audio or images) for further integration of human knowledge in digital maintenance.
    - Develop adaptive content formats according to user and environmental conditions (e.g. performance or light) for further decreasing reporting time and improving efficiency.

- Recommendation facets:

    - Extend the proposed recommendation framework to different techniques (e.g. collaborative filtering) and implement automatic data collection methods (e.g. content-tracing or eye-tracking) for improving recommendations accuracy.

- Applications and experiments:

    - Experiment with the proposed methods and real-life maintainers in real-life conditions to study the correlation between AR content and recommendations in maintenance reporting operations and study their independent effects on reporting workload, errors and time.

- Develop new content formats and recommendation facets for different maintenance reporting operations (e.g. service logs) to extend integration of human knowledge in digital maintenance.

AR technologies are information visualisation tools that can smooth knowledge transfer between humans and digital systems. Future works aim to find necessary research towards a framework for automatic recommender authoring for AR knowledge transfer and capture applications. Thus, envisioning a future where maintenance digital systems can integrate human knowledge to its full extent.

Chapter 6
Thesis discussion

6.1 Discussion on contributions to knowledge

This PhD thesis hypothesised that

"by automatically creating, adapting and recommending augmented content, AR applications can transfer, capture and re-use knowledge in different tasks and contexts to improve efficiency and effectiveness of diagnosis operations while reducing implementation costs for enhancing human knowledge integration in digital maintenance".

In order to prove the hypothesis, this PhD thesis proposed a series of objectives (Section 1.3.2). These have been conducted as independent research works and presented in a journal-paper format in previous chapters. Their conclusions and contributions to knowledge regarding this thesis hypothesis are discussed in the following subsections.

6.1.1 First PhD thesis objective

The first thesis objective aimed to *"identify the relation between AR content-related techniques and knowledge transfer, capture and re-use"* for finding relevant research gaps regarding human knowledge integration in digital maintenance. In order to accomplish this objective, Chapter 2 presented a systematic literature review of AR content-related methods for knowledge transfer, capture and re-use in maintenance applications. This review utilised thematic and numerical analysis to evaluate the utilisation of AR content-related techniques in 74 publications according to the concepts defined in Table 2-6: asset, operation, task, knowledge, authoring, context-awareness and interaction-analysis. These analyses helped to classify existing types of AR content-related methods and identify their maturity levels regarding different maintenance operations. Besides, they also helped to determine the relation between content-related methods and AR knowledge transfer, capture and re-use capabilities and the maintenance operations which those were achieved. These results can be considered to contribute to academic literature as they resulted in future lines of research work for content-related methods in AR-maintenance applications. Those results that are relevant for this thesis are listed below:

1. There is lack of research in AR content-related methods focusing on certain maintenance operations like diagnosis, and management.
2. There is lack of AR content-related methods that are applicable to more than one maintenance operation.
3. Advancements in context-awareness and interaction-analysis methods appear to be directly correlated to advancements in authoring techniques.
4. Most advanced techniques in authoring, context-awareness and interaction-analysis are ‘automatic knowledge-based’, ‘multiple-context knowledge-based’ and ‘automatic data acquisition and analysis’, respectively to each technique.
5. Automation of content-related techniques appear to have a positive effect on cost-effectiveness of AR-maintenance systems and applications.
6. Advancements in content-related methods seemed to be linked to the existence of knowledge-domain representations of maintenance operations.
7. Authoring methods can enable knowledge transfer or capture by creating content that either displays or retrieves information.
8. AR knowledge re-use capabilities seem to require of context-awareness and interaction-analysis methods that can adapt content to maintenance, user and environmental contexts.
9. Evidence of AR knowledge capture and re-use applications seemed to mostly focus on maintenance repair and assembly operations.

These conclusions helped to identify this PhD thesis objectives and their scope regarding maintenance operations.

First, these conclusions identified a research gap in content-related techniques for diagnosis operations (1), which also required of knowledge-domain models for further AR advancements (6). These served to focus the second thesis objective on developing a knowledge-domain model of diagnosis tasks and demonstrating its validity for knowledge capture and re-use in diverse actions (Chapter 3). Contributions to knowledge and to the thesis aim of the second thesis objective are discussed in Section 6.1.2.

Second, literature review conclusions showed that AR content-related methods should be applicable to more than one operation (2) and automated (5) for being cost effective. They also implied that authoring and context-awareness methods can enable AR to transfer and capture knowledge (7). Besides, they suggested that progress in authoring methods act as basis for advancements in interaction-analysis (3). Based on these, the third thesis objective was determined to develop an automatic and context-aware authoring (4) method for knowledge transfer in multiple AR maintenance applications (Chapter 4). So, it could serve as the basis for more advanced context-awareness and interaction-analysis methods for knowledge capture and re-use while demonstrating cost-effectiveness of AR technologies. The third thesis objective's contributions to knowledge and to the thesis aim are discussed in Section 6.1.3.

Finally, these literature review conclusions proposed that AR requires interaction-analysis and context-awareness to enable knowledge re-use (8). Also, they implied that there was little evidence of AR knowledge re-use applications in diagnosis operations (9). Therefore, the fourth thesis objective was set to develop a context-aware, interaction-analysis recommender method for knowledge re-use in diagnosis reporting operations (Chapter 5). This method closes the loop regarding human knowledge integration in maintenance through AR because of the following reasons. First, it builds upon the methods proposed in previous chapters to enable diagnosis knowledge re-use by analysing diagnosis contexts and suggesting probable faults. Second, it enables to demonstrate that diagnosis knowledge can be re-used for the same actions but in different contexts, while the first thesis objective demonstrated that knowledge can be re-used for different actions. This is important because academic definitions state that knowledge re-use should be applicable to different operations and contexts (Pérez-Salazar *et al.*, 2019). The fourth thesis objective's contributions to knowledge and to the thesis aim are discussed in Section 6.1.4.

6.1.2 Second PhD thesis objective

The second objective aimed to *"develop ontology-based reporting and monitoring methods and validate their ability to capture and re-use knowledge in failure diagnosis operations"*. From a thesis perspective, this objective had two main requisites:

1. to create a knowledge-domain model of human diagnosis tasks that could serve as basis for further advancements in AR content-related techniques, and

2. to validate this model's ability to capture and re-use knowledge in diagnosis operations.

For fulfilling those, Chapter 3 proposed a knowledge-domain of diagnosis tasks and ontology-based reporting and monitoring methods to capture and re-use knowledge for improving efficiency of diagnosis operations. This research utilised ontology-structural metrics, expert interviews, reporting tests, monitoring experiments and usability surveys for validating the ontology and methods mentioned above. Results of ontology-structural metrics comparisons, expert interviews and reporting tests showed that the proposed ontology-based reporting method can capture knowledge in a comprehensive and structured manner. Monitoring experiments and usability surveys results indicated that the proposed monitoring method improves efficiency of fault-finding tasks better than common, data-driven methods These results imply that the proposed methods for knowledge capture and re-use contribute to fill an important research gap towards the integration of expert knowledge in condition monitoring. From a thesis perspective, they helped to demonstrate that the proposed ontology in diagnosis operations serve as basis for further advancements in AR content-related methods for knowledge capture and re-use. It also helped to validate knowledge re-use applicability for different operations like monitoring.

6.1.3 Third PhD thesis objective

The third objective aimed to *"develop an ontology-based method for automatic and adaptive authoring and validate its ability for effective knowledge transfer in diverse maintenance operations"*. From a thesis perspective, this objective had two main requisites:

1. develop an authoring technique applicable to multiple maintenance operations for cost-effective knowledge transfer, and
2. develop an automatic authoring technique that would allow to advance other content-related methods towards AR knowledge capture and re-use.

For fulfilling those, Chapter 4 proposed an ontology-based, pattern-matching algorithm and programmable content formats for automatic adaptive authoring in AR maintenance applications. This research utilised efficiency experiments and usability surveys to validate the proposed authoring technique through its content in two maintenance operations: repair and remote diagnosis. For validation purposes, the proposed method was compared through its content against that of operation-specific authoring solutions and non-AR alternatives. Experimental results confirmed that the proposed authoring method achieved similar efficiency improvements to specific authoring solutions when compared to non-AR methods. Also, surveys results indicated similar perceived usability for all three authoring solutions' content. These results imply that the proposed method contributes to fill an important research gap towards industrial implementation of AR in maintenance applications. From a thesis perspective, they helped to demonstrate that the proposed authoring method can automate AR content creation processes for cost-effective implementation while achieving sufficient knowledge transfer effectiveness. Besides, the applicability of proposed method to multiple operations implied that it could be used as a framework for AR human knowledge integration in digital maintenance. Although, it would require of advanced context-awareness and interaction-analysis methods to enable knowledge capture and re-use.

6.1.4 Fourth PhD thesis objective

The fourth and final objective aimed to *"develop an ontology-based, context-aware, interaction-analysis AR recommender method and validate its ability to capture and re-use knowledge in diagnosis reporting"*. From a thesis perspective, this objective had to main requisites:

1. extend the proposed authoring method with context-awareness interaction-analysis techniques for enabling AR knowledge re-use in different contexts than captured, and

2. demonstrate the proposed framework for AR human knowledge integration works in maintenance diagnosis applications.

For fulfilling those, Chapter 5 proposed a hybrid recommender technique with dynamic authoring for AR knowledge capture and re-use in diagnosis reporting. It uses interaction-analysis methods to identify the target case and ontology-based, context-aware similarity functions to assess cases. This research employed effectiveness and efficiency experiments and usability surveys to validate the proposed recommender method through comparison against alternative non-AR and non-recommender reporting tools. Experimental results indicated that the proposed method reduces reporting time and errors compared to alternative approaches. Survey results also suggested that the proposed reporting tool was perceived as more effective and satisfactory. These results imply that the proposed recommender method contributes to fill an important research gap towards the integration of human knowledge in diagnosis operations. From a thesis perspective, it provides a programmable recommender framework that enables AR technologies to capture knowledge and also to re-use existing knowledge in similar operations but different contexts. Therefore, these contributions along with those from earlier thesis objectives demonstrate that AR enables human knowledge integration in digital maintenance through knowledge transfer, capture and re-use in different operations and contexts.

6.1.5 Contributions to knowledge

Previous subsections discussed this thesis' contributions to knowledge according to previous chapters' conclusions and thesis objectives they aimed to fulfil. Based on them, this thesis contributions to knowledge can be summarised as follows:

C-1. A theory on the AR technologies that act as enablers of knowledge capture, transfer and re-use that serves as guideline for future AR research to enable human knowledge integration in digital maintenance. [O1 – Ch2]

C-2. A combination of ontology-based reporting and monitoring methods that can capture expert diagnosis knowledge in failure reporting tasks and re-use it to improve efficiency of fault-finding tasks by integrating data- and knowledge-driven monitoring approaches. [O2 – Ch3]

C-3. An automatic adaptive authoring method for cost-effective AR deployment with sufficient knowledge transfer effectiveness that can serve as the basis of a framework for AR knowledge transfer in digital maintenance. [O3 – Ch4]

C-4. An AR hybrid recommender method for expert knowledge re-use in diverse diagnosis contexts to improve efficiency and effectiveness of reporting tasks that can serve as an extension to the previous framework that enables knowledge capture and re-use in digital maintenance. [O4 – Ch5]

According to these contributions, it can be said that this PhD has demonstrated how AR content-related techniques can enable knowledge transfer, capture and re-use to improve efficiency and effectiveness of diagnosis operations in a cost-effective manner. Figure 6-1 presents an overview of this framework taking into consideration this thesis contributions and their scopes of application. Besides, the proposed methods can serve as the basis of a framework for human knowledge integration in digital maintenance in the context of Industry 4.0. Nevertheless, that would require to further validate these methods in other maintenance related operations like training or repair and extend its capabilities for knowledge transfer, capture and re-use in other contexts. The following section discusses this thesis researches limitations to identify the next steps to advance towards a framework for human knowledge integration in digital maintenance.

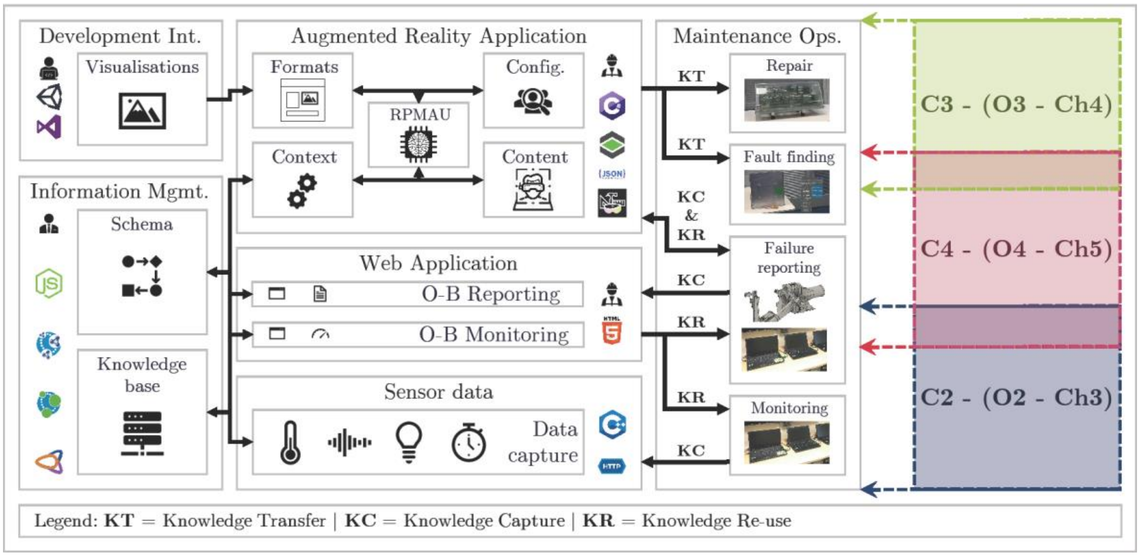


Figure 6-1. Overview of the proposed framework for human knowledge integration in digital maintenance according to this thesis contributions – The contributions presented in this thesis chapter's are combined into a single system architecture (left-side), including the logos of the languages and tools utilised to prototype it. These point towards the different maintenance operations they demonstrated to support, along with the knowledge management processes to do so (knowledge transfer, capture and re-use). Besides, the limits of each contributions are marked through the boxes on the right-side of the figure.

6.2 Discussion on research limitations

This PhD thesis contains several chapters to describe the contributions proposed to achieve its objectives. These chapters include discussion sections that analysed the validity of their contributions according to their specific research objectives. Nevertheless, it is still necessary to evaluate the validity of those contributions according to the thesis aim. This PhD thesis aimed to develop AR content-related methods that could create, adapt and analyse content for knowledge transfer, capture and re-use to improve efficiency and effectiveness of diagnosis operations. The ulterior motive was to combine these methods to propose a framework for human knowledge integration in digital maintenance through cost-effective AR technologies. The proposed contributions and their validation results may have certain limitations to demonstrate their validity for such purpose.

Overall, these limitations involve the validity of the proposed contributions as a framework for knowledge integration according to different factors like cost-effectiveness or user and environment conditions. These limitations are discussed in the following subsections according to the factors they are associated with.

6.2.1 Validity for diagnosis operations

This PhD thesis aimed to validate the proposed AR content-related methods for knowledge transfer, capture, and re-use in diagnosis operations. Table 6-1 presents this thesis case studies in terms of failures and diagnosis tasks according to the knowledge integration methods (transfer, capture and re-use) to validate.

Table 6-1. Overview of thesis contributions validity for diagnosis operations regarding cases of study failures and tasks.

Knowledge	Chapter	Objective	Case study failure	Diagnosis operation
Transfer	4	3	Mechanical	Collaborative fault-finding and manual repair
Capture	3	2	Electro-mechanical, electric and electronic	Failure diagnosis reporting
	5	4	Electric and electronic	Failure diagnosis reporting
Re-use	3	2	Electric and electronic	Control monitoring
	5	4	Electric and electronic	Failure diagnosis reporting

Table 6-1 indicates that AR knowledge transfer, capture and re-use were validated with mechanical, electronic and electrical failures. The proposed methods were designed agnostically to failure types and so, there is no reason to believe they would not be valid for others like hydraulic or pneumatic. Still, future works could corroborate so and also extend the proposed methods regarding capture of different heterogeneous data sources like images or audio for improving human knowledge integration.

Table 6-1 also implies that human knowledge integration through AR knowledge transfer, capture and re-use was validated for different diagnosis tasks. Knowledge transfer was proven in fault-finding and repair, knowledge capture in diagnosis reporting tasks and knowledge re-use in monitoring and diagnosis reporting tasks. On one side, these validations demonstrate that the proposed methods allowed to re-use knowledge in other operations (monitoring) and contexts (reporting) than those were it was captured. On the other side, it has not been proven that knowledge captured can be re-applied in other diagnosis tasks like fault-finding. Future works could investigate knowledge re-use applicability not only to other diagnosis tasks but also to other maintenance tasks like repair (e.g. to suggest report procedures based on faults identified).

Overall, it can be said that the proposed contributions for human knowledge integration improve efficiency and effectiveness of diagnosis operations, but further investigation is required to demonstrate its impact on other maintenance related operations.

6.2.2 Validity for cost-effective Augmented Reality implementation

Another requisite established by this PhD thesis aim was to demonstrate that the proposed contributions could transfer, capture and re-use knowledge in a cost-effective manner. As discussed in Chapter 4, the proposed method for automatic and adaptive authoring separates content creation and information management processes and automates the former. Thus, reducing the costs associated with duplicating existing maintenance information in the form of augmented content since maintenance experts do not need to re-create it. Besides, additional content-related methods proposed in Chapter 5 were designed so they could be integrated in the proposed authoring framework. Thus, providing additional gains in terms of efficiency and effectiveness improvements without increasing content-creation costs. Therefore, it can be said that the proposed framework is more cost-effective since it provides higher added value with reduced costs than conventional AR frameworks where augmented content needs to be manually created (Bottani and Vignali, 2019). Nevertheless, it is still necessary to quantify in terms of monetary costs and gains what are the benefits of the proposed AR framework for human knowledge integration. Future works can investigate these aspects by developing and testing cost analysis models for comparison of the proposed framework with other AR implementation systems.

6.2.3 Validity in real-life conditions

This PhD thesis has proposed content-related methods that adapt content to maintenance, user and environmental contextual factors. These have been validated in terms of efficiency and effectiveness improvements against other AR and non-AR solutions in laboratory conditions. There were two main reasons for conducting validation through laboratory experiments. First, they were necessary to keep certain factors (e.g. ergonomics, light conditions) constant as they were out-of-scope of the validation methods proposed. Second, they were necessary to enable validation through comparison with other solutions, which were not ready for testing in industrial environments. Nevertheless, the proposed contributions are part of a system prototype which was tested in industrial scenarios operated by this thesis industrial sponsor. First impressions were positive in terms of usability and impact on operators' performance, but further experimentation is required to evaluate the impact of non-considered factors in operational efficiency and effectiveness.

Overall, it can be said that the proposed framework for knowledge integration could work in real-life conditions, but further validation is needed to corroborate laboratory results. These future research works should include the following:

- Investigate proposed methods adaptiveness to other user and environment contextual factors like user's expertise, ergonomics or light conditions. Future works could also include advancements in interaction-analysis techniques like eye-tracking or content-tracing regarding automatic data collection and analysis of those factors. Future works should also include experiments in industrial environments with real-life conditions and maintainers.

- Study how the proposed AR framework for human knowledge integration can communicate with existing maintenance information systems for easing its real-life implementation. Future works can include translation mechanisms between ontology-based and other types of databases like SQL or graphical. Besides, future works could also study how to apply ontology-based information systems for knowledge integration in Industry 4.0 contexts. The PhD student is the co-author of a publication (Erkoyuncu *et al.*, 2020) that evaluated ontology-based information systems for Digital Twins. It utilised this thesis prototypes for validation purposes.

6.2.4 Validity for extended applicability

This PhD thesis focused on demonstrating that AR knowledge transfer, capture and re-use can improve efficiency and effectiveness of diagnosis operations. The motivation was to demonstrate the added value of human knowledge integration in digital maintenance. There were three main reasons to select diagnosis as the maintenance area were to focus this thesis efforts. First, it was a relevant area for knowledge integration due to current challenges with structure and heterogeneity of existing sources (Yazdi, 2019). Second, there was lack of advanced research in AR content-related techniques for knowledge capture and re-use compared to other maintenance operations (Fernández Del Amo *et al.*, 2018). Third, it was challenging from an AR perspective to augment diagnosis knowledge due to its abstraction and implicitness compared to other maintenance operations (Longo, Nicoletti and Padovano, 2019).

Although this thesis objectives focused in diagnosis operations, its contributions utilise ontologies to represent the operational domain. This thesis third (Chapter 4) and fourth (Chapter 5) objectives proposed their content-related methods generically, so they are applicable to any maintenance operation. Furthermore, Chapter 4 also validated its contributions in repair operations. Thus, indicating that the proposed AR framework can be applicable to other maintenance tasks.

Also, there is also no reason to believe that this framework could not be applicable to other asset-life phases (e.g. manufacturing) or other fields of AR application (e.g. medicine). Nevertheless, it would require of additional research to extend the proposed framework's applicability. These future works include:

- Future works for framework's extension to other maintenance operations or asset's life-cycle phases:
  - Study operations knowledge domains and develop ontologies regarding their human-related tasks.
  - Investigate operational requirements for knowledge transfer and capture and implement content formats according to new datatypes (e.g. image, audio) for enabling automatic authoring.
  - Investigate operational requirements for knowledge re-use and implement new recommendation facets according to new contextual features (e.g. condition of components, etc.) for enabling interaction-analysis.
- Future works for framework's extension to other AR applications:
  - Study similar future works to those above but focused in application's human-related tasks (e.g. medical ultrasound diagnosis, tourism navigation, etc.).
  - Study viability of implementing a similar infrastructure that this framework proposes for AR application in terms of other AR technologies like tracking and registration.
  - Investigate the added value of this AR framework in terms of operational gains and implementation costs in the selected field of application.

It is worthy to note that as a result of this PhD thesis, an MSc group project (Wali *et al.*, 2020) has been conducted to study the proposed framework applicability to thermographic assessment operations.

6.3 Discussion on thesis real-world impact

This PhD thesis aimed to develop AR content-related methods for cost-effective human knowledge integration in digital maintenance. The combination of these methods can serve as basis of an AR framework (Figure 6-1) for industrial systems. For validation purposes, these methods have been implemented in a system prototype that is online within the network of the OpEx laboratory, where this thesis experiments were conducted. Therefore, this prototype can be assessed to determine the practical implications of this thesis contributions. Following subsections analyse the prototype's impact in future research works, research's industrial implementation and associated data security and privacy implications.

6.3.1 Impact in future research works

Previous Sections 3.7, 4.7, 5.7 and 6.2 already discussed future research works related to this thesis's contributions to knowledge, which are later summarised in Section 7.2. Besides proposing these works, the thesis system prototype can help to carry them out. This system prototype has already been utilised in two further research projects:

1. *"A design framework for Digital Twins"* (Erkoyuncu *et al.*, 2020): the system's prototype ontology-based information system was used to validate the Digital Twin design framework on a case study in manufacturing retrofitting.
2. *"A Framework for A Dynamic AR Based Degradation Assessment Digital Twin for Enhancing Maintenance-related Activities"* (Wali *et al.*, 2020): the system's prototype information system and its AR application based on the proposed authoring method were utilised for validating knowledge transfer effectiveness in degradation assessment tasks.

These projects have helped to demonstrate the proposed framework's applicability to future research works in knowledge transfer, capture and re-use applications. Although this thesis contributions do not directly involve personal data collection, future works do and so, it is necessary to consider their data privacy implications. These implications are later assessed in Subsection 6.3.3.

6.3.2 Impact in research's industrial implementation

Another area where this system prototype can have an impact involves this thesis commercial applications. This thesis industrial sponsor was involved on informal expert interviews regarding the system's prototype implementation design. These permitted to consider certain aspects for the prototype's industrial feasibility. There are three main aspects of the prototype's design which were influenced by these interviews:

1. Ergonomics: the industrial sponsor does maintenance of big, complex assets. Therefore, using tablet-based AR applications to support their operations may have caused interference with manual tasks. So, the prototype was designed assuming the use of Head-Mounted devices (HoloLens 1 and 2). However, the use of standard libraries to manage device interactions (MixedRealityToolkit) would allow for simple re-programming for hand-held devices.

2. Limited connectivity: the industrial sponsor aimed to utilise the outcomes of this thesis to support maintenance operations that normally occur in areas with limited connectivity. That is the main reason for the proposed framework to utilise cloud rather than web services. The prototype was designed to use JSON objects (3.5.1, 4.5.1 and 5.5.1) for transferring ontological entities between devices and store them locally. Thus, allowing AR and web browser applications (Figure 6-1) to run for limited periods of time without internet connection without replicating cloud server's ontological storage and inference capabilities.

3. Network security: the industrial sponsor belongs to the Defence Industry and so, it requires high standards for data security. Data security had an impact on the cloud implementation chosen for the prototype. Because the system prototype store case study received from the industrial sponsor, this could not be shared publicly. Therefore, the system prototype was implemented as a private cloud within the University's network (Subsection 3.5.1) to ensure that case study data was stored securely.

Besides informal interviews for implementation's design, the industrial sponsor also contributed with human and facility resources to conduct initial testing on the proposed AR framework in industrial environments. First impressions were positive regarding usability and impact in diagnosis efficiency and effectiveness. Nevertheless, an unexpected power shortage on sponsor's facilities did not allow to conduct more in-depth experiments within the thesis projected plan.

The laboratory's system prototype and initial tests conducted in industrial environments suggest that the proposed framework can be implemented as a commercial information system. According to NASA's Technology Readiness Level (TRL) (Mankins, 2009), initial tests in industrial environments position the system's prototype at TRL-7, with TRL-9 referring to commercial deployment. In order to close the gap between these levels, there are several routes to follow regarding future development works:

1. To develop the system as an independent one and introduce communication methods with existing information systems. These would include future works discussed above regarding translation techniques between ontology-based and other types (e.g. SQL) database systems.

2. To extend the system for developing a complete Computerised Maintenance Management System (CMMS). In the context of Industry 4.0, this would include future research works to integrate human knowledge in Digital Twins. This area of work includes the publication referred above (Erkoyuncu *et al.*, 2020) from which the thesis student is a co-author and which can be considered a resultant work from this thesis research.

3. To implement the system as a software library and publish it within other standard libraries in AR development like Mixed Reality Toolkit (Microsoft Corporation, 2020).

These routes can be complementary, and their selection depend on the business strategy determined to bridge the route to market. However, commercial software to be developed as a result of this thesis work should take into consideration the data security implications presented by the industrial sponsor. Their implications into the thesis work are assessed in the following subsection.

6.3.3 Data security and privacy implications

This thesis contributions relate to AR and ontology-based software for knowledge transfer, capture and re-use. AR technologies use sensors to track the real-world and their users and so, they can collect personal data (Rauschnabel, He and Ro, 2018). Ontology-based methods normally involve knowledge capture and can be considered to collect personal data only if captured knowledge can be traced back to its owner (Bertino *et al.*, 2006). Besides, it is usual for both, AR and ontology-based methods, to require internet connection and so, require features to avoid possible data security threats. Therefore, the student found relevant to assess the data security and privacy implications of this thesis contributions.

6.3.3.1 Data security

Data security relates to the protection of digital information from unauthorised uses and attacks (Bertino *et al.*, 2006). This thesis focused its research in ontology-based and AR content-related methods, which do not involve research in data security per se. However, this thesis research employed confidential and personal data for validation purposes that needed to be secured.

Personal data was obtained in paper forms to demographically assess participants who conducted experiments, interviews and surveys described in Subsections 3.5, 4.5 and 5.5. As per University's regulations based on GDPR directives, original paper forms were stored in locked areas and destroyed after 6 months of their creation. Their digital transcripts were anonymised and stored in encrypted devices.

Confidential data refers to part of the case study data utilised in Chapter 3 and Chapter 5, given by the industrial sponsor for validation purposes. As per their regulations, this digital data had to be stored in secured devices to prevent unauthorised access. The use of this data had an impact on the system prototype design (Figure 6-1). This resulted on the selection of a private cloud implementation within the University's network. It is accessible only by authorised researches for extended prototype use in other researches as explained in Subsection 3.5.1. Although this implementation can be considered sufficient for future research works, extended commercial applications (Subsection 6.3.2) may require the use of public networks and so, further measures to ensure data security. These measures include developing additional software features to ensure data security such as end-to-end encryption, secure coding, virtual private networks, etc.

6.3.3.2 Data privacy

Data privacy relates to the implications of collecting and using personal data and the rights of people to keep control over that (Rauschnabel, He and Ro, 2018). This thesis has two areas to consider regarding data privacy. One is how this thesis contributions and consequent future research developments make use of personal data. The other is the use of personal data collected as part of validation.

Validation procedures from Chapter 3, Chapter 4 and Chapter 5 collected personal data to assess this thesis contributions. It included demographic information (e.g. date of birth, years of experience, etc.) as well as performance measures from their experiments (e.g. completion time, errors, etc.). The assessment consisted of demographic and statistical analysis to understand the impact of this thesis contributions over pre-determined maintenance procedures. According to University's regulations based on GDPR directives, this thesis made use of that data in an anonymous manner. Besides, participants were given the chance to withdraw their data for a period of 6 months. The validation protocols, including data collection and analysis, were approved through the Cranfield University Research Ethics Systems (CURES) with reference numbers 2392/2017, 3623/2017, 3680/2018, 9203/2019, 9145/2019 and 9447/2019.

The main implications of this thesis regarding data privacy relate to its research contributions and consequent future works.

With respect to this thesis' contributions, it seems relevant to assess the types of personal data and the methods employed to collect and use it. The ontology-based knowledge capture methods proposed in Chapter 3 and Chapter 5 collect knowledge (data) from users. This captured knowledge mainly relates to assets, failure modes and diagnosis procedures and is collected anonymously. Because its anonymity, it cannot be considered to breach data privacy regulations as it cannot be traced back to the user from whom it was captured (Bertino *et al.*, 2006). Besides, the AR authoring methods proposed in Chapters 4 and 5 also collect data from users. This data includes the times when users asked for content to be created and preferences of content types to be created. With respect to the system prototype, content preferences are stored anonymously and deleted when the application is turned off. Content creation times are also stored anonymously and can serve to analyse user performance resultant from content usage. Future research and commercial implementations should consider the implications of analysing user performance as it can pose a risk to data privacy if not done anonymously (Rauschnabel, He and Ro, 2018).

With respect to this thesis' proposed future works, some presented in Chapter 4 and Chapter 5 suggest analysing user data to further enhance knowledge transfer, capture and re-use. For example, Subsection 4.7.2 proposes to utilise eye-tracking methods for enhancing augmented content adaptiveness through user expertise analysis. Besides, Subsection 5.7.2 proposes to utilise natural language processing methods to analyse data correctness for reducing reporting errors. The use of such methods can also pose privacy risks because they can reveal sensitive data (Liebling and Preibusch, 2014). Therefore, future research works should also consider the ethical and legal implications of analysing personal sensitive data.

Overall, it can be said that this thesis contributions avoid data privacy concerns because their anonymous treatment of data. Nevertheless, future research and commercial implementations should consider legal implications of non-anonymous data treatment as well as ethical ones of sensitive data analysis for avoiding any data privacy concerns.

Chapter 7
Thesis conclusions and future works

7.1 Conclusions

This PhD thesis hypothesised that advanced and automatic AR content-related techniques can transfer, capture and re-use knowledge to improve efficiency and effectiveness of diagnosis operations and decrease implementation costs for enhancing human knowledge in digital maintenance.

For proving such hypothesis, this PhD thesis has proposed:

A. Ontology-based reporting and monitoring techniques to enable knowledge capture and re-use for improving efficiency of diverse diagnosis tasks (Chapter 3).

B. A real-time, ontology-based and pattern-matching authoring technique to automate content creation for reducing AR implementation costs and gaining sufficient knowledge transfer effectiveness for improving operational efficiency and effectiveness of diverse maintenance operations (Chapter 4).

C. An ontology-based, context-aware interaction-analysis AR recommendation technique to enable knowledge re-use for improving efficiency and effectiveness of diagnosis reporting tasks in different contexts (Chapter 5).

Through diverse validation methods including stopwatch experiments, expert interviews and usability surveys, these techniques have proven that:

A.1. Ontology domains regarding human diagnosis tasks can be used to capture knowledge through reporting tools, improving the accuracy and structure of knowledge being captured. So, they can be used to advanced AR content-related techniques to enable enhanced knowledge re-use.

A.2. Knowledge captured in diagnosis reporting tasks can be re-used to improve efficiency of monitoring tasks. So, knowledge can be re-used in different operations than those in which was captured.

B.1. Authoring techniques can automate content-creation processes for reducing AR implementation costs. So, human knowledge integration can be done cost-effectively through AR.

B.2. Ontology-based authoring can transfer knowledge of multiple maintenance operations and maintain efficiency and effectiveness improvements. So, authoring automation does not affect expected AR benefits.

C.1. Context-aware and interaction-analysis recommender methods can re-use knowledge of diagnosis reporting tasks to improve their efficiency and effectiveness. So, knowledge can be re-used in similar operations that those in which was captured but in different contexts.

Altogether, these conclusions indicate that the thesis contributions stated above enable knowledge transfer, capture and re-use for improving effectiveness and efficiency of diagnosis tasks. Besides, they imply that augmented content for knowledge transfer, capture and re-use can be created and adapted automatically to reduce AR implementation costs. Thus, proving the validity of this thesis hypothesis regarding enhancements of human knowledge integration in digital maintenance. Moreover, these contributions help to fulfil important research gaps according to the literature reviews presented in this thesis:

A. Ontology-based reporting and monitoring methods increase structure and accuracy of knowledge capture to enhance integrated data management and to re-use it for improving complex fault-finding tasks.

B. The ontology-based, pattern-matching authoring method standardises and contextualises content-creation processes for diverse maintenance operations to facilitate integration of AR in maintenance information systems.

C. The ontology-based, context-aware, interaction-analysis AR recommendation method contextualises and standardises expert knowledge capture and re-use to enhance digitalisation of diagnosis reporting tasks.

Collectively, these contributions enable to create and adapt augmented content automatically as well as to input maintenance data and knowledge through web and AR applications. Hence, they can be considered to conform an AR framework for human knowledge integration in digital maintenance (Figure 6-1). This framework has been implemented as a system prototype and utilised for validation in subsequent research works including integrated data management for Digital Twins and AR-based degradation assessment. Besides, this framework's prototype has conducted initial tests in industrial environments (TRL-7) showing promising results in terms of usability and diagnosis efficiency improvements in real-life conditions. Thus, advancing the potential real-world impact of these contributions in terms of human knowledge integration in digital maintenance. Nevertheless, validation results presented in previous chapters (Sections 3.6, 4.6 and 5.6) have only demonstrated their benefits in laboratory conditions and for diagnosis operations. Also, specific discussions of these results (Section 6.2) regarding the thesis aim have indicated the limitations of this thesis contributions. They would require of future research and development works to obtain the full potential of human knowledge integration in digital maintenance and beyond.

7.2 Future works

The proposed framework and thesis contributions have demonstrated potential to improve efficiency and effectiveness of diagnosis operations by integrating human knowledge in digital systems. In the context of Industry 4.0, this could be utilised to achieve similar gains not only in other maintenance operations but also in other phases of assets' lifecycle. In order to achieve so and based on previous chapters discussions (Sections 3.7, 4.7, 5.7 and 6.2), this PhD thesis future works can be summarised as follows:

1. Enhance knowledge-domain models and AR content-related techniques to extend this framework's applicability to other maintenance operations:

a. Investigate knowledge-domain ontologies for new maintenance operations and relevant heterogeneous datatypes (e.g. audio or images) and develop advanced authoring methods for improving captured knowledge structure and accuracy.

b. Investigate operational requirements from user and environment and develop advanced context-awareness methods to automate contextual data collection (e.g. eye- or content-tracking) for enhancing knowledge transfer effectiveness.

c. Study relations among operation-domains ontologies and develop advanced interaction-analysis methods to improve maintenance context analysis for enhancing knowledge re-use among different operations.

2. Extend the proposed framework's applicability to other asset's lifecycle phases, Industry 4.0 processes and fields of AR application:

a. Investigate the applicability of AR as this framework proposes in terms of usability and infrastructure and develop advanced AR techniques like tracking for accommodating this framework to new cases of use.

b. Identify human-related tasks in new selected framework's applications (e.g. manufacturing retrofitting or medical surgery) and study the research declared above (1) to extend this framework's applicability to other operations.

3. Investigate real-world impact of the proposed framework and associated future research works required to achieve so:

a. Study monetary costs and gains of implementing the proposed framework and develop cost-analysis models to compare it with AR commercial systems.

b. Investigate the proposed framework integration with existing information systems and develop advanced methods for easing its real-life implementation such as translation mechanisms between ontology-based and other types of databases.

This PhD thesis has researched the impact of ontology-based AR content-related techniques for knowledge transfer, capture and re-use in maintenance diagnosis. Its contributions have shown potential to exchange heterogeneous data sources among digital systems and humans and vice versa to improve diagnosis efficiency and effectiveness for enhancing asset availability in a cost-effective manner. These contributions have also prepared the ground of a framework for human knowledge integration in digital information systems. Future research can reveal its full potential through integration with Digital Twins and other relevant technologies for improving assets availability and sustainment throughout their entire lifecycle.

References

Abele, L., Grimm, S., Zillner, S. and Kleinsteuber, M. (2014) 'An ontology-based approach for decentralized monitoring and diagnostics', in *Proceedings of the 12th IEEE International Conference on Industrial Informatics, INDIN 2014*, pp. 706–712. doi: 10.1109/indin.2014.6945600.

Adomavicius, G. and Tuzhilin, A. (2015) 'Context-aware recommender systems', in Ricci, F., Rokach, L., and Shapira, B. (eds) *Recommender Systems Handbook.* 2nd edn. Boston, MA, USA: Springer, pp. 191–226. doi: 10.1007/978-1-4899-7637-6_6.

Aggarwal, C. C. (2016) 'Knowledge-Based Recommender Systems', in Aggarwal, C. C. (ed.) *Recommender Systems: The Textbook.* 1st edn. New York, NY, USA: Springer, pp. 167–197. doi: 10.1007/978-3-319-29659-3_5.

Akbari, A., Setayeshmehr, A., Borsi, H., Gockenbach, E. and Fofana, I. (2010) 'Intelligent agent-based system using dissolved gas analysis to detect incipient faults in power transformers', *IEEE Electrical Insulation Magazine.* IEEE, 26(6), pp. 27–40. doi: 10.1109/MEI.2010.5599977.

Akbarinasaji, S. and Homayounvala, E. (2017) 'A novel context-aware augmented reality framework for maintenance systems', *Journal of Ambient Intelligence and Smart Environments*, 9(3), pp. 315–327. doi: 10.3233/AIS-170435.

Akçayir, M. and Akçayir, G. (2017) 'Advantages and challenges associated with augmented reality for education: A systematic review of the literature', *Educational Research Review*, 20, pp. 1–11. doi: 10.1016/j.edurev.2016.11.002.

Alam, M. F., Katsikas, S., Beltramello, O. and Hadjiefthymiades, S. (2017) 'Augmented and virtual reality based monitoring and safety system: A prototype IoT platform', *Journal of Network and Computer Applications*, 89, pp. 109–119. doi: 10.1016/j.jnca.2017.03.022.

Angelopoulos, A., Michailidis, E. T., Nomikos, N., Trakadas, P., Hatziefremidis, A., Voliotis, S. and Zahariadis, T. (2020) 'Tackling faults in the industry 4.0 era—a survey of machine-learning solutions and key aspects', *Sensors (Switzerland)*, 20(1), pp. 1–34. doi: 10.3390/s20010109.

Aruna, T., Saranya, K. and Bhandari, C. (2011) 'A survey on ontology evaluation tools', in *Proceedings of the 2011 International Conference on Process Automation, Control and Computing*. Coimbatore, India: IEEE. doi: 10.1109/PACC.2011.5978931.

Avalle, G., De Pace, F., Fornaro, C., Manuri, F. and Sanna, A. (2019) 'An Augmented Reality System to Support Fault Visualization in Industrial Robotic Tasks', *IEEE Access*. IEEE, 7, pp. 132343–132359. doi: 10.1109/ACCESS.2019.2940887.

Azuma, R. T. (2016) 'The Most Important Challenge Facing Augmented Reality', *Presence: Teleoperators and Virtual Environments*, 25(3), pp. 234–238. doi: 10.1162/PRES_a_00264.

Bacca, J., Baldiris, S., Fabregat, R. and Graf, S. (2014) 'Augmented Reality Trends in Education: A Systematic Review of Research and Applications', *Educational Technology & Society*, 17(4), pp. 133–149. doi: ISSN 1436-4522 (online).

Barrasa, J. (2019) 'neosemantics'. Available at: https://github.com/neo4j-labs/neosemantics.

Bautista-Zambrana, M. R. (2015) 'Methodologies to Build Ontologies for Terminological Purposes', in *Procedia - Social and Behavioral Sciences*. Elsevier B.V., pp. 264–269. doi: 10.1016/j.sbspro.2015.02.063.

Becerra-Fernandez, I. and Sabherwal, R. (2010) *Knowledge Management: Systems and Processes*. 1st edn, *M.E. Sharpe, Inc.* 1st edn. Armonk, NY: M.E. Sharpe. Available at: https://ethio.wikispaces.com/file/view/book-KM-Systems+and+Processes.pdf.

Behravan, A., Meckel, S. and Obermaisser, R. (2019) 'Generic Fault-Diagnosis Strategy based on Diagnostic Directed Acyclic Graphs using Domain Ontology in Automotive Applications', in *Proceedings of the 10th GMM Symposium*. Dortmund: VDE VERLAG GMBH, Berlin, pp. 95–99. Available at: https://ieeexplore.ieee.org/document/8727850.

Bekkaoui, M., Karray, M. H., Bekaddour, F. and Meliani, S. M. (2017) 'Semantic correlation for alarms classification in maintenance process', *EEA - Electrotehnica, Electronica, Automatica*, 65(3).

Bekkaoui, M., Karray, M. H. and Sari, Z. (2015) 'Knowledge formalization for experts' selection into a collaborative maintenance platform', *IFAC-PapersOnLine*. Elsevier Ltd., 28(3), pp. 1445–1450. doi: 10.1016/j.ifacol.2015.06.290.

Bernasconi, A., Kharshiduzzaman, M., Anodio, L. F., Bordegoni, M., Re, G. M., Braghin, F. and Comolli, L. (2014) 'Development of a monitoring system for crack growth in bonded single-lap joints based on the strain field and visualization by augmented reality', *Journal of Adhesion*, 90(5–6), pp. 496–510. doi: 10.1080/00218464.2013.857606.

Bernhardt, S., Nicolau, S. A., Soler, L. and Doignon, C. (2017) 'The status of augmented reality in laparoscopic surgery as of 2016', *Medical Image Analysis*. Elsevier B.V., 37, pp. 66–90. doi: 10.1016/j.media.2017.01.007.

Bertino, E., Khan, L. R., Sandhu, R. and Thuraisingham, B. (2006) 'Secure knowledge management: Confidentiality, trust, and privacy', *IEEE Transactions on Systems, Man, and Cybernetics Part A:Systems and Humans*, 36(3), pp. 429–438. doi: 10.1109/TSMCA.2006.871796.

Bhattacharya, B. and Winer, E. H. (2019) 'Augmented reality via expert demonstration authoring (AREDA)', *Computers in Industry*. Elsevier B.V., 105, pp. 61–79. doi: 10.1016/j.compind.2018.04.021.

Blattgerste, J., Renner, P. and Pfeiffer, T. (2019) 'Authorable augmented reality instructions for assistance and training in work environments', *ACM International Conference Proceeding Series*. doi: 10.1145/3365610.3365646.

Bleser, G., Damen, D., Behera, A., Hendeby, G., Mura, K., Miezal, M., Gee, A., Petersen, N., Maçães, G., Domingues, H., Gorecky, D., Almeida, L., Mayol-Cuevas, W., Calway, A., Cohn, A. G., Hogg, D. C. and Stricker, D. (2015) 'Cognitive learning, monitoring and assistance of industrial workflows using egocentric sensor networks', *PLoS ONE*, 10(6), pp. 1–41. doi: 10.1371/journal.pone.0127769.

Booth, A., Papaioannou, D. and Sutton, A. (2012) *Systematic approaches to a successful literature review.*, *Sage*. Thousand Oaks, CA, USA: Sage Publications.

Bottani, E. and Vignali, G. (2019) 'Augmented reality technology in the manufacturing industry: A review of the last decade', *IISE Transactions*. Taylor & Francis, 51(3), pp. 284–310. doi: 10.1080/24725854.2018.1493244.

Brank, J., Grobelnik, M. and Mladenić, D. (2005) 'A survey of ontology evaluation techniques', in *Proceedings of 8th Int. multi-conf. Information Society*, pp. 166–170. doi: 10.1.1.101.4788.

Breitman, K., Casanova, M. A. and Truszkowski, W. (2007) *Semantic Web: Concepts, Technologies and Applications*, *Springer-Verlag*. London, UK: Springer-Verlag. doi: 10.1007/978-1-84628-710-7.

Candell, O., Karim, R. and Söderholm, P. (2009) 'eMaintenance-Information logistics for maintenance support', *Robotics and Computer-Integrated Manufacturing*, 25(6), pp. 937–944. doi: 10.1016/j.rcim.2009.04.005.

Cao, Y., Wang, T., Qian, X., Rao, P. S., Wadhawan, M., Huo, K. and Ramani, K. (2019) 'GhostAR: A time-space editor for embodied authoring of human-robot collaborative task with augmented reality', *UIST 2019 - Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology*, pp. 521–534. doi: 10.1145/3332165.3347902.

Capgemini Research Institute (2018) *Augmented and Virtual Reality in Operations*, *Capgemini Research Institute*. Available at: https://www.capgemini.com/research/augmented-and-virtual-reality-in-operations/.

Carrera, Á., Iglesias, C. A., García-Algarra, J. and Kolařík, D. (2014) 'A real-life application of multi-agent systems for fault diagnosis in the provision of an Internet business service', *Journal of Network and Computer Applications*, 37(1), pp. 146–154. doi: 10.1016/j.jnca.2012.11.004.

Ceruti, A., Liverani, A. and Bombardi, T. (2017) 'Augmented vision and interactive monitoring in 3D printing process', *International Journal on Interactive Design and Manufacturing*. Springer Paris, 11(2), pp. 385–395. doi: 10.1007/s12008-016-0347-y.

Chang, M. M. L., Nee, A. Y. C. and Ong, S. K. (2020) 'Interactive AR-assisted product disassembly sequence planning (ARDIS)', *International Journal of Production Research*. Taylor & Francis, 0(0), pp. 1–16. doi: 10.1080/00207543.2020.1730462.

Chi, H. L., Chen, Y. C., Kang, S. C. and Hsieh, S. H. (2012) 'Development of user interface for tele-operated cranes', *Advanced Engineering Informatics*. Elsevier Ltd, 26(3), pp. 641–652. doi: 10.1016/j.aei.2012.05.001.

Chi, H. L., Kang, S. C. and Wang, X. (2013) 'Research trends and opportunities of augmented reality applications in architecture, engineering, and construction', *Automation in Construction*. Elsevier B.V., 33, pp. 116–122. doi: 10.1016/j.autcon.2012.12.017.

Chou, C.-C., Chiang, C.-T., Wu, P.-Y., Chu, C.-P. and Lin, C.-Y. (2016) 'Spatiotemporal analysis and visualization of power consumption data integrated with building information models for energy savings', *Resources, Conservation and Recycling*. Elsevier B.V., 123, pp. 219–229. doi: 10.1016/j.resconrec.2016.03.008.

Cimino, C., Negri, E. and Fumagalli, L. (2019) 'Review of digital twin applications in manufacturing', *Computers in Industry*. Elsevier B.V., 113, pp. 103–130. doi: 10.1016/j.compind.2019.103130.

Cohen, J. (1992) 'A power primer', *Physicological Bulletin*, 112(1), pp. 155–159. doi: 10.1037/0033-2909.112.1.155.

Corrêa dos Santos, A. C., Delamaro, M. E. and Nunes, F. L. S. (2013) 'The Relationship between Requirements Engineering and Virtual Reality Systems: A Systematic Literature Review', in *Virtual and Augmented Reality (SVR), 2013 XV Symposium on*, pp. 53–62. doi: 10.1109/SVR.2013.52.

Cullot, N., Ghawi, R. and Yétongnon, K. (2007) 'DB2OWL : A Tool for Automatic Database-to-Ontology Mapping', *Proceedings of the Fifteenth Italian Symposium on Advanced Database Systems, SEBD*, pp. 1–4.

Das, M. P., Dong, Z. and Scherer, S. (2018) 'Joint Point Cloud and Image Based Localization for Efficient Inspection in Mixed Reality', *IEEE International Conference on Intelligent Robots and Systems*, pp. 6357–6363. doi: 10.1109/IROS.2018.8594318.

Dendani, N., Khadir, M. and Guessoum, S. (2012) 'Hybrid approach for fault diagnosis based on CBR and ontology: Using jCOLIBRI framework', in *International Conference on Complex Systems*. IEEE, pp. 1–8. doi: 10.1109/ICoCS.2012.6458509.

Dey, A., Billinghurst, M., Lindeman, R. W. and Swan, J. E. (2018) 'A systematic review of 10 Years of Augmented Reality usability studies: 2005 to 2014', *Frontiers Robotics AI*, 5(APR). doi: 10.3389/frobt.2018.00037.

Dibowski, H., Holub, O. and Rojíček, J. (2017) 'Ontology-based fault propagation in building automation systems', *International Journal of Simulation: Systems, Science and Technology*, 18(3), pp. 1.1-1.14. doi: 10.5013/IJSSST.a.18.03.01.

Dini, G. and Dalle Mura, M. (2015) 'Application of Augmented Reality Techniques in Through-life Engineering Services', *Procedia CIRP*. Elsevier B.V., 38, pp. 14–23. doi: 10.1016/j.procir.2015.07.044.

Djordjevic, L., Petrovic, N. and Tosic, M. (2019) 'Ontology based approach to development of augmented reality applications', *Proceedings of the 27th Telecommunications Forum, TELFOR 2019*, pp. 0–3. doi: 10.1109/TELFOR48224.2019.8971208.

Dong, S., Feng, C. and Kamat, V. R. (2013) 'Sensitivity analysis of augmented reality-assisted building damage reconnaissance using virtual prototyping', *Automation in Construction*. Elsevier B.V., 33, pp. 24–36. doi: 10.1016/j.autcon.2012.09.005.

Ebrahimipour, V. and Yacout, S. (2015) 'Ontology-based schema to support maintenance knowledge representation with a case study of a pneumatic valve', *IEEE Transactions on Systems, Man, and Cybernetics: Systems*. IEEE, 45(4), pp. 702–712. doi: 10.1109/TSMC.2014.2383361.

Eernisse, M. (2015) 'EJS'.

Egger, J. and Masood, T. (2020) 'Augmented reality in support of intelligent manufacturing – A systematic literature review', *Computers and Industrial Engineering*. Elsevier, 140(May 2019), p. 106195. doi: 10.1016/j.cie.2019.106195.

ElSayed, N. A. M., Thomas, B. H., Marriott, K., Piantadosi, J. and Smith, R. T. (2016) 'Situated Analytics: Demonstrating immersive analytical tools with Augmented Reality', *Journal of Visual Languages and Computing*. Elsevier, 36, pp. 13–23. doi: 10.1016/j.jvlc.2016.07.006.

Erkoyuncu, J. A., Fernández del Amo, I., Ariansyaha, D., Bulka, D., Vrabič, R. and Roy, R. (2020) 'A Design Framework for Adaptive Digital Twins', *CIRP Annals - Manufacturing Technology*, 00, pp. 8–11. doi: 10.1016/j.cirp.2020.04.086.

Erkoyuncu, J. A., Fernández del Amo, I., Dalle Mura, M., Roy, R. and Dini, G. (2017) 'Improving efficiency of industrial maintenance with context aware adaptive authoring in augmented reality', *CIRP Annals - Manufacturing Technology*. doi: 10.1016/j.cirp.2017.04.006.

Espíndola, D. B., Fumagalli, L., Garetti, M., Pereira, C. E., Botelho, S. S. C. and Ventura Henriques, R. (2013) 'A model-based approach for data integration to improve maintenance management by mixed reality', *Computers in Industry*, 64(4), pp. 376–391. doi: 10.1016/j.compind.2013.01.002.

Fang, H. C., Ong, S. K. and Nee, A. Y. C. (2012) 'Interactive robot trajectory planning and simulation using augmented reality', *Robotics and Computer-Integrated Manufacturing*. Elsevier, 28(2), pp. 227–237. doi: 10.1016/j.rcim.2011.09.003.

Fernández-López, M., Gómez-Pérez, A. and Juristo, N. (1997) 'METHONTOLOGY: From Ontological Art Towards Ontological Engineering', in *Proceedings of the 1997 AAAI Spring Symposium Series*. Stanford, CAL, EEUU, pp. 33–40. doi: 10.1109/AXMEDIS.2007.19.

Fernández Del Amo, I., Erkoyuncu, J. A., Roy, R., Palmarini, R. and Onoufriou, D. (2018) 'A systematic review of Augmented Reality content-related techniques for knowledge transfer in maintenance applications', *Computers in Industry*, 103, pp. 47–71. doi: 10.1016/j.compind.2018.08.007.

Fernández del Amo, I., Erkoyuncu, J., Vrabič, R., Frayssinet, R., Vazquez Reynel, C. and Roy, R. (2019) 'Structured authoring for AR-based communication to enhance efficiency in remote diagnosis for complex equipment', *Advanced Engineering Informatics*. Elsevier, 45(March), p. 15. doi: 10.1016/j.aei.2020.101096.

Ferrari, R., Dibowski, H. and Baldi, S. (2017) 'A Message Passing Algorithm for Automatic Synthesis of Probabilistic Fault Detectors from Building Automation Ontologies', *IFAC-PapersOnLine*. Elsevier B.V., 50(1), pp. 4184–4190. doi: 10.1016/j.ifacol.2017.08.809.

Fiorentino, M. ., Uva, A. E. ., Monno, G. . and Radkowski, R. . (2016) 'Natural interaction for online documentation in industrial maintenance', *International Journal of Computer Aided Engineering and Technology*, 8(1–2), pp. 56–79. doi: 10.1504/IJCAET.2016.073269.

Fiorentino, M., Uva, A. E., Gattullo, M., Debernardis, S. and Monno, G. (2014) 'Augmented reality on large screen for interactive maintenance instructions', *Computers in Industry*. Elsevier B.V., 65(2), pp. 270–278. doi: 10.1016/j.compind.2013.11.004.

Flatt, H., Koch, N., Röcker, C., Günter, A. and Jasperneite, J. (2015) 'A context-aware assistance system for maintenance applications in smart factories based on augmented reality and indoor localization', in *International Conference on Emerging Technologies and Factory Automation*. Luxembourg City, Luxembourg 8-11 September 2015: IEEE, pp. 8–11. doi: 10.1109/ETFA.2015.7301586.

Flotyński, J. and Walczak, K. (2015) 'Conceptual knowledge-based modeling of interactive 3D content', *Visual Computer*, 31(10), pp. 1287–1306. doi: 10.1007/s00371-014-1011-9.

Flotyński, J. and Walczak, K. (2017) 'Ontology-Based Representation and Modelling of Synthetic 3D Content: A State-of-the-Art Review', *Computer Graphics Forum*, 36(8), pp. 329–353. doi: 10.1111/cgf.13083.

Franceschini, F., Galetto, M., Maisano, D. and Mastrogiacomo, L. (2016) 'Towards the use of augmented reality techniques for assisted acceptance sampling', *Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture*, 230(10), pp. 1870–1884. doi: 10.1177/0954405415624360.

García-Peñalvo, F. J., Colomo-Palacios, R., García, J. and Therón, R. (2012) 'Towards an ontology modeling tool. A validation in software engineering scenarios', *Expert Systems with Applications*, 39(13), pp. 11468–11478. doi: 10.1016/j.eswa.2012.04.009.

García, J., García-Peñalvo, F. J. and Therón, R. (2010) 'A survey on ontology metrics', in *Communications in Computer and Information Science*. Springer, Berlin, Heidelberg, pp. 22–27. doi: 10.1007/978-3-642-16318-0_4.

Gattullo, M., Scurati, G. W., Fiorentino, M., Uva, A. E., Ferrise, F. and Bordegoni, M. (2019) 'Towards augmented reality manuals for industry 4.0: A methodology', *Robotics and Computer-Integrated Manufacturing*. Elsevier Ltd, 56(March 2018), pp. 276–286. doi: 10.1016/j.rcim.2018.10.001.

Ghimire, R., Pattipati, K. R. and Luh, P. B. (2016) 'Fault diagnosis and augmented reality-based troubleshooting of HVAC systems', *AUTOTESTCON (Proceedings)*. IEEE, 2016-Octob, pp. 1–10. doi: 10.1109/AUTEST.2016.7589590.

Gimeno, J., Morillo, P., Orduña, J. M. and Fernández, M. (2013) 'A new AR authoring tool using depth maps for industrial procedures', *Computers in Industry*, 64(9), pp. 1263–1271. doi: 10.1016/j.compind.2013.06.012.

Golański, P., Perz-Osowska, M. and Szczekala, M. (2014) 'A Demonstration Model of a Mobile Expert System with Augmented Reality User Interface Supporting M-28 Aircraft Maintenace/Prototyp Przenośnego-Ekspertowego Systemu Z Rzeczywistością Rozszerzoną Do Obsługi Samolotu M-28', *Journal of KONBiN*, 31(1), pp. 23–31. doi: 10.2478/jok-2014-0020.

Golparvar-Fard, M. and Ham, Y. (2014) 'Automated diagnostics and visualization of potential energy performance problems in existing buildings using energy performance augmented reality models', *Journal of Computing Civil Engineering*, 28(1), pp. 17–29. doi: 10.1061/(ASCE)CP.1943-5487.0000311.

Gómez-Pérez, A. (2004) 'Ontology Evaluation', in Staab, S. and Studer, R. (eds) *Handbook on Ontologies, International Handbooks on Information Systems.* Karlsruhe: Springer Berlin Heidelberg, pp. 251–273. doi: 10.1007/978-3-540-24750-0_13.

Gong, Y. and Janssen, M. (2013) 'An interoperable architecture and principles for implementing strategy and policy in operational processes', *Computers in Industry*. Elsevier B.V., 64(8), pp. 912–924. doi: 10.1016/j.compind.2013.06.018.

Grant, M. J. and Booth, A. (2009) 'A typology of reviews: An analysis of 14 review types and associated methodologies', *Health Information and Libraries*

Journal, 26(2), pp. 91–108. doi: 10.1111/j.1471-1842.2009.00848.x.

Gunawardana, A. and Shani, G. (2015) ‘Evaluating recommender systems’, in Ricci, F., Rokach, L., and Shapira, B. (eds) *Recommender Systems Handbook*. 2nd edn. Boston, MA, USA: Springer, pp. 265–308. doi: 10.1007/978-1-4899-7637-6_8.

Hadar, E., Shtok, J., Cohen, B., Tzur, Y. and Karlinsky, L. (2017) ‘Hybrid remote expert - An emerging pattern of industrial remote support’, in *CEUR Workshop Proceedings*. CEUR-WS, pp. 33–40.

Hall, S. and Takahashi, R. (2017) ‘Augmented and virtual reality: The promise and peril of immersive technologies’, *McKinsey&Company*, (October), p. 9. Available at: https://www.mckinsey.com/industries/technology-media-and-telecommunications/our-insights/augmented-and-virtual-reality-the-promise-and-peril-of-immersive-technologies.

Ham, Y. and Golparvar-Fard, M. (2013) ‘EPAR: Energy Performance Augmented Reality models for identification of building energy performance deviations between actual measurements and simulation results’, *Energy and Buildings*, 63, pp. 15–28. doi: 10.1016/j.enbuild.2013.02.054.

Hart, S. G. (2006) ‘NASA-task load index (NASA-TLX); 20 years later’, *Proceedings of the Human Factors and Ergonomics Society*, pp. 904–908. doi: 10.1177/154193120605000909.

He, F., Ong, S. K. and Nee, A. Y. C. (2019) ‘A mobile solution for augmenting a manufacturing environment with user-generated annotations’, *Information (Switzerland)*, 10(2). doi: 10.3390/info10020060.

Hejlsberg, A. (2011) *The C# programming language*. Addison-Wesley.

Henriquez, P., Alonso, J. B., Ferrer, M. A. and Travieso, C. M. (2014) ‘Review of automatic fault diagnosis systems using audio and vibration signals’, *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 44(5), pp. 642–652. doi: 10.1109/TSMCC.2013.2257752.

Higgins, J. and Green, S. (2008) *Cochrane Handbook for Systematic Cochrane Handbook for Systematic Reviews of*. 1st edn, *Wiley-Blackwell*. 1st edn. Edited by J. Higgins and S. Green. Chichester, England: Wiley-Blackwell. doi: 10.1002/9780470712184.

Hlomani, H. and Stacey, D. (2014) ‘Approaches, methods, metrics, measures, and subjectivity in ontology evaluation: A survey’, *Semantic Web Journal*, 1, pp. 1–11. Available at: http://www.semantic-web-journal.net/system/files/swj657.pdf.

Hou, L., Wang, Y., Wang, X., Maynard, N., Cameron, I., Zhang, S. and Maynard, Y. (2014) 'Combining photogrammetry and augmented reality towards an integrated facility management system for the oil industry', *Proceedings of the IEEE*, 102(2), pp. 204–220. doi: 10.1109/JPROC.2013.2295327.

Huang, J. M., Ong, S. K. and Nee, A. Y. C. (2017) 'Visualization and interaction of finite element analysis in augmented reality', *CAD Computer Aided Design*. Elsevier Ltd, 84, pp. 1–14. doi: 10.1016/j.cad.2016.10.004.

Irizarry, J., Gheisari, M., Williams, G. and Walker, B. N. (2013) 'InfoSPOT: A mobile Augmented Reality method for accessing building information through a situation awareness approach', *Automation in Construction*, 33, pp. 11–23. doi: 10.1016/j.autcon.2012.09.002.

Jardine, A. K. S., Lin, D. and Banjevic, D. (2006) 'A review on machinery diagnostics and prognostics implementing condition-based maintenance', *Mechanical Systems and Signal Processing*, 20(7), pp. 1483–1510. doi: 10.1016/j.ymssp.2005.09.012.

Javier, I., Masoud, G., Graceline, W. and Kathy, R. (2014) 'Ambient intelligence environments for accessing building information', *Facilities*, 32(3/4), pp. 120–138. doi: 10.1108/F-05-2012-0034.

Jiao, Y., Zhang, S., Li, Y., Wang, Y. and Yang, B. (2013) 'Towards cloud Augmented Reality for construction application by BIM and SNS integration', *Automation in Construction*. Elsevier B.V., 33, pp. 37–47. doi: 10.1016/j.autcon.2012.09.018.

Jin, H., Ibrahim, S., Bell, T., Gao, W., Huang, D. and Wu, S. (2010) 'Cloud Types and Services', in Furht, B. and Escalante, A. (eds) *Handbook of Cloud Computing*. Boston, MA, USA: Springer, pp. 335–355. doi: 10.1007/978-1-4419-6524-0_14.

Khadir, M. T. and Klai, S. (2010) 'A steam turbine diagnostic maintenance system based on an evolutive domain ontology', in *ICMWI 2010 - Proceedings*. IEEE, pp. 360–367. doi: 10.1109/ICMWI.2010.5648079.

Khalil, M., Bergs, C., Papadopoulos, T., Wuchner, R., Bletzinger, K. U. and Heizmann, M. (2019) 'IIoT-based fatigue life indication using augmented reality', *IEEE International Conference on Industrial Informatics (INDIN)*, 2019-July, pp. 746–751. doi: 10.1109/INDIN41052.2019.8972114.

Khan, S., Phillips, P., Jennions, I. and Hockley, C. (2014) 'No Fault Found events in maintenance engineering Part 1: Current trends, implications and organizational practices', *Reliability Engineering and System Safety*. Elsevier, 123, pp. 183–195. doi: 10.1016/j.ress.2013.11.003.

Khan, S. and Yairi, T. (2018) 'A review on the application of deep learning in system health management', *Mechanical Systems and Signal Processing*. Elsevier Ltd, 107, pp. 241–265. doi: 10.1016/j.ymssp.2017.11.024.

Kim, C., Park, T., Lim, H. and Kim, H. (2013) 'On-site construction management using mobile computing technology', *Automation in Construction*, 35, pp. 415–423. doi: 10.1016/j.autcon.2013.05.027.

Kitchenham, B. (2004) *Procedures for performing systematic reviews*, *Keele University*. Keele, UK. Available at: http://www.inf.ufsc.br/~aldo.vw/kitchenham.pdf.

Kong, X. T. R., Luo, H., Huang, G. Q. and Yang, X. (2019) 'Industrial wearable system: the human-centric empowering technology in Industry 4.0', *Journal of Intelligent Manufacturing*. Springer US, 30(8), pp. 2853–2869. doi: 10.1007/s10845-018-1416-9.

Kwon, O. S., Park, C. S. and Lim, C. R. (2014) 'A defect management system for reinforced concrete work utilizing BIM, image-matching and augmented reality', *Automation in Construction*. Elsevier B.V., 46, pp. 74–81. doi: 10.1016/j.autcon.2014.05.005.

Lakshantha, E. and Egerton, S. (2016) 'A diagrammatic framework for intuitive human robot interaction', *Journal of Ambient Intelligence and Smart Environments*, 8(1), pp. 21–33. doi: 10.3233/AIS-150358.

Lambrecht, J., Kleinsorge, M., Rosenstrauch, M. and Krüger, J. (2013) 'Spatial programming for industrial robots through task demonstration', *International Journal of Advanced Robotic Systems*, 10. doi: 10.5772/55640.

Laroche, F., Dhuieb, M. A., Belkadi, F. and Bernard, A. (2016) 'Accessing enterprise knowledge: A context-based approach', *CIRP Annals - Manufacturing Technology*. CIRP, 65(1), pp. 189–192. doi: 10.1016/j.cirp.2016.04.136.

Lee, S. G., Ma, Y.-S., Thimm, G. L. and Verstraeten, J. (2008) 'Product lifecycle management in aviation maintenance, repair and overhaul', *Computers in industry*. Elsevier, 59(2), pp. 296–303. doi: 10.1016/j.compind.2007.06.022.

Lee, W. H. and Lee, H. K. (2016) 'The usability attributes and evaluation measurements of mobile media AR (augmented reality)', *Cogent Arts and Humanities*. Cogent, 3(1). doi: 10.1080/23311983.2016.1241171.

Li, D., Fast-Berglund, Å. and Paulin, D. (2019) 'Current and future Industry 4.0 capabilities for information and knowledge sharing: Case of two Swedish SMEs', *International Journal of Advanced Manufacturing Technology.* The International Journal of Advanced Manufacturing Technology, 105(9), pp. 3951–3963. doi: 10.1007/s00170-019-03942-5.

Li, W., Nee, A. and Ong, S. K. (2017) 'A State-of-the-Art Review of Augmented Reality in Engineering Analysis and Simulation', *Multimodal Technologies and Interaction*, 1(3), p. 17. doi: 10.3390/mti1030017.

Liebling, D. J. and Preibusch, S. (2014) 'Privacy considerations for a pervasive eye tracking world', *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, pp. 1169–1177. doi: 10.1145/2638728.2641688.

Lin, P. H. and Chen, S. Y. (2020) 'Design and Evaluation of a Deep Learning Recommendation Based Augmented Reality System for Teaching Programming and Computational Thinking', *IEEE Access.* IEEE, 8, pp. 45689–45699. doi: 10.1109/ACCESS.2020.2977679.

Liu, B., Ding, Z., Wu, J. and Yao, L. (2019) 'Ontology-based fault diagnosis: A decade in review', *ACM International Conference Proceeding Series*, pp. 112–116. doi: 10.1145/3307363.3307381.

Liu, C., Cao, S., Tse, W. and Xu, X. (2017) 'Augmented Reality-assisted Intelligent Window for Cyber-Physical Machine Tools', *Journal of Manufacturing Systems*, 44(2), pp. 280–286. doi: 10.1016/j.jmsy.2017.04.008.

Liu, Y., Li, S. Q., Wang, J. F., Zeng, H. and Lu, J. P. (2014) 'A computer vision-based assistant system for the assembly of narrow cabin products', *International Journal of Advanced Manufacturing Technology*, 76(1–4), pp. 281–293. doi: 10.1007/s00170-014-6274-9.

Liu, Y., Li, S. and Wang, J. (2014) 'Interactive operation of physically-based slender flexible parts in an augmented reality environment', *Science China Technological Sciences*, 57(7), pp. 1383–1391. doi: 10.1007/s11431-014-5522-4.

Liyanage, C., Elhag, T., Ballal, T. and Li, Q. (2009) 'Knowledge communication and translation – a knowledge transfer model', *Journal of Knowledge Management*, 13(3), pp. 118–131. doi: 10.1108/13673270910962914.

Longo, F., Nicoletti, L. and Padovano, A. (2019) 'Ubiquitous knowledge empowers the Smart Factory: The impacts of a Service-oriented Digital Twin on enterprises' performance', *Annual Reviews in Control.* Elsevier Ltd, 47, pp. 221–

236. doi: 10.1016/j.arcontrol.2019.01.001.

van Lopik, K., Sinclair, M., Sharpe, R., Conway, P. and West, A. (2020) ‘Developing augmented reality capabilities for industry 4.0 small enterprises: Lessons learnt from a content authoring case study’, *Computers in Industry*. Elsevier B.V., 117, p. 103208. doi: 10.1016/j.compind.2020.103208.

Makris, S., Karagiannis, P., Koukas, S. and Matthaiakis, A.-S. (2016) ‘Augmented reality system for operator support in human–robot collaborative assembly’, *CIRP Annals - Manufacturing Technology*, 65(1), pp. 61–64. doi: 10.1016/j.cirp.2016.04.038.

Makris, S., Pintzos, G., Rentzos, L. and Chryssolouris, G. (2013) ‘Assembly support using AR technology based on automatic sequence generation’, *CIRP Annals - Manufacturing Technology*. CIRP, 62(1), pp. 9–12. doi: 10.1016/j.cirp.2013.03.095.

Mankins, J. C. (2009) ‘Technology readiness assessments: A retrospective’, *Acta Astronautica*. Elsevier, 65(9–10), pp. 1216–1223. doi: 10.1016/j.actaastro.2009.03.058.

Manuri, F. and Sanna, A. (2016) ‘A Survey on Applications of Augmented Reality’, *Advances in Computer Science*, 5(1), pp. 18–27. doi: 10.17485/ijst/2016/v9i2/86369.

De Marchi, L., Ceruti, A., Marzani, A. and Liverani, A. (2013) ‘Augmented reality to support on-field post-impact maintenance operations on thin structures’, *Journal of Sensors*, 2013. doi: 10.1155/2013/619570.

Markus, M. L. (2001) ‘Toward a theory of knowledge reuse: Types of knowledge reuse situations and factors in reuse success’, *Journal of Management Information Systems*, 18(1), pp. 57–93. doi: 10.1080/07421222.2001.11045671.

Martínez, H., Fabry, T., Laukkanen, S., Mattila, J. and Tabourot, L. (2014) ‘Augmented reality aiding collimator exchange at the LHC’, *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 763, pp. 354–363. doi: 10.1016/j.nima.2014.06.037.

Masoni, R., Ferrise, F., Bordegoni, M., Gattullo, M., Uva, A. E., Fiorentino, M., Carrabba, E. and Di Donato, M. (2017) ‘Supporting Remote Maintenance in Industry 4.0 through Augmented Reality’, *Procedia Manufacturing*. doi: 10.1016/j.promfg.2017.07.257.

Masood, T., Roy, R., Harrison, A., Xu, Y., Gregson, S. and Reeve, C. (2015)

‘Integrating through-life engineering service knowledge with product design and manufacture’, *International Journal of Computer Integrated Manufacturing*. Taylor & Francis, 28(1), pp. 59–74. doi: 10.1080/0951192X.2014.900858.

McDaniel, M. and Storey, V. C. (2019) ‘Evaluating domain ontologies: Clarification, classification, and challenges’, *ACM Computing Surveys*, 52(4), pp. 70:2-70:44. doi: 10.1145/3329124.

McGurk, S., Abela, C. and Debattista, J. (2017) ‘Towards Ontology Quality Assessment’, in *Proceedings of the 4th Workshop on Linked Data Quality*. Available at: http://ceur-ws.org/Vol-1824/ldq_paper_2.pdf.

Medina-Oliva, G., Voisin, A., Monnin, M. and Leger, J. B. (2014) ‘Predictive diagnosis based on a fleet-wide ontology approach’, *Knowledge-Based Systems*. Elsevier B.V., 68, pp. 40–57. doi: 10.1016/j.knosys.2013.12.020.

Mehta, R. (2015) *UK Support and Service industry a high value employer and a net exporter*, *TES Centre*. Cranfield, United Kingdom. Available at: https://www.through-life-engineering-services.org/downloads/Final_UK_Support_and_Service_Industry.pdf.

Meža, S., Turk, Ž. and Dolenc, M. D. (2014) ‘Component based engineering of a mobile BIM-based augmented reality system’, *Automation in Construction*, 42, pp. 1–12. doi: 10.1016/j.autcon.2014.02.011.

Microsoft Corporation (2019) ‘Visual Studio’, *https://visualstudio.microsoft.com*. Available at: https://visualstudio.microsoft.com.

Microsoft Corporation (2020) *Mixed Reality Toolkit Documentation*, *Microsoft Corporation*. Available at: https://microsoft.github.io/MixedRealityToolkit-Unity/Documentation/GettingStartedWithTheMRTK.html (Accessed: 29 April 2020).

Miguelañez, E., Brown, K. E., Lewis, R., Roberts, C. and Lane, D. M. (2009) ‘Fault diagnosis of a train door system based on semantic knowledge representation’, in *Proceedings of the 4th IET International Conference on Railway Condition Monitoring*, pp. 27–27. doi: 10.1049/ic:20080333.

Mishra, M. and Thaduri, A. (2016) ‘Ontology Based Diagnosis for Maintenance Decisions of Paper Mill Roller Using Dynamic Response’, in *Current Trends in Reliability, Availability, Maintainability and Safety*, pp. 187–196. doi: 10.1007/978-3-319-23597-4.

Moher, D., Liberati, A., Tetzlaff, J. and Altman, D. G. (2010) ‘Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement’, *British*

Medical Journal, 8, pp. 336–341. doi: 10.1016/j.ijsu.2010.02.007.

Mourtzis, D., Vlachou, A. and Zogopoulos, V. (2017a) 'Cloud-based augmented reality remote maintenance through shop-floor monitoring: A product-service system approach', *Journal of Manufacturing Science and Engineering, Transactions of the ASME*, 139(6), pp. 1–11. doi: 10.1115/1.4035721.

Mourtzis, D., Vlachou, A. and Zogopoulos, V. (2017b) 'Cloud-Based Augmented Reality Remote Maintenance Through Shop-Floor Monitoring: A Product-Service System Approach', *Journal of Manufacturing Science and Engineering*, 139(6), p. 061011. doi: 10.1115/1.4035721.

Musen, M. A. (2015) 'The Protégé Project: A Look Back and a Look Forward', *AI Matters*, 1(4), pp. 4–12. doi: 10.1145/2757001.2757003.

Naets, F., Cosco, F. and Desmet, W. (2017) 'Improved human-computer interaction for mechanical systems design through augmented strain / stress Frank Naets Francesco Cosco Wim Desmet', *International Journal of Intelligent Engineering Informatics*, 5(1), pp. 50–66. doi: 10.1504/IJIEI.2017.082565.

Nagy, J., Oláh, J., Erdei, E., Máté, D. and Popp, J. (2018) 'The role and impact of industry 4.0 and the internet of things on the business strategy of the value chain-the case of hungary', *Sustainability (Switzerland)*, 10(10). doi: 10.3390/su10103491.

Nakai, A. and Suzuki, K. (2016) 'Instructional Information System using AR Technology for Chemical Plants', *Chemical Engineering Transactions*, 53, pp. 199–204. doi: 10.3303/CET1653034.

Nawaz, Z., Aftab, S. and Anwer, F. (2017) 'Simplified FDD Process Model', *International Journal of Modern Education and Computer Science*, 9(9), pp. 53–59. doi: 10.5815/ijmecs.2017.09.06.

Nee, A. Y. C., Ong, S. K., Chryssolouris, G. and Mourtzis, D. (2012) 'Augmented reality applications in design and manufacturing', *CIRP Annals - Manufacturing Technology*. CIRP, 61(2), pp. 657–679. doi: 10.1016/j.cirp.2012.05.010.

Ng, L. X., Wang, Z. B., Ong, S. K. and Nee, A. Y. C. (2013) 'Integrated product design and assembly Assembly planning and evaluation in an augmented reality environment', *International Journal of Production Research*, 51(23–24), pp. 7388–7404. doi: 10.1080/00207543.2013.837986.

Nielsen, J. (1993) *Usability Engineering*. 1st edn, *Morgan-Kaufmann*. 1st edn. San Francisco, California, USA: Morgan Kaufmann. Available at: https://www.nngroup.com/books/usability-engineering/.

Nonaka, I. (1994) 'A dynamic theory of organizational knowledge creation', *Organization science*. Informs, 5(1), pp. 14–37. doi: 10.1287/orsc.5.1.14.

Nuñez, D. L. and Borsato, M. (2017) 'An ontology-based model for prognostics and health management of machines', *Journal of Industrial Information Integration*. Elsevier Inc., 6, pp. 33–46. doi: 10.1016/j.jii.2017.02.006.

Okazaki, T. and Takaseki, R. (2017) 'Override ship maneuvering simulator using AR toolkit', *Intelligent Automation & Soft Computing*. Taylor & Francis, 23(1), pp. 167–174. doi: 10.1080/10798587.2016.1142709.

Ong, S. K. and Nee, A. Y. C. (2004) *Virtual and Augmented Reality Applications in Manufacturing*. 1st edn, *Springer-Verlag*. 1st edn. London: Springer-Verlag. doi: 10.1007/978-1-4471-3873-0.

Ong, S. K., Yuan, M. L. and Nee, A. Y. C. (2008) 'Augmented reality applications in manufacturing: a survey', *International Journal of Production Research*, 46(10), pp. 2707–2742. doi: 10.1080/00207540601064773.

Ong, S. K. and Zhu, J. (2013) 'A novel maintenance system for equipment serviceability improvement', *CIRP Annals - Manufacturing Technology*, 62(1), pp. 39–42. doi: 10.1016/j.cirp.2013.03.091.

Oyekan, J., Prabhu, V., Tiwari, A., Baskaran, V., Burgess, M. and Mcnally, R. (2017) 'Remote real-time collaboration through synchronous exchange of digitised human–workpiece interactions', *Future Generation Computer Systems*. Elsevier B.V., 67, pp. 83–93. doi: 10.1016/j.future.2016.08.012.

Oztemel, E. and Gursev, S. (2020) 'Literature review of Industry 4.0 and related technologies', *Journal of Intelligent Manufacturing*. Springer US, 31(1), pp. 127–182. doi: 10.1007/s10845-018-1433-8.

Palmarini, R., Erkoyuncu, J. A., Roy, R. and Torabmostaedi, H. (2018) 'A systematic review of augmented reality applications in maintenance', *Robotics and Computer-Integrated Manufacturing*. Elsevier Ltd, 49(July 2017), pp. 215–228. doi: 10.1016/j.rcim.2017.06.002.

Panzarino, O. (2014) *Learning Cypher*. 1st edn. Packt Publishing.

Parida, A. and Kumar, U. (2004) 'Managing information is key to maintenance effectiveness', *Intelligent Maintenance Systems*. Godkänd. Available at: http://www.diva-portal.org/smash/record.jsf?pid=diva2%3A1001234&dswid=-8340.

Park, C.-S., Lee, D.-Y., Kwon, O.-S. and Wang, X. (2013) 'A framework for proactive construction defect management using BIM, augmented reality and ontology-based data collection template', *Automation in Construction*, 33, pp. 61–71. doi: 10.1016/j.autcon.2012.09.010.

Pedersen, M. R., Nalpantidis, L., Andersen, R. S., Schou, C., Bøgh, S., Krüger, V. and Madsen, O. (2016) 'Robot skills for manufacturing: From concept to industrial deployment', *Robotics and Computer-Integrated Manufacturing*. Elsevier, 37, pp. 282–291. doi: 10.1016/j.rcim.2015.04.002.

Peffers, K., Tuunanen, T., Rothenberger, M. A. and Chatterjee, S. (2008) 'A Design Science Research Methodology for Information Systems Research', *Journal of Management Information Systems*, 24(3), pp. 45–77. doi: 10.2753/mis0742-1222240302.

Pérez-Salazar, M. del R., Aguilar-Lasserre, A. A., Cedillo-Campos, M. G., Juárez-Martínez, U. and Posada-Gómez, R. (2019) 'Processes and measurement of knowledge management in supply chains: an integrative systematic literature review', *International Journal of Production Research*, 57(7), pp. 2136–2159. doi: 10.1080/00207543.2018.1521530.

Pierdicca, R., Frontoni, E., Zingaretti, P., Mancini, A., Malinverni, E. S., Tassetti, A. N., Marcheggiani, E. and Galli, A. (2016) 'Smart maintenance of riverbanks using a standard data layer and Augmented Reality', *Computers and Geosciences*. Elsevier, 95, pp. 67–74. doi: 10.1016/j.cageo.2016.06.018.

Pistofidis, P., Emmanouilidis, C., Papadopoulos, A. and Botsaris, P. N. (2016) 'Management of linked knowledge in industrial maintenance', *Industrial Management & Data Systems*, 116(8), pp. 1741–1758. doi: 10.1108/IMDS-10-2015-0409.

Plauger, P. J. (2002) 'The C/C++ programming language', *C/C++ Users Journal*.

Pokric, B., Krco, S., Drajic, D., Pokric, M., Rajs, V., Mihajlovic, Z., Knezevic, P. and Jovanovic, D. (2015) 'Augmented reality enabled iot services for environmental monitoring utilising serious gaming concept', *Journal of Wireless Mobile Networks, Ubiquitous Computing, and Dependable Applications*, 6(1), pp. 37–55. Available at: http://www.scopus.com/inward/record.url?eid=2-s2.0-84926287740&partnerID=tZOtx3y1.

Priya, M. H. and Vasudevan, S. K. (2018) 'An innovative application for car engine disparity check - A novel attempt', *Proceedings of IEEE International Conference on Signal Processing and Communication, ICSPC 2017*, 2018-Janua(July), pp. 175–180. doi: 10.1109/CSPC.2017.8305833.

PTC Corporation (2020) *Vuforia SDK, PTC Corporation.* Available at: https://library.vuforia.com/getting-started/overview.html (Accessed: 29 April 2020).

Raad, J. and Cruz, C. (2015) 'A survey on ontology evaluation methods', in *IC3K 2015 - Proceedings of the 7th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management.* Lisbon, Portugal: ScitePress, pp. 179–186. doi: 10.5220/0005591001790186.

Radkowski, R., Herrema, J. and Oliver, J. (2015) 'Augmented Reality-Based Manual Assembly Support With Visual Features for Different Degrees of Difficulty', *International Journal of Human-Computer Interaction*, 31(5), pp. 337–349. doi: 10.1080/10447318.2014.994194.

Rajpathak, D. and Chougule, R. (2011) 'A generic ontology development framework for data integration and decision support in a distributed environment', *International Journal of Computer Integrated Manufacturing*, 24(2), pp. 154–170. doi: 10.1080/0951192X.2010.531291.

Rajpathak, D. G. and Singh, S. (2014) 'An ontology-based text mining method to develop D-matrix from unstructured text', *IEEE Transactions on Systems, Man, and Cybernetics: Systems.* IEEE, 44(7), pp. 966–977. doi: 10.1109/TSMC.2013.2281963.

Rambach, J., Pagani, A., Schneider, M., Artemenko, O. and Stricker, D. (2018) '6DoF Object Tracking based on 3D Scans for Augmented Reality Remote Live Support', *Computers.* MDPI AG, 7(1), p. 6. doi: 10.3390/computers7010006.

Ramirez-Amaro, K., Beetz, M. and Cheng, G. (2017) 'Transferring skills to humanoid robots by extracting semantic representations from observations of human activities', *Artificial Intelligence*, 247, pp. 95–118. doi: 10.1016/j.artint.2015.08.009.

Rauschnabel, P. A., He, J. and Ro, Y. K. (2018) 'Antecedents to the adoption of augmented reality smart glasses: A closer look at privacy risks', *Journal of Business Research.* Elsevier, 92(April 2016), pp. 374–384. doi: 10.1016/j.jbusres.2018.08.008.

Re, G. M., Oliver, J. and Bordegoni, M. (2016) 'Impact of monitor-based augmented reality for on-site industrial manual operations', *Cognition, Technology and Work.* Springer London, 18(2), pp. 379–392. doi: 10.1007/s10111-016-0365-3.

Reina, A., Cope, A. J., Nikolaidis, E., Marshall, J. A. R. and Sabo, C. (2017)

‘ARK : Augmented Reality for Kilobots’, *IEEE Robotics and Automation Letters*, 2(3), pp. 1755–1761. doi: 10.1109/LRA.2017.2700059.

Renu, R., Visotsky, D., Knackstedt, S., Mocko, G., Summers, J. D. and Schulte, J. (2016) ‘A Knowledge Based FMEA to Support Identification and Management of Vehicle Flexible Component Issues’, *Procedia CIRP*. Elsevier B.V., 44, pp. 157–162. doi: 10.1016/j.procir.2016.02.112.

Rødseth, H., Schjølberg, P. and Marhaug, A. (2017) ‘Deep digital maintenance’, *Advances in Manufacturing*. Shanghai University, 5(4), pp. 299–310. doi: 10.1007/s40436-017-0202-9.

Rohacs, J., Rohacs, D. and Jankovics, I. (2016) ‘Conceptual development of an advanced air traffic controller workstation based on objective workload monitoring and augmented reality’, *Proceedings of the Institution of Mechanical Engineers, Part G: Journal of Aerospace Engineering*, 230(9), pp. 1747–1761. doi: 10.1177/0954410016636154.

Roy, R., Stark, R., Tracht, K., Takata, S. and Mori, M. (2016) ‘Continuous maintenance and the future - Foundations and technological challenges’, *CIRP Annals - Manufacturing Technology*. CIRP, 65(2), pp. 667–688. doi: 10.1016/j.cirp.2016.06.006.

Schaeffer, S. E., Garza, S. E., Espinosa, J. C., Urbina, S. C., Nurmi, P. and Cruz-Reyes, L. (2018) ‘A framework for informing consumers on the ecological impact of products at point of sale’, *Behaviour and Information Technology*. Taylor & Francis, 37(6), pp. 607–621. doi: 10.1080/0144929X.2018.1470254.

Schall, G., Zollmann, S. and Reitmayr, G. (2013) ‘Smart Vidente: Advances in mobile augmented reality for interactive visualization of underground infrastructure’, *Personal and Ubiquitous Computing*, 17(7), pp. 1533–1549. doi: 10.1007/s00779-012-0599-x.

Sebillo, M., Vitiello, G., Paolino, L. and Ginige, A. (2016) ‘Training emergency responders through augmented reality mobile interfaces’, *Multimedia Tools and Applications*. Multimedia Tools and Applications, 75(16), pp. 9609–9622. doi: 10.1007/s11042-015-2955-0.

Shaaban, O. A., Che Mat, R. and Mahayudin, M. H. (2015) ‘The development of mobile augmented reality for laptop maintenance (MAR4LM)’, *Jurnal Teknologi*, 77(29), pp. 91–96. doi: 10.11113/jt.v77.6842.

Shatte, A., Holdsworth, J. and Lee, I. (2014) ‘Mobile augmented reality based context-aware library management system’, *Expert Systems with Applications*.

Elsevier Ltd, 41(5), pp. 2174–2185. doi: 10.1016/j.eswa.2013.09.016.

Smith, M. K., Welty, C. and McGuinness, D. L. (2004) *OWL Web Ontology Language Guide, W3C Recommendation*. Available at: https://www.w3.org/TR/2004/REC-owl-guide-20040210/ (Accessed: 23 July 2019).

Smyth, B. (2007) 'Case-Based Recommendation', in Brusilovsky, P., Kobsa, A., and Nejdl, W. (eds) *The Adaptive Web: Methods and Strategies of Web Personalization*. 1st edn. Berlin, Heidelberg: Springer, pp. 342–376. doi: 10.1007/978-3-540-72079-9_11.

Suárez-Figueroa, M. C., Gómez-Pérez, A. and Fernández-López, M. (2015) 'The NeOn Methodology framework: Ascenario-based methodology for ontologydevelopment', *Applied Ontology*, 10(2), pp. 107–145. doi: 10.3233/AO-150145.

Suarez-Warden, F. and González Mendívil, E. (2017) 'Problem solving in transference of procedural knowledge aided by AR to support aeronautical transportation decisions', *International Journal on Interactive Design and Manufacturing*. Springer Paris, pp. 1–8. doi: 10.1007/s12008-017-0384-1.

Suárez-Warden, F., Mendívil, E. G., Ramírez, A. F. and García-Lumbreras, S. (2015) 'Profit Model for Incorporating AR Technology in Assembly Tasks of Aeronautical Maintenance', *Procedia Computer Science*. Elsevier Masson SAS, 75(Vare), pp. 113–122. doi: 10.1016/j.procs.2015.12.227.

Surhone, L. M., Tennoe, M. T. and Henssonow, S. F. (2010) *Node.Js*. Beau Bassin, MUS: Betascript Publishing.

Tartir, S. and Arpinar, I. B. (2007) 'Ontology evaluation and ranking using OntoQA', in *International Conference on Semantic Computing*, pp. 185–192. doi: 10.1109/ICSC.2007.19.

Tartir, S., Arpinar, I., Moore, M., Sheth, A. and Aleman-Meza, B. (2005) 'OntoQA: Metric-Based Ontology Quality Analysis', in *IEEE Workshop on Knowledge Acquisition from Distributed, Autonomous, Semantically Heterogeneous Data and Knowledge Sources*, pp. 45–53. doi: 10.1.1.131.2087.

Tatic, D. and Tešic, B. (2017) 'The application of augmented reality technologies for the improvement of occupational safety in an industrial environment', *Computers in Industry*, 85, pp. 1–10. doi: 10.1016/j.compind.2016.11.004.

Torres-Ruiz, M., Mata, F., Zagal, R., Guzmán, G., Quintero, R. and Moreno-Ibarra, M. (2020) 'A recommender system to generate museum itineraries applying augmented reality and social-sensor mining techniques', *Virtual Reality*. Springer London, 24(1), pp. 175–189. doi: 10.1007/s10055-018-0366-z.

Tsai, M. K. and Yau, N. J. (2013) 'Enhancing usability of augmented-reality-based mobile escape guidelines for radioactive accidents', *Journal of Environmental Radioactivity*. Elsevier Ltd, 118, pp. 15–20. doi: 10.1016/j.jenvrad.2012.11.001.

Tzimas, E., Vosniakos, G. C. and Matsas, E. (2019) 'Machine tool setup instructions in the smart factory using augmented reality: a system construction perspective', *International Journal on Interactive Design and Manufacturing*. Springer Paris, 13(1), pp. 121–136. doi: 10.1007/s12008-018-0470-z.

Unity Technologies (2019) *Unity Game Engine*, *https://unity.com*. Available at: https://unity.com.

Veas, E., Grasset, R., Ferencik, I., Grünewald, T. and Schmalstieg, D. (2013) 'Mobile augmented reality for environmental monitoring', *Personal and Ubiquitous Computing*, 17(7), pp. 1515–1531. doi: 10.1007/s00779-012-0597-z.

Vincent, H., Benoit, J., Xavier, S. and David, B. (2017) 'Inoovas - Industrial ontology for operation in virtual and augmented scene: The architecture', *Proceedings of the 4th International Conference on Control, Decision and Information Technologies, CoDIT 2017*. IEEE, 2017-Janua, pp. 300–305. doi: 10.1109/CoDIT.2017.8102608.

Vogl, G. W., Weiss, B. A. and Helu, M. (2019) 'A review of diagnostic and prognostic capabilities and best practices for manufacturing', *Journal of Intelligent Manufacturing*. Springer US, 30(1), pp. 79–95. doi: 10.1007/s10845-016-1228-8.

Vrandečić, D. (2009) 'Ontology Evaluation', in Staab, S. and Studer, R. (eds) *Handbook on Ontologies, International Handbooks on Information Systems*. Springer-Verlag Berlin Heidelberg, pp. 293–313. doi: 10.1007/978-3-540-92673-3_13.

Walczak, K. and Flotyński, J. (2019) 'Inference-based creation of synthetic 3D content with ontologies', *Multimedia Tools and Applications*, 78(9), pp. 12607–12638. doi: 10.1007/s11042-018-6788-5.

Wali, F., Huang, J., Luo, W., Li, G. and Planta, A. K. de (2020) *A Framework for A Dynamic AR Based Degradation Assessment Digital Twin for Enhancing Maintenance-related Activities*. Cranfield University.

Wan, S., Li, D., Gao, J. and Li, J. (2019) 'A knowledge based machine tool maintenance planning system using case-based reasoning techniques', *Robotics and Computer-Integrated Manufacturing*, 58(January), pp. 80–96. doi: 10.1016/j.rcim.2019.01.012.

Wang, D., Tang, W. H. and Wu, Q. H. (2010) 'Ontology-based fault diagnosis for power transformers', in *IEEE PES General Meeting, PES 2010*, pp. 1–8. doi: 10.1109/PES.2010.5589575.

Wang, M., Qin, G. and Hu, N. (2012) 'Research on knowledge representation for health management of aerospace based on ontology', in *International Conference on Quality, Reliability, Risk, Maintenance, and Safety Engineering*. IEEE, pp. 587–591. doi: 10.1109/ICQR2MSE.2012.6246302.

Wang, S., Guo, R., Wang, H., Ma, Y. and Zong, Z. (2018) 'Manufacture assembly fault detection method based on deep learning and mixed reality', *Proceedings of the IEEE International Conference on Information and Automation, ICIA 2018*. IEEE, (20172), pp. 808–813. doi: 10.1109/ICInfA.2018.8812577.

Wang, X., Love, P. E. D., Kim, M. J. and Wang, W. (2014) 'Mutual awareness in collaborative design: An Augmented Reality integrated telepresence system', *Computers in Industry*. Elsevier B.V., 65(2), pp. 314–324. doi: 10.1016/j.compind.2013.11.012.

Wang, X., Ong, S. K. and Nee, A. Y. C. (2015) 'Real-virtual interaction in AR assembly simulation based on component contact handling strategy', *Assembly Automation*, 35(4), pp. 376–394. doi: 10.1108/AA-02-2015-012.

Wang, X., Ong, S. K. and Nee, A. Y. C. (2016) 'A comprehensive survey of augmented reality assembly research', *Advances in Manufacturing*. Shanghai University, 4(1), pp. 1–22. doi: 10.1007/s40436-015-0131-4.

Wang, X, Ong, S. K. and Nee, A. Y. C. (2016a) 'Multi-modal augmented-reality assembly guidance based on bare-hand interface', *Advanced Engineering Informatics*, 30(3), pp. 406–421. doi: 10.1016/j.aei.2016.05.004.

Wang, X, Ong, S. K. and Nee, A. Y. C. (2016b) 'Real-virtual components interaction for assembly simulation and planning', *Robotics and Computer-Integrated Manufacturing*, 41, pp. 102–114. doi: 10.1016/j.rcim.2016.03.005.

Wang, X., Yew, A. W. W., Ong, S. K. and Nee, A. Y. C. (2020) 'Enhancing smart shop floor management with ubiquitous augmented reality', *International Journal of Production Research*. Taylor & Francis, 58(8), pp. 2352–2367. doi: 10.1080/00207543.2019.1629667.

Wang, Z. B., Ng, L. X., Ong, S. K. and Nee, A. Y. C. (2013) 'Assembly planning and evaluation in an augmented reality environment', *International Journal of Production Research*, 51(23–24), pp. 7388–7404. doi: 10.1080/00207543.2013.837986.

Wang, Z. B., Ong, S. K. and Nee, A. Y. C. (2013) 'Augmented reality aided interactive manual assembly design', *International Journal of Advanced Manufacturing Technology*, 69(5–8), pp. 1311–1321. doi: 10.1007/s00170-013-5091-x.

Webel, S., Bockholt, U., Engelke, T., Gavish, N., Olbrich, M. and Preusche, C. (2013) 'An augmented reality training platform for assembly and maintenance skills', *Robotics and Autonomous Systems*, 61(4), pp. 398–403. doi: 10.1016/j.robot.2012.09.013.

Weijters, B., Cabooter, E. and Schillewaert, N. (2010) 'The effect of rating scale format on response styles: The number of response categories and response category labels', *International Journal of Research in Marketing*. Elsevier B.V., 27(3), pp. 236–247. doi: 10.1016/j.ijresmar.2010.02.004.

Westerfield, G., Mitrovic, A. and Billinghurst, M. (2013) 'Intelligent Augmented Reality Training for Assembly Tasks', in *International Conference on Artificial Intelligence in Education*. Berlin: Springer-Verlag Berlin Heidelberg. doi: 10.1007/978-3-642-39112-5-55.

Westerfield, G., Mitrovic, A. and Billinghurst, M. (2015) 'Intelligent Augmented Reality Training for Motherboard Assembly', *International Journal of Artificial Intelligence in Education*, 25, pp. 157–172. doi: 10.1007/s40593-014-0032-x.

Wójcicki, T. (2014) 'Supporting the diagnostics and the maintenance of technical devices with augmented reality', *Diagnostyka*, 15(1), pp. 43–47.

Xiao, Y. and Watson, M. (2017) 'Guidance on Conducting a Systematic Literature Review', *Journal of Planning Education and Research*, p. 20. doi: 10.1177/0739456X17723971.

Xu, F., Liu, X., Chen, W., Zhou, C. and Cao, B. (2018) 'Ontology-based method for fault diagnosis of loaders', *Sensors*, 18(729). doi: 10.3390/s18030729.

Yang, S.-R., Chang, J.-R., Lin, H.-M. and Lin, K.-H. (2015) 'Identification of Manholes Beneath Pavements Using RFID and AR Technologies', *Journal of Testing and Evaluation*, 43(2), p. 20140136. doi: 10.1520/JTE20140136.

Yazdi, M. (2019) 'Acquiring and Sharing Tacit Knowledge in Failure Diagnosis Analysis Using Intuitionistic and Pythagorean Assessments', *Journal of Failure Analysis and Prevention*. Springer US, 19(2), pp. 369–386. doi: 10.1007/s11668-

019-00599-w.

Yew, A. W. W., Ong, S. K. and Nee, A. Y. C. (2016) ‘Towards a griddable distributed manufacturing system with augmented reality interfaces’, *Robotics and Computer-Integrated Manufacturing.* Elsevier, 39, pp. 43–55. doi: 10.1016/j.rcim.2015.12.002.

Zenati-Henda, N., Bellarbi, A., Benbelkacem, S. and Belhocine, M. (2014) ‘Augmented reality system based on hand gestures for remote maintenance’, in *International Conference on Multimedia Computing and Systems (ICMCS).* Marrakech, Morroco. doi: 10.1109/ICMCS.2014.6911258.

Zhong, Z., Xu, T., Wang, F. and Tang, T. (2018) ‘Text Case-Based Reasoning Framework for Fault Diagnosis and Predication by Cloud Computing’, *Mathematical Problems in Engineering*, 2018, pp. 1–10. doi: 10.1155/2018/9464971.

Zhou, A., Yu, D. and Zhang, W. (2015) ‘A research on intelligent fault diagnosis of wind turbines based on ontology and FMECA’, *Advanced Engineering Informatics.* Elsevier Ltd, 29(1), pp. 115–125. doi: 10.1016/j.aei.2014.10.001.

Zhou, Y., Li, Q. and Zuo, Y. (2009) ‘Fault knowledge management in aircraft maintenance’, in *Proceedings of the 8th International Conference on Reliability, Maintainability and Safety.* IEEE, pp. 645–649. doi: 10.1109/ICRMS.2009.5270111.

Zhou, Y., Luo, H. and Yang, Y. (2017) ‘Implementation of augmented reality for segment displacement inspection during tunneling construction’, *Automation in Construction*, 82, pp. 112–121. doi: 10.1016/j.autcon.2017.02.007.

Zhu, J., Ong, S. K. and Nee, A. Y. C. (2013) ‘An authorable context-aware augmented reality system to assist the maintenance technicians’, *The International Journal of Advanced Manufacturing Technology*, 66(9), pp. 1699–1714. doi: 10.1007/s00170-012-4451-2.

Zhu, J., Ong, S. K. and Nee, A. Y. C. (2014) ‘A context-aware augmented reality system to assist the maintenance operators’, *International Journal on Interactive Design and Manufacturing*, 8(4), pp. 293–304. doi: 10.1007/s12008-013-0199-7.

Zhu, J., Ong, S. K. and Nee, A. Y. C. (2015) ‘A context-aware augmented reality assisted maintenance system’, *International Journal of Computer Integrated Manufacturing.* Taylor & Francis, 28(2), pp. 213–225. doi: 10.1080/0951192X.2013.874589.

Zhu, Z., Zhou, X. and Shao, K. (2019) ‘A novel approach based on Neo4j for multi-constrained flexible job shop scheduling problem', *Computers and Industrial Engineering.* doi: 10.1016/j.cie.2019.03.022.

Zollmann, S., Hoppe, C. and Kluckner, S. (2014) 'Augmented Reality for Construction Site Monitoring and Documentation', *Proceedings of the IEEE*, 102(2), pp. 137–154. doi: 10.1109/JPROC.2013.2294314.

Zubizarreta, J., Aguinaga, I. and Amundarain, A. (2019) 'A framework for augmented reality guidance in industry', *International Journal of Advanced Manufacturing Technology.* The International Journal of Advanced Manufacturing Technology, 102(9–12), pp. 4095–4108. doi: 10.1007/s00170-019-03527-2.

Appendices

Appendix A
Ontology expert interviews: procedure and questionnaire

CURES Reference: CURES/3680/2018

Data repository: http://doi.org/10.17862/cranfield.rd.12279152

A relevant method to assess the validity of an ontology is to interview its targeted users regarding its intended purpose (Aruna, Saranya and Bhandari, 2011; Bautista-Zambrana, 2015). As part of this PhD thesis, the student proposed a semi-structured interview for experts in diagnosis activities to validate the ability of diagont (Subsection 3.4.1) to describe their diagnosis rationale. This semi-structured interview, which was designed to be conducted individually, comprised the following steps:

1. Ontology description (10 minutes): to present interviewees diagont, its purpose and its schema. This description was based on a document containing the definitions of diagont's classes, attributes and relationships and a slide-presentation describing its schema.
2. Ontology improvement suggestions (40 minutes): to ask interviewees about their opinions regarding diagont's classes, attributes and relationships and suggest them to propose any modifications they may consider improving the description of their diagnosis rationale.

Interviewees answers were collected as free-text responses while guided by the student (interviewer) using a questionnaire. The questionnaire helped to guide the semi-structured interview for ensuring all diagont's classes, attributes and relationships were considered by interviewees regarding their ability to describe their diagnosis rationale. The questionnaire consisted of two parts. The first part (Table A-1) aimed to collect demographical data regarding the purpose of the study such as years of experience, roles, etc.

Table A-1. Demographic questionnaire from ontology expert interviews[5]

Participant ID	*Response (numeric)*
Role	*Response (free-text)*
Years of experience	*Response (numeric)*
Organisation	*Response {A, B}*
Department	*Response (free-text)*
Case study	*Response {Loading Arm, Helicopter Mission System}*

The second part aimed to guide interviewees through diagont's scheme for proposing changes. Table A-2 presents the data collection form for this interview's part. It includes all diagont's classes, attributes and relationships and the types of changes to be made either to the element's name, schema (e.g. domain, range, etc.) or definition.

Table A-2. Ontological questionnaire from ontology expert interviews[5]

Type	Name	Change	Description
Class	Step	*{Name, Schema, Definition}*	*Change (free-text)*
Relationship	belongsTo	*{Name, Schema, Definition}*	*Change (free-text)*
Relationship	evaluates	*{Name, Schema, Definition}*	*Change (free-text)*
Relationship	diagnoses	*{Name, Schema, Definition}*	*Change (free-text)*
Attribute	isCritical	*{Name, Schema, Definition}*	*Change (free-text)*
Attribute	isContributory	*{Name, Schema, Definition}*	*Change (free-text)*

[5] Elements of the table that represent interviewees' responses are in *Italic*.

Type	Name	Change	Description
Attribute	hasObject	*{Name, Schema, Definition}*	*Change (free-text)*
Attribute	hasMethod	*{Name, Schema, Definition}*	*Change (free-text)*
Attribute	hasComparison	*{Name, Schema, Definition}*	*Change (free-text)*
Class	Task	*{Name, Schema, Definition}*	*Change (free-text)*
Relationship	conductedBy	*{Name, Schema, Definition}*	*Change (free-text)*
Attribute	hasDescription	*{Name, Schema, Definition}*	*Change (free-text)*
Class	State	*{Name, Schema, Definition}*	*Change (free-text)*
Relationship	causes	*{Name, Schema, Definition}*	*Change (free-text)*
Relationship	describes	*{Name, Schema, Definition}*	*Change (free-text)*
Relationship	measuredBy	*{Name, Schema, Definition}*	*Change (free-text)*
Attribute	hasStatus	*{Name, Schema, Definition}*	*Change (free-text)*
Attribute	hasDomain	*{Name, Schema, Definition}*	*Change (free-text)*
Attribute	hasPhenomenon	*{Name, Schema, Definition}*	*Change (free-text)*
Attribute	hasMeasureValue	*{Name, Schema, Definition}*	*Change (free-text)*
Attribute	hasMeasureUnit	*{Name, Schema, Definition}*	*Change (free-text)*
Attribute	hasMeasureDate	*{Name, Schema, Definition}*	*Change (free-text)*
Class	Failure	*{Name, Schema, Definition}*	*Change (free-text)*
Relationship	affectsTo	*{Name, Schema, Definition}*	*Change (free-text)*
Attribute	hasDescription	*{Name, Schema, Definition}*	*Change (free-text)*
Attribute	hasImpact	*{Name, Schema, Definition}*	*Change (free-text)*
Attribute	hasDomain	*{Name, Schema, Definition}*	*Change (free-text)*
Attribute	hasPhenomenon	*{Name, Schema, Definition}*	*Change (free-text)*
Attribute	hasImage	*{Name, Schema, Definition}*	*Change (free-text)*
Attribute	hasAudio	*{Name, Schema, Definition}*	*Change (free-text)*
Class	Auditor	*{Name, Schema, Definition}*	*Change (free-text)*
Relationship	evaluates	*{Name, Schema, Definition}*	*Change (free-text)*
Relationship	monitors	*{Name, Schema, Definition}*	*Change (free-text)*
Attribute	isValidated	*{Name, Schema, Definition}*	*Change (free-text)*
Attribute	hasComparison	*{Name, Schema, Definition}*	*Change (free-text)*
Class	Monitor	*{Name, Schema, Definition}*	*Change (free-text)*
Relationship	encounters	*{Name, Schema, Definition}*	*Change (free-text)*

Type	Name	Change	Description
Relationship	Considers	*{Name, Schema, Definition}*	*Change (free-text)*
Attribute	hasDescription	*{Name, Schema, Definition}*	*Change (free-text)*
Class	Agent	*{Name, Schema, Definition}*	*Change (free-text)*
Class	Asset	*{Name, Schema, Definition}*	*Change (free-text)*
Class	System	*{Name, Schema, Definition}*	*Change (free-text)*
Class	Component	*{Name, Schema, Definition}*	*Change (free-text)*
Class	Device	*{Name, Schema, Definition}*	*Change (free-text)*
Datatype	impact	*{Set, Schema, Definition}*	*Change (free-text)*
Datatype	status	*{Set, Schema, Definition}*	*Change (free-text)*
Datatype	domain	*{Set, Schema, Definition}*	*Change (free-text)*
Datatype	phenomenon	*{Set, Schema, Definition}*	*Change (free-text)*
Datatype	object	*{Set, Schema, Definition}*	*Change (free-text)*
Datatype	method	*{Set, Schema, Definition}*	*Change (free-text)*
Datatype	comparison	*{Set, Schema, Definition}*	*Change (free-text)*
Datatype	unit	*{Set, Schema, Definition}*	*Change (free-text)*

The complete interview sheets including the information sheet, the informed consent form and the interview data collection form and documents can be found at http://doi.org/10.17862/cranfield.rd.12279152.

Appendix B
Reporting usability surveys: procedure and questionnaire

CURES Reference: CURES/3680/2018

Data repository: http://doi.org/10.17862/cranfield.rd.12279152

A relevant method to assess the usability of a software tool is to allow targeted users to try it and then ask them to state their opinion regarding such matter (Nielsen, 1993). The student proposed to conduct usability surveys on the reporting tool after diagnosis experts tried it for reporting two different failures. Their aim was to evaluate the tool from an ontological perspective using relevant criteria as described in Subsection 3.5.2.4. Table A-1 presents the survey that diagnosis experts were asked to complete after conducting usability tests. It comprises the different statements that interviewees were asked to state their agreement with. The survey utilises a Likert Scale format of 1-7 to take into consideration the results presented by Weijters, Cabooter and Schillewaert (2010). These suggested that such format maximises potential information transmission when surveying expert populations.

Table B-1. Survey questionnaire on reporting tool's usability[6]

Statements	Likert Scale						
The vocabulary used in the ontology is accurate	*1*	*2*	*3*	*4*	*5*	*6*	*7*
The vocabulary used in the ontology is complete	*1*	*2*	*3*	*4*	*5*	*6*	*7*
The vocabulary used in the ontology is concise	*1*	*2*	*3*	*4*	*5*	*6*	*7*
The vocabulary used in the ontology is consistent	*1*	*2*	*3*	*4*	*5*	*6*	*7*
The hierarchy of the ontology is accurate	*1*	*2*	*3*	*4*	*5*	*6*	*7*
The hierarchy of the ontology is complete	*1*	*2*	*3*	*4*	*5*	*6*	*7*
The hierarchy of the ontology is concise	*1*	*2*	*3*	*4*	*5*	*6*	*7*
The hierarchy of in the ontology is consistent	*1*	*2*	*3*	*4*	*5*	*6*	*7*
The semantics of the ontology are accurate	*1*	*2*	*3*	*4*	*5*	*6*	*7*
The semantics of the ontology are complete	*1*	*2*	*3*	*4*	*5*	*6*	*7*
The semantics of the ontology are concise	*1*	*2*	*3*	*4*	*5*	*6*	*7*
The semantics of the ontology are consistent	*1*	*2*	*3*	*4*	*5*	*6*	*7*
The context of the ontology is accurate	*1*	*2*	*3*	*4*	*5*	*6*	*7*
The context of the ontology is complete	*1*	*2*	*3*	*4*	*5*	*6*	*7*
The context of the ontology is concise	*1*	*2*	*3*	*4*	*5*	*6*	*7*
The context of in the ontology is consistent	*1*	*2*	*3*	*4*	*5*	*6*	*7*

The complete interview sheets including the information sheet, the informed consent form and the interview data collection form and documents can be found at http://doi.org/10.17862/cranfield.rd.12279152.

[6] Elements of the table that represent survey's responses are in *Italic*.

Appendix C
Monitoring usability surveys: procedure and questionnaire

CURES Reference: CURES/9203/2019

Data repository: http://doi.org/10.17862/cranfield.rd.12279152

A relevant method to assess the usability of a software tool is to allow targeted users to try it and then ask them to state their opinion regarding such matter (Nielsen, 1993). The student proposed to conduct usability surveys regarding the monitoring tool after testers tried it in fault-finding tasks as part of monitoring efficiency experiments (Subsection 3.5.2.5). Surveys aim to evaluate the tool from a usability perspective regarding their ease-of-use and effectiveness (Subsection 3.5.2.6) according to the criterions proposed by (Nielsen, 1993). Table A-1 presents the survey that testers were asked to complete after conducting monitoring efficiency experiments. It comprises the different statements about ease-of-use and effectiveness that testers were asked to state their agreement with. The survey utilises a Likert Scale format of 1-5 to take into consideration the results presented by Weijters, Cabooter and Schillewaert (2010). These suggested that such format maximises potential information transmission when surveying non-expert populations.

Table C-1. Survey questionnaire on monitoring tool's usefulness[7]

Criterion	Statement	Likert Scale				
Ease-of-use	I found the monitoring tool easy to use	*1*	*2*	*3*	*4*	*5*
	I liked using the monitoring tool to conduct fault-finding tasks	*1*	*2*	*3*	*4*	*5*
	I found easy to understand the data shown by the monitoring tool	*1*	*2*	*3*	*4*	*5*
Effectiveness	The monitoring tool helped to conduct fault-finding tasks	*1*	*2*	*3*	*4*	*5*
	The monitoring tool identified failures accurately	*1*	*2*	*3*	*4*	*5*
	The monitoring tool represented clearly the condition of the asset	*1*	*2*	*3*	*4*	*5*

The complete interview sheets including the information sheet, the informed consent form and the interview data collection form and documents can be found at http://doi.org/10.17862/cranfield.rd.12279152.

[7] Elements of the table that represent survey's responses are in *Italic*.

Appendix D
Repair case study: experimental procedure and Augmented Reality solution

CURES Reference: CURES/9145/2019

Data repository: https://doi.org/10.17862/cranfield.rd.12213380

This case study comprises an experimental maintenance operation to support and an AR solution (ARAUM) to compare against the proposal made in Chapter 4.

The case study proposed by Erkoyuncu *et al.* (2017), considers repair operations in complex engineering assets for the Defence Industry. These are focused mainly in mechanical, electric and hydraulic systems and assembly and replacement procedures. The case study equipment is a laboratory prototype of a gearbox for studying gear-wheels degradation that is utilised to represent real-life conditions of asset-repair scenarios. The experimental operation focuses on a specific repair operation composed of several assembly, disassembly and replacement steps involving mechanical components. The instructions for those steps are presented in Table D-1 and apply to the gearbox displayed in Figure D-1. The experiment consists of following the steps' instructions to complete the repair operation using either the case study's AR solution (ARAUM), non-AR information-delivery methods (maintenance manuals) or Chapter 4's AR proposal (PMAU).

Table D-1. Experimental procedure of the repair case study

Step	Instruction
R1	Unscrew and remove transparent cover
R2	Unscrew and remove wheel break
R3	Re-place and re-screw wheel brake
R4	Place back and re-screw transparent cover
Operation	"Change the brake"
Tools	Screwdriver no.57, Allen Key no.24
Items	Wire no. 18, Gloves
Safety precautions	Maintenance works may require lifting heavy loads. The conditions for these works are not always ideal. Parts may not be within easy reach, access may be poor or there may not be sufficient space to move. Floors may be slippery or cables might be in the way, and work may be performed at low heights.

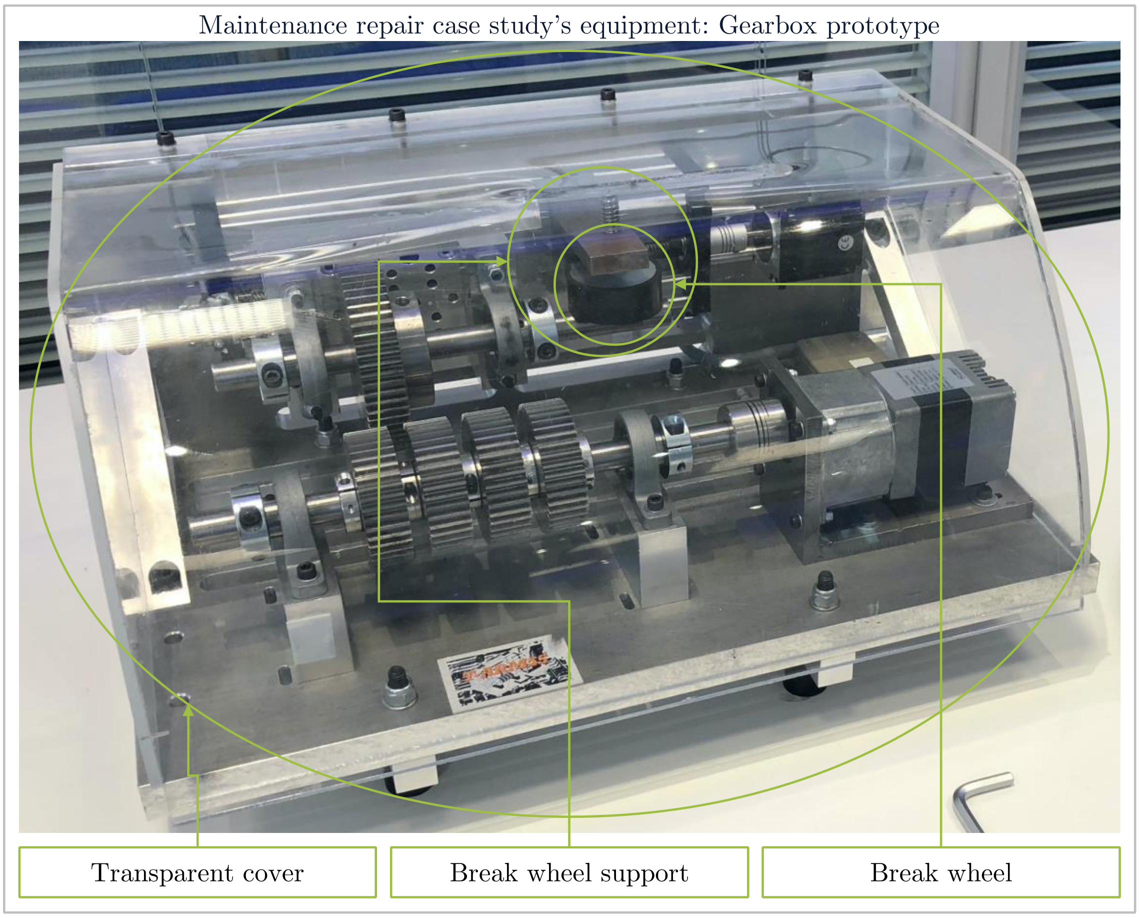


Figure D-1. Picture of repair case study's equipment – The image shows the gearbox prototype and indicates its components related to the case study's experimental operation.

The case study's AR solution (ARAUM) consists of an AR application built based on a conventional expert authoring solution (Figure 4-1) for maintenance repair operations. This AR application is meant for Hand-Held Devices (HHD) and utilises a specific database for experts to generate AR content. This is the database (Figure D-2) that was utilised by the student to produce the case study's ontology (repont – Figure 4-7).

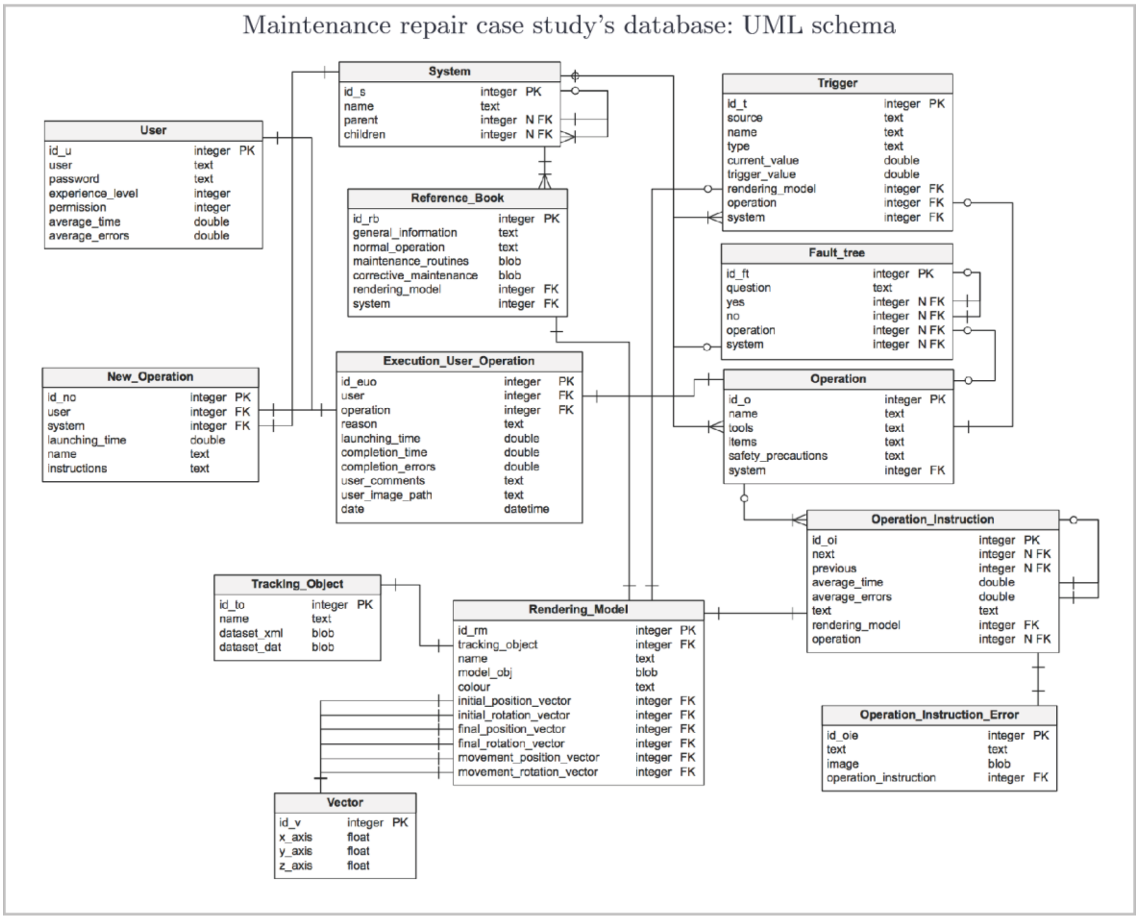


Figure D-2. UML representation of repair case study's database – The figure shows the UML schema for the case study's database in which the case study's ontology is based upon (Figure 4-7).

The case study's database includes more information that the data utilised within the experiment. This is because the experiment focused on a repair scenario following pre-determined instructions while the database includes additional data regarding asset's normal functioning (e.g. 'Books of Reference'), failure identification (e.g. 'Fault Tree') and some other repair-related data.

The experimental repair scenario is defined to follow the 'Operation' "Change the brake". This 'Operation' aims at replacing one of the gearbox's components when it has been worn away. Table D-1 presents the repair 'Operation' and its steps as 'Operation_Instructions'. 'Operation_Instructions' are augmented through text descriptions ('text') and animations ('Rendering_Model').

Additional data from the database (Figure D-2) is also augmented for experiment purposes. Each 'Operation' includes a 'tools', 'items' and 'safety_precautions' that are displayed as text in the AR application. Each 'instruction' is also delivered by AR means through a textual description and an additional animation overlaid on top of the real-world object imitating the movement to be done for conducting the repair step. Time and errors for each 'instruction' are measured while conducting the experiment. Figure D-3 presents an example of an augmented 'instruction' ('Remove the brake piece and replace by new one') using this case study's AR solution.

Maintenance repair case study's AR solution

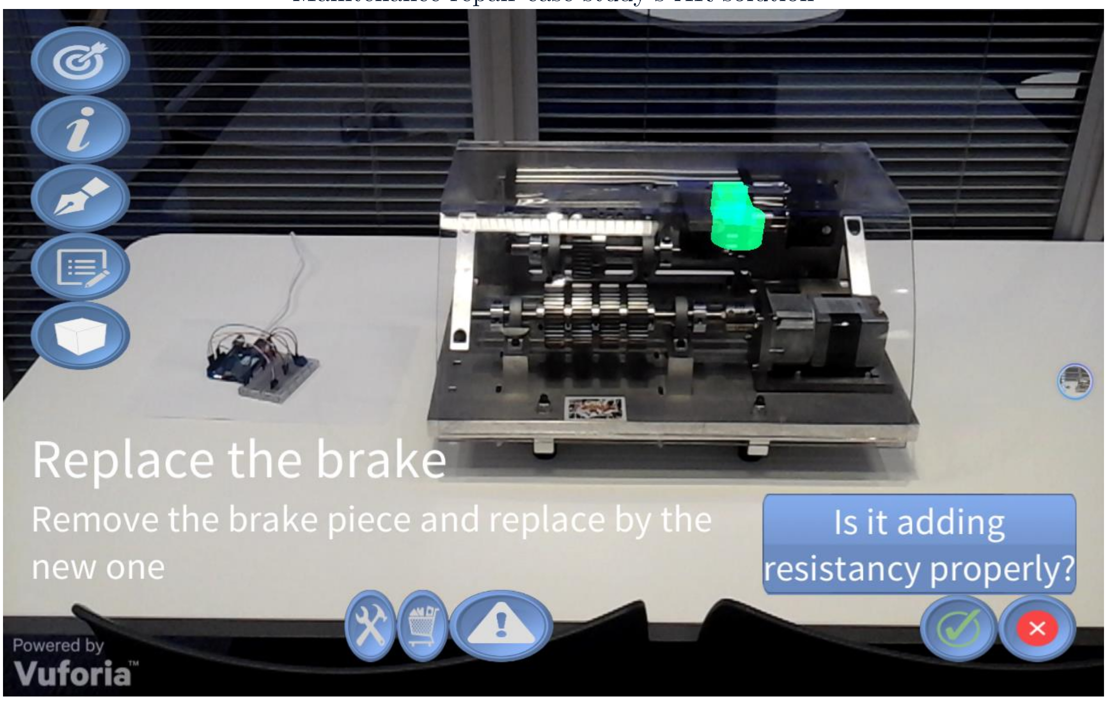


Figure D-3. Screenshot of repair case study's AR solution (ARAUM) – The figure displays an instruction of the case study's experimental operation (R3 – Table D-1) using the case study's AR solution.

A more detailed description of the case study's AR solution can be found at http://doi.org/10.1016/j.cirp.2017.04.006.

The complete experimental procedures including information sheets, informed consent and data collection forms and additional documents can be found at https://doi.org/10.17862/cranfield.rd.12213380.

Appendix E
Remote diagnosis case study: experimental procedure and Augmented Reality solution

CURES Reference: CURES/9145/2019

Data repository: https://doi.org/10.17862/cranfield.rd.12213380

This case study comprises an experimental maintenance operation to support and an AR solution (SMAARRC) to compare against the proposal in Chapter 4.

The case study proposed by Fernández del Amo *et al.* (2019), considers remote diagnosis operations in complex engineering assets for the Aerospace Industry. The focus of these operations is purely in mechanical systems. The case study's equipment is an aircraft's fuel hatch prototype with unidentified imperfections that are the diagnosis target. The experimental operation focuses on a remote diagnosis operation that comprises inspection, measurement and repair of mechanical components. The instructions for those steps are presented in Table D-1 and apply to the fuel hatch displayed in Figure D-1. The experiment consists following the steps to be sent by the expert to complete a remote diagnosis operation either through usual communication methods (phone calls and emails), Chapter's 4 AR proposal (PMAU) or the case study's AR solution (SMAARRC).

Table E-1. Experimental procedure of the remote diagnosis case study

Step	Instruction
D1	Expert asks to unscrew the screws of the front panel of the fuel hatch and open it
D2	Expert asks to visually inspect the right and left sides of the hatch and to take a photograph of every defect found
D3	[Two defects should be found by tester] Expert asks to repair by placing the patch
D4	Expert asks to take a photograph of the previous reparation result and send it

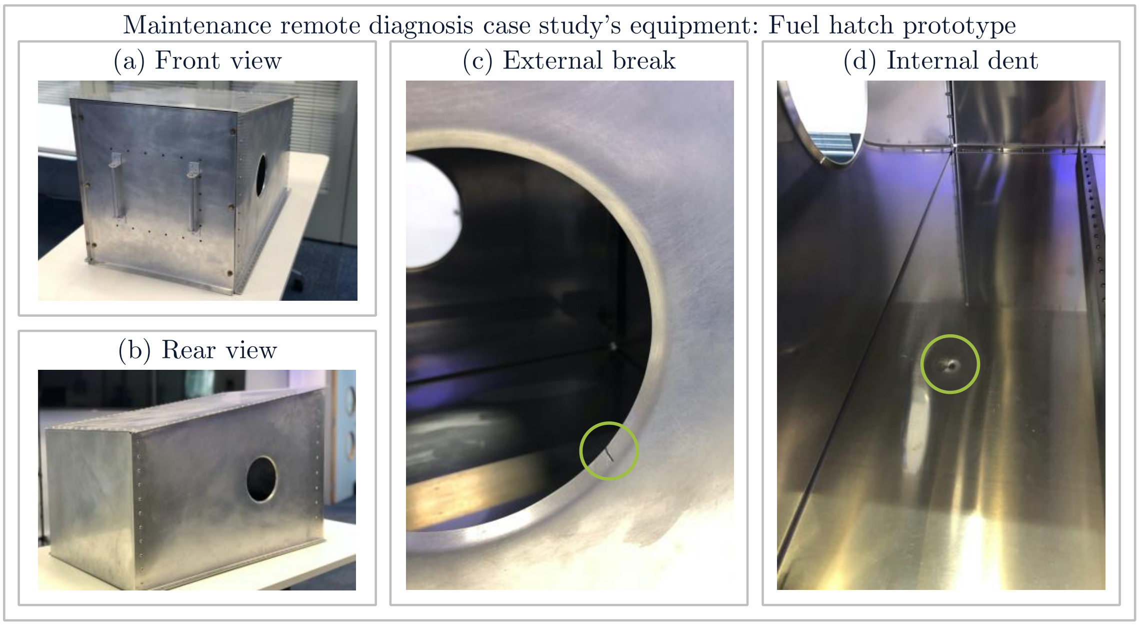


Figure E-1. Pictures of remote diagnosis case study's equipment – The figures shows the fuel hatch prototype from different views and indicates its defects to be found (c and d - D2 - Table D-1) and repaired (c - D3 - Table D-1) as part of the case study's experimental operation.

The case study's AR solution (SMAARRC) consists of an AR application built based on a real-time expert authoring solution (Figure 4-1) for maintenance remote diagnosis operations. This AR application is meant for Head-Mounted Devices (HMD) and utilises a specific interface for experts to generate AR content. As a communication-support AR tool, the case study's database represents the elements declared for the specific remote diagnosis messages. This is the database (Table E-2) that was utilised by the student to produce the case study's ontology (remont – Figure 4-7).

Table E-2. Tabular representation of remote diagnosis case study's database – The table includes the different elements that conforms a message, their definitions, the potential values these can get and their AR visualisation modes (Text (T), Holograms (H), Pictures (P) and Measurements (M)).

Elements	Definition	Values	T	H	P	M
Sender	Person that sends a message	Expert or Technician	X			
Recipient	Person(s) that receive(s) a message	Expert(s) or Technician(s)	X			
Type	Aim for which the message is being sent	Action, Confirmation, Question, or Response	X			
Component	Equipment's part a message refers to	Component's name in equipment's CAD model	X	X		
Location	Place to where a message refers to	3D position and rotation from equipment's origin	X	X	X	
Identifier	Order of a message in a call	Integer / timestamp	X			
Category	Call's context in which a message is being sent	Definition or Validation	X			
Action	Method to conduct a procedure being defined	Pull, Push, Screw, Inspect, Measure, Photograph...	X			
Measure	Magnitude with which the method is applied	Quantitative measure	X	X	X	X
		Qualitative measure	X	X	X	
Object	Additional elements that complete a message	Free text	X			

The database does not store pre-identified information *per se,* but the reported messages generated by the AR-supported communication. Therefore, the message elements and the visualisation modes for each of them declare the database structure. Table E-2 presents both. Each message element has different visualisation modes, which are the real-time authoring rules given to the expert as a desktop application to send messages to the technician for conducting the remote diagnosis using a head-mounted device. An example of a message being sent through SMAARRC's AR remote-communication solution is presented in Figure D-2, where the expert asks the technician to identify defects in the fuel hatch's interior (Table D-1 – D2).

As the repair case study, the remote diagnosis case study considers an interface to input data for content creation. This is a desktop application for the remote expert (Figure D-2-a) to send the messages to the AR-supported technician (Figure D-2-b). This desktop application comprises a 3D model view where to generate the messages and the technician's live streaming. For this research experiments, the expert application will comprise an ontology interface to send messages, a 3D model view and the technician's live streaming.

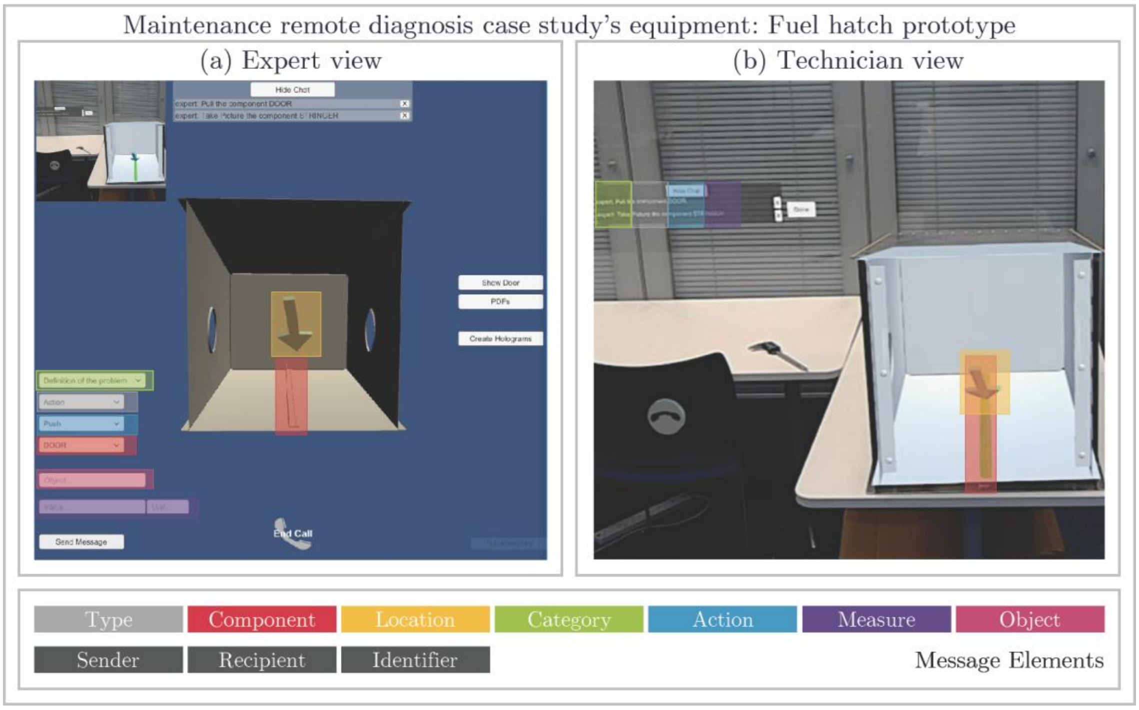


Figure E-2. Screenshot of remote diagnosis case study's AR solution (SMAARRC) – The figure displays an instruction of the case study's experimental operation (D1 – Figure D-1) using the case study's AR solution from both, (a) expert and (b) technician views, including the different message elements to be sent.

A more detailed description of the case study's AR solution can be found at https://doi.org/10.1016/j.aei.2020.101096.

The complete experimental procedures including information sheets, informed consent and data collection forms and additional documents can be found at https://doi.org/10.17862/cranfield.rd.12213380.

Appendix F
Usability surveys on manually versus automatically authored AR solutions: procedure and questionnaire

CURES Reference: CURES/9145/2019

Data repository: https://doi.org/10.17862/cranfield.rd.12213380

A relevant method to assess the usability of a software tool is to allow targeted users to try it and then ask them to state their opinion regarding such matter (Nielsen, 1993). The student proposed to conduct usability surveys (Subsection 4.5.2.2) on all AR applications utilised as part of the stopwatch time and errors experiments described in Subsection 4.5.2.1. The surveys' aim was to compare the usability of automatically authored AR applications (PMAU) against manually authored applications (ARAUM and SMAARRC) to identify any perceived differences by testers concerning their content. So, it could be demonstrated that PMAU automatically created content was as usable as manually created content from alternative AR solutions (ARAUM and SMAARRC). Table A-1 presents the survey that testers were asked to complete after conducting the abovementioned experiments. It comprises the statements that testers were asked to state their agreement with according to the usability criterions considered. The survey utilises a Likert Scale format of 1-5 to take into consideration the results presented by Weijters, Cabooter and Schillewaert (2010). These suggested that such format maximises potential information transmission when surveying non-expert populations.

Table F-1. Survey questionnaire on manually versus automatically authored AR solutions' usability[8]

In this survey, you will be questioned about the usability of the AR solutions you experimented with for each of the maintenance operations conducted. This survey utilises five aspects of usability to identify your opinion on each AR solution tested:	
1	Ease-to-learn: ability of the solution to show its functioning by itself.
2	Ease-to-use: ability of the solution to be self-understandable.
3	Accuracy: ability of the solution to display AR content in its right position.
4	Effectiveness: ability of the solution to support the maintenance operation.
5	Satisfaction: your overall impression after using the AR solution.
Each of these aspects is evaluated from an overall perspective and specifically for each AR content type used. Please, complete the survey below using a ranking from 1 (worst / strongly disagree) to 5 (best / strongly agree) to punctuate each AR solution according to each of these aspects. When comparing Solution 1 versus Solution 2, please re-call which solution you used in your first and second experiments: ARAUM, SMAARCC or PMAU.	

EASE-TO-LEARN						
Please, rank your answers from 1 (worst) to 5 (best).						
How easy was to use the AR solution when the experiment started?	Solution 1	1	2	3	4	5
	Solution 2	1	2	3	4	5
How easy was to use the AR solution when the experiment finished?	Solution 1	1	2	3	4	5
	Solution 2	1	2	3	4	5
How intuitive was to use the AR solution?	Solution 1	1	2	3	4	5
	Solution 2	1	2	3	4	5
EASE-TO-USE						
Please, evaluate each AR solution's feature according to its ease of use.						
Buttons	Solution 1	1	2	3	4	5
	Solution 2	1	2	3	4	5
Gestures	Solution 1	1	2	3	4	5
	Solution 2	1	2	3	4	5
Dictation	Solution 1	1	2	3	4	5
	Solution 2	1	2	3	4	5

[8] Elements of the table that represent survey's responses are in *Italic*.

Text	Solution 1	1	2	3	4	5
	Solution 2	1	2	3	4	5
Images	Solution 1	1	2	3	4	5
	Solution 2	1	2	3	4	5
Holograms	Solution 1	1	2	3	4	5
	Solution 2	1	2	3	4	5
Animations	Solution 1	1	2	3	4	5
	Solution 2	1	2	3	4	5
What did you find more frustrating about the AR solution?	Solution 1					
	Solution 2					
What did you like the most about the AR solution?	Solution 1					
	Solution 2					
ACCURACY						
Please, rank you answers from 1 (strongly disagree) to 5 (strongly agree).						
The content is overlaid accurately in the AR solution.	Solution 1	1	2	3	4	5
	Solution 2	1	2	3	4	5
The content does not shake when moving around.	Solution 1	1	2	3	4	5
	Solution 2	1	2	3	4	5
The content does not occlude the view of the equipment while maintaining it.	Solution 1	1	2	3	4	5
	Solution 2	1	2	3	4	5
The content was easy to watch or read.	Solution 1	1	2	3	4	5
	Solution 2	1	2	3	4	5
The AR solution has a correct response time (e.g. when clicking a button).	Solution 1	1	2	3	4	5
	Solution 2	1	2	3	4	5
EFFECTIVENESS						
Please, rank you answers from 1 (strongly disagree) to 5 (strongly agree).						
The content was easy to understand.	Solution 1	1	2	3	4	5
	Solution 2	1	2	3	4	5
The AR solution can help to reduce errors in maintenance tasks.	Solution 1	1	2	3	4	5
	Solution 2	1	2	3	4	5
The AR solution can help to not miss instructions for maintenance tasks.	Solution 1	1	2	3	4	5
	Solution 2	1	2	3	4	5

The AR solution can help to improve efficiency of maintenance tasks.	Solution 1	1	2	3	4	5
	Solution 2	1	2	3	4	5
The AR solution can help to increase your confidence when performing a task.	Solution 1	1	2	3	4	5
	Solution 2	1	2	3	4	5
The AR solution provides content that is suitable for the tasks to be performed.	Solution 1	1	2	3	4	5
	Solution 2	1	2	3	4	5
SATISFACTION						
Overall, how well designed do you believe the AR solution is?	Solution 1	1	2	3	4	5
	Solution 2	1	2	3	4	5
Overall, how useful did you find the AR solution?	Solution 1	1	2	3	4	5
	Solution 2	1	2	3	4	5
Overall, how satisfied were you with the AR solution?	Solution 1	1	2	3	4	5
	Solution 2	1	2	3	4	5

Additional comments	Solution 1	
	Solution 2	

The complete survey sheets including the information sheet, the informed consent form and the survey data collection form and documents can be found at https://doi.org/10.17862/cranfield.rd.12213380.

Appendix G
Failure diagnosis reporting experiments: procedure and reporting solutions

CURES Reference: CURES/9447/2019

Data repository: https://doi.org/10.17862/cranfield.rd.12382604

Failure diagnosis reporting experiments aim at evaluating testers reporting time and errors on two failure scenarios (Subsection 5.5.3) when utilising different reporting solutions (Subsection 5.5.2.1).

Although testing diverse failure scenarios, diagnosis reporting experiments consist of the same steps for testers. These include the completion of up to three reports using individual-based forms from two diagont classes (*'Step'* and *'State'*) as seen in Figure 5-13 and Figure 5-14. Table D-1 presents the forms (classes) and its items (attributes and relationships) that testers are required to fulfil (assert) as part of diagnosis reporting experiments.

Table G-1. Ontology-based forms from failure diagnosis reporting experiments[9]

Form	Type	Name	Response
1	Class	Step	*Automatic*
	Attribute	isCritical	*{true, false}*
	Attribute	isContributory	*{true, false}*
	Attribute	hasObject	*{symptom, trace, cause}*
	Attribute	hasMethod	*{inspect, measure, repair, replace}*

[9] Elements of the table that represent testers responses are in *Italic*.

Form	Type	Name	Response
	Attribute	hasComparison	*{equalTo, notEqualTo, greaterThan, ...[10]}*
	Relationship	belongsTo	*{new, list[11]}*
	Relationship	evaluates	*{new[12], list}*
	Relationship	diagnoses	*{new[13], recommendations[14]}*
2	Class	State (evaluates)	*Automatic*
	Attribute	hasStatus	*{normal, safe-degraded, unsafe-degraded, faulty}*
	Attribute	hasDomain	*{mechanics, electrics, electronics, hydraulics, ...}*
	Attribute	hasPhenomenon	*{fracture, fatigue, corrosion, impact, blockage, ...}*
	Attribute	hasMeasureValue	*Numeric (free-text)*
	Attribute	hasMeasureUnit	*{metre, degree, kilogram, second, newton, ...}*
	Attribute	hasMeasureDate	*Automatic*
	Relationship	causes	*{new, list}*
	Relationship	describes	*{new, list}*
	Relationship	measuredBy	*{new, list}*
3	Relationship	State (diagnoses)	*Automatic*
	Attribute	hasStatus	*{normal, safe-degraded, unsafe-degraded, faulty}*
	Attribute	hasDomain	*{mechanics, electrics, electronics, hydraulics, ...}*
	Attribute	hasPhenomenon	*{fracture, fatigue, corrosion, impact, blockage, ...}*
	Attribute	hasMeasureValue	*Numeric (free-text)*
	Attribute	hasMeasureUnit	*{metre, degree, kilogram, second, newton, ...}*
	Attribute	hasMeasureDate	*Automatic*
	Relationship	causes	*{new, list}*
	Relationship	describes	*{new, list}*
	Relationship	measuredBy	*{new, list}*

[10] All responses including "..." refer to the additional values presented in

Table 3-2.

[11] Refers to the list of all available individuals asserted to the 'State' class in the knowledge base.

[12] If new is selected, then the correspondent new form (2) is generated.

[13] If new is selected, then the correspondent new form (3) is generated.

[14] Refers to the list of recommended individuals when the reporting solution is either TBR or ARR.

Chapter 5 proposes an AR solution with hybrid recommendations that can enhance efficiency and effectiveness of diagnosis reporting operations. In order to demonstrate so, diagnosis reporting experiments evaluate several variables (time, errors and recommendations accuracy) to compare the effect of diverse reporting tools as described in Subsection 5.5.2.1. These reporting tools include the one proposed in Section 5.4, along with others proposed in previous Chapters. These tools are the following:

- ARR (RPMAU): the proposed AR solution presented in Chapter 5. It includes AR visualisation and hybrid recommendations for diagnosis reporting.
- ARN (PMAU): the proposed AR solution presented in Chapter 4. It includes AR visualisation but does not include recommendations.
- TBN: the proposed reporting tool presented in Chapter 3. It creates ontology-based forms but does not include AR visualisation nor recommendations.
- TBR: a solution implemented to compare ARR with non-AR recommender solutions. It is similar to TBN but includes ontology-based recommendations as described in Figure 5-8.

Table D-1 presents an overview of the reporting tools utilised in these experiments according to their features: AR visualisation and recommendations.

Table G-2. Classification of experimental reporting tools according to inclusion of AR visualisation and/or recommendations as part of their features.

<table>
<tr><td colspan="2"></td><td colspan="2">Recommendations</td></tr>
<tr><td colspan="2"></td><td>No</td><td>Yes</td></tr>
<tr><td rowspan="2">AR visualisation</td><td>No</td><td>TBN</td><td>TBR</td></tr>
<tr><td>Yes</td><td>ARN</td><td>ARR</td></tr>
</table>

The reasons to tests four different solutions are two. First, recommendations for diagnosis reporting operations can be convenient because they can help to reduce reporting time and errors by suggesting pre-determined responses to be used by reporters. Second, AR technologies can help not only to reduce reporting time and errors by increasing visualisation, but also to enhance recommendations accuracy by utilising contextual data. Therefore, these experiments aim to analyse the impact of these two effects by testing the four abovementioned reporting tools.

Figure D-1 presents examples of these tools to provide an overview of the experimental procedure. It includes screenshots of AR and non-AR solutions when being used to complete each of the forms described in Table D-1. These three steps comprise the experimental procedure. Recommendations are only used in Step 1 (Figure D-1), where the recommender solutions (ARR and TBR) utilised previously reported forms (Figure 5-13 and Figure 5-14) as the context to suggest which *'State'* to *'diagnose'* (Step 3). If the tester selects a recommendation, then it will save reporting time and will also reduce errors in case the recommendation selected is the correct one.

The complete experimental procedures including information sheets, informed consent and data collection forms and additional documents can be found at https://doi.org/10.17862/cranfield.rd.12382604.

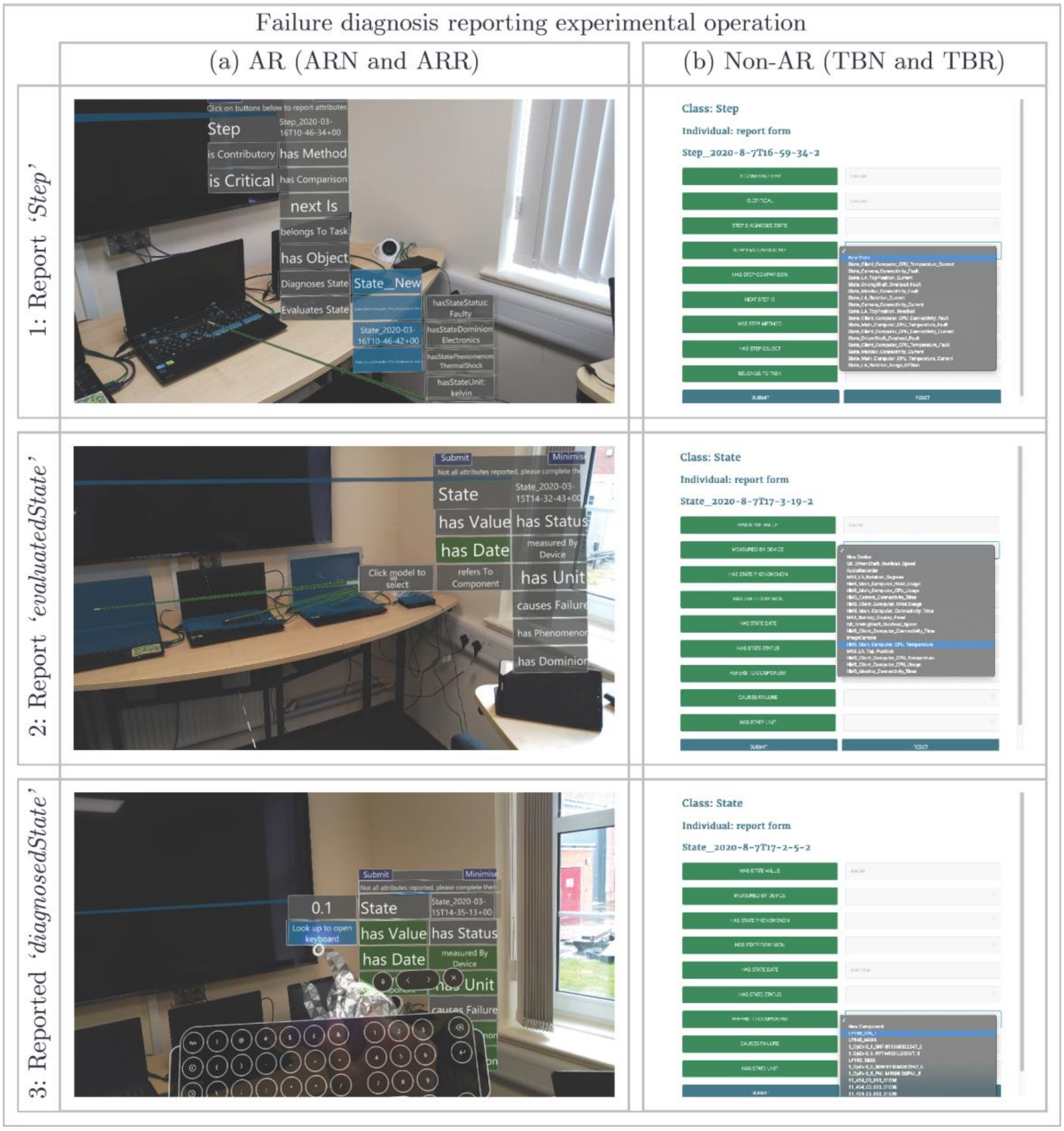


Figure G-1. Failure diagnosis reporting experimental operation – The figure displays several screenshots of (a) AR and (b) non-AR solutions being used at each stage of the experimental operation. These stages are three and each aims to report a diagont individual of each class involved in reporting a fault: (1) report 'Step', (2) report 'evaluatedState' and (3) report 'diagnosedState'. Hybrid recommendations, which differ among TBR and ARR solutions, are used on step 1. If the relevant 'diagnosedState' is found within the given recommendations, then step 3 is automatically completed saving from reporting time and errors.

Appendix H
Workload surveys on diagnosis reporting tasks: procedure and questionnaire

CURES Reference: CURES/9447/2019

Data repository: https://doi.org/10.17862/cranfield.rd.12382604

A relevant method to assess testers perceived performance on a given task is using workload surveys (Hart, 2006). NASA Task Load indeX (NASA-TLX) surveys. NASA-TLX is a standard questionnaire developed by NASA Ames Research (Hart, 2006) to collect workload self-evaluation data from experimental testers. Workload surveys consist of two, self-rating steps. Prior to experiments, testers are asked to evaluate the relative importance of each workload aspect given an understanding of the tasks to perform. Table H-1 presents the questionnaire that testers were given to provide that evaluation. After the experiments, testers are asked to complete a second questionnaire to quantitatively evaluate the importance of each aspect independently. Table H-2 presents the statements that testers were asked to complete for that evaluation.

Table H-1. Pre-experimental NASA TLX survey on workload factors' relative importance (Hart, 2006)

Regarding the reporting tasks you were explained to conduct in the experiment:

Please, determine which workload aspects may be more important for this task.

Please, circle the most important aspect for each pair according to your opinion.

Pair 1	Effort	or	Performance
Pair 2	Temporal Demand	or	Frustration
Pair 3	Temporal Demand	or	Effort
Pair 4	Physical Demand	or	Frustration
Pair 5	Performance	or	Frustration
Pair 6	Physical Demand	or	Temporal Demand
Pair 7	Physical Demand	or	Performance
Pair 8	Temporal Demand	or	Mental Demand
Pair 9	Frustration	or	Effort
Pair 10	Performance	or	Mental Demand
Pair 11	Performance	or	Temporal Demand
Pair 12	Mental Demand	or	Effort
Pair 13	Mental Demand	or	Physical Demand
Pair 14	Effort	or	Physical Demand
Pair 15	Frustration	or	Mental Demand

Workload aspect	Definition
Mental Demand	How much mental and perceptual activity is required (e.g. thinking. deciding, calculating, remembering. looking, searching, etc)? Is the task easy or demanding, simple or complex, exacting or forgiving?
Physical Demand	How much physical activity is required (e.g. pushing, pulling, turning. controlling, activating, etc.)? Is the task easy or demanding, slow or brisk, slack or strenuous restful or laborious?
Temporal Demand	How much time pressure do you feel due to the rate or pace at which the tasks or task elements occurr? Is the pace slow and leisurely or rapid and frantic?
Performance	How successful do you think you are in accomplishing the goals of the task set by the experimenter? How satisfied are you with your performance in accomplishing these goals?
Effort	How hard do you have to work (mentally and physically) to accomplish your level of performance?
Frustration	How insecure, discouraged, irritated. stressed and annoyed versus secure, gratified, content, relaxed and complacent do you feel during the task?

Table H-2. Post-experimental NASA TLX survey on workload factors' importance (Hart, 2006)

Please, on a scale from very low to very high place a mark on the box that better states your opinion on the questions below.

Mental Demand																			How mentally demanding was the task?
Very low																			Very high

Physical Demand																			How physically demanding was the task?
Very low																			Very high

Temporal Demand																			How hurried or rushed was the pace of the task?
Very low																			Very high

Performance																			How successful were you in accomplishing what you were asked to?
Very low																			Very high

Effort																			How hard did you have to work to accomplish your level of performance?
Very low																			Very high

Frustration																			How insecure, discouraged, irritated, stressed and annoyed were you?
Very low																			Very high

The complete survey sheets including the information sheet, the informed consent form and the survey data collection form and documents can be found at https://doi.org/10.17862/cranfield.rd.12382604.

Appendix I
Usability surveys on AR versus non-AR reporting solutions: procedure and questionnaire

CURES Reference: CURES/9447/2019

Data repository: https://doi.org/10.17862/cranfield.rd.12382604

A relevant method to assess the usability of a software tool is to allow targeted users to try it and then ask them to state their opinion regarding such matter (Nielsen, 1993). The student proposed to conduct usability surveys (Subsection 5.5.2.3) on all reporting tools utilised as part of diagnosis reporting experiments described in Subsection 5.5.2.1. The surveys' aim was to compare the usability of AR (ARR, ARN) against non-AR tools (TBR, TBN) to identify any perceived differences by testers concerning their content. So, it could be demonstrated that AR content was as usable as web content to perform diagnosis reporting tasks. Table A-1 presents the survey that testers were asked to complete after conducting the abovementioned experiments. It comprises the statements that testers were asked to state their agreement with according to the usability criterions considered. The survey utilises a Likert Scale format of 1-5 to take into consideration the results presented by Weijters, Cabooter and Schillewaert (2010). These suggested that such format maximises potential information transmission when surveying non-expert populations.

Table I-1. Survey questionnaire on AR vs non-AR reporting tools' usability[15]

In this survey, you will be questioned about the usability of the reporting tools you experimented with for each of the reporting operations conducted. This survey utilises five aspects of usability to identify your opinion on each reporting tool tested:							
1	Ease-to-learn: ability of the tool to show its functioning by itself.						
2	Ease-to-use: ability of the tool to be self-understandable.						
3	Accuracy: ability of the tool to display content in its right position.						
4	Effectiveness: ability of the tool to support the maintenance operation.						
5	Satisfaction: your overall impression after using the tool.						
Each of these aspects is evaluated from an overall perspective and specifically for each AR content type used. Please, complete the survey below using a ranking from 1 (worst / strongly disagree) to 5 (best / strongly agree) to punctuate each AR solution according to each of these aspects. When comparing Tool 1 versus Tool 2, please re-call which solution you used in your first and second experiments: TBN, TBR, ARN or ARR.							
EASE-TO-LEARN							
Please, rank your answers from 1 (worst) to 5 (best).							
How easy was to use the tool when the experiment started?		*Tool 1*	1	2	3	4	5
		Tool 2	1	2	3	4	5
How easy was to use the tool when the experiment finished?		*Tool 1*	1	2	3	4	5
		Tool 2	1	2	3	4	5
How intuitive was to use the tool?		*Tool 1*	1	2	3	4	5
		Tool 2	1	2	3	4	5
EASE-TO-USE							
Please, evaluate each tool's feature according to its ease of use.							
Buttons		*Tool 1*	1	2	3	4	5
		Tool 2	1	2	3	4	5
Gestures		*Tool 1*	1	2	3	4	5
		Tool 2	1	2	3	4	5
Dictation		*Tool 1*	1	2	3	4	5
		Tool 2	1	2	3	4	5
Keyboard		*Tool 1*	1	2	3	4	5
		Tool 2	1	2	3	4	5

[15] Elements of the table that represent survey's responses are in *Italic*.

Text	Tool 1	1	2	3	4	5
	Tool 2	1	2	3	4	5
3D Models (Holograms)	Tool 1	1	2	3	4	5
	Tool 2	1	2	3	4	5
What did you find more frustrating about the tool?	Tool 1					
	Tool 2					
What did you like the most about the tool?	Tool 1					
	Tool 2					
ACCURACY						
Please, rank you answers from 1 (strongly disagree) to 5 (strongly agree).						
The content is overlaid accurately in the tool.	Tool 1	1	2	3	4	5
	Tool 2	1	2	3	4	5
The content does not shake when moving around.	Tool 1	1	2	3	4	5
	Tool 2	1	2	3	4	5
The content does not occlude the view of the equipment while maintaining it.	Tool 1	1	2	3	4	5
	Tool 2	1	2	3	4	5
The content was easy to watch or read.	Tool 1	1	2	3	4	5
	Tool 2	1	2	3	4	5
The tool has a correct response time (e.g. when clicking a button).	Tool 1	1	2	3	4	5
	Tool 2	1	2	3	4	5
EFFECTIVENESS						
Please, rank you answers from 1 (strongly disagree) to 5 (strongly agree).						
The content was easy to understand.	Tool 1	1	2	3	4	5
	Tool 2	1	2	3	4	5
The tool can help to not miss steps in diagnosis reporting tasks.	Tool 1	1	2	3	4	5
	Tool 2	1	2	3	4	5
The tool can help to accurately report steps of diagnosis tasks.	Tool 1	1	2	3	4	5
	Tool 2	1	2	3	4	5
The tool can help to improve efficiency of diagnosis reporting tasks.	Tool 1	1	2	3	4	5
	Tool 2	1	2	3	4	5
The tool can help to increase your confidence when performing a task.	Tool 1	1	2	3	4	5
	Tool 2	1	2	3	4	5

The tool provides content that is suitable for the tasks to be performed.	Tool 1	1	2	3	4	5
	Tool 2	1	2	3	4	5
SATISFACTION						
Overall, how well designed do you believe the tool is?	Tool 1	1	2	3	4	5
	Tool 2	1	2	3	4	5
Overall, how useful did you find the tool?	Tool 1	1	2	3	4	5
	Tool 2	1	2	3	4	5
Overall, how satisfied were you with the tool?	Tool 1	1	2	3	4	5
	Tool 2	1	2	3	4	5

Additional comments	Tool 1	
	Tool 2	

The complete survey sheets including the information sheet, the informed consent form and the survey data collection form and documents can be found at https://doi.org/10.17862/cranfield.rd.12382604.