Sustainable Production through Advanced Manufacturing, Intelligent Automation and Work Integrated Learning, J. Andersson et al. (Eds.)

doi:10.3233/ATDE240151

Combining Ontology and Large Language Models to Identify Recurring Machine Failures in Free-Text Fields

Marcus BENGTSSON[a, b, 1], Ricky Stanley D'CRUZE[b], Mobyen Uddin AHMED[b], Tomohiko SAKAO[c], Peter FUNK[b], Rickard SOHLBERG[b]
[a] *Volvo Construction Equipment Operations Eskilstuna Sweden*
[b] *School of Innovation, Design and Engineering, Mälardalen University*
[c] *Department of Management and Engineering, Linköping University*
ORCiD ID: Marcus Bengtsson https://orcid.org/0000-0002-0729-0122

Abstract. Companies must enhance total maintenance effectiveness to stay competitive, focusing on both digitalization and basic maintenance procedures. Digitalization offers technologies for data-driven decision-making, but many maintenance decisions still lack a factual basis. Prioritizing efficiency and effectiveness require analyzing equipment history, facilitated by using Computerized Maintenance Management Systems (CMMS). However, CMMS data often contains unstructured free-text, leading to manual analysis, which is resource-intensive and reactive, focusing on short time periods and specific equipment. Two approaches are available to solve the issue: minimizing free-text entries or using advanced methods for processing them. Free-text allows detailed descriptions but may lack completeness, while structured reporting aids automated analysis but may limit fault description richness. As knowledge and experience are vital assets for companies this research uses a hybrid approach by combining Natural Language Processing with domain specific ontology and Large Language Models to extract information from free-text entries, enabling the possibility of real-time analysis e.g., identifying recurring failure and knowledge sharing across global sites.

Keywords. Industrial Maintenance, Artificial Intelligence, Natural Language Processing, Large Language Models, Experience Reuse

1. Introduction

Ensuring optimal machine utilization and availability in various lifecycle phases such as production and operation is a crucial prerequisite for manufacturing companies aiming to maintain a competitive edge. The principles of lean production, such as flow layout and minimizing buffers, have highlighted the importance of this endeavor, as has the growing emphasis on sustainability. Maintaining machine conditions and repairing them when necessary is inevitable, too, since physical deterioration cannot be avoided in the physical world. Also, maintenance of products plays a significant role from the environmental sustainability perspective [1]. How maintenance is performed influences significantly the service life of a product, equipment, or asset. Lifetime extension is an

[1] Corresponding Author: Marcus Bengtsson, marcus.bengtsson@mdu.se.

effective way to increase the efficiency of the resource of the products and is considered preferred to material recycle, sitting at the most prioritized position of the waste hierarchy (reduce, reuse, and recycle) [2]. With the increasing interest in the access to material resources (including rare metals) [3], the role of maintenance is getting larger.

To remain competitive, companies must constantly enhance and improve their total maintenance effectiveness. It is important to note that these improvements should not be limited to digitalization efforts alone but should also encompass improvements in basic maintenance procedures [4-5]. The good news is that advancements in digitalization often offer innovative technologies that also can improve basic maintenance, particularly in terms of data-driven decision-making [6]. Despite this potential, many maintenance decisions continue to rely on non-fact-based approaches [7]. Also, it has been reported that reusing experiences as reported by maintenance employees are limited [8]. To improve this situation, companies are advised to analyze their current state and the historical behavior of their equipment [9]. Employing a Computerized Maintenance Management System (CMMS) is essential as a tool in this analytical process [10-11].

Effective logging of failure causes in the CMMS is key to reducing machine failures and recurring issues [12]. However, the utilization of CMMS for maintenance record analysis remains limited [11]. One major challenge associated with CMMS data is that it often consists of free-text, making it difficult to analyze using automated systems. As a result, manual analysis is frequently employed, consuming valuable resources and time, with limited outcomes in terms of time-period coverage and cross-case analysis [13]. Typically, only short time periods and specific machine equipment are analyzed, rather than considering a broader selection of machines and their operational history. Furthermore, manual analysis is also reactive, as recurring failures have usually occurred by the time data is extracted for analysis.

Moving forward, there are two potential approaches to improve the situation [14]. The first involves minimizing or avoiding the use of free-text entries in failure logs, while the second focuses on adopting sophisticated and advanced methods to process and interpret free-text data. Each approach has its own strengths and weaknesses. Allowing free-text entries in failure logs enables technicians to describe rare or complex faults in detail, facilitating the transfer of valuable experience to co-workers. However, it also opens the possibility of incomplete or insufficient information, leading to variations in content quality. Reducing free-text and implementing highly structured reporting would enhance the accessibility of failure reports for automated processing and analysis. Nevertheless, this approach could restrict the description of faults and actions, potentially resulting in the loss of valuable experience and knowledge, especially in complex domains. Knowledge and experience are increasingly crucial factors in companies and embedded in company's value, and maintenance knowledge may well be classified as a part of the intellectual capital as defined in [15].

In our research, we have employed Natural Language Processing (NLP) techniques with a combination of a domain specific Ontology, leveraging computational linguistics and Artificial Intelligence (AI), and Machine Learning (ML), to understand the meaning of free-text entries and extract essential information from them. This approach takes advantage of Large Language Models (LLM) with their multilingual capabilities and support. It enables the real-time analysis of failure reports not only within a single manufacturing site but also at an international multi-site level. For instance, if a recurring failure has been previously reported in a plant in Sweden, this knowledge can assist maintenance technicians in identifying the root-cause of a similar failure when it occurs for the first time in a plant located in a different part of the world using a different language. By harnessing the power of language models, valuable insights and experiences can be shared across geographical boundaries, facilitating efficient troubleshooting and problem-solving on a global scale.

This paper's purpose is to offer a thorough account of our research methods and findings. Through the analysis of an industrial case study, the paper will delve into the requirements and possibilities of mitigating recurring failures by analysis of free-text fields.

2. Methodology

2.1. Case Company

The production plant of the case company accommodates an approximate workforce of 1,100 employees, engaging in diverse manufacturing processes encompassing machining, curing, assembling, testing, and painting of driveline components tailored for the heavy automotive industry. The plant features approximately 300 manufacturing machines, a dedicated heat treatment facility, an array of assembly equipment including presses, torch wrenches, and manually guided vehicles, alongside test benches and a paint shop process. In the Computerized Maintenance Management System (CMMS) there were a little over 3,200 pieces of equipment registered in the production plant in 2023.

Within the case company, about 100 employees are actively involved in maintenance. Maintenance repairmen play a pivotal role in executing hands-on tasks, which encompass corrective and preventive maintenance, including activities such as condition monitoring and improvement work. Additional roles encompass maintenance engineers, developers, procurers, and storage personnel, each contributing to the overarching maintenance framework.

The case company has relied on the same CMMS since 1999, but system upgrades have been implemented. In Sweden, three other production plats belonging to the same corporation use the same system. Within a few years though it is planned that all production plants in the corporation, globally (more than 10 plants), will change to a standardized system. Today, at the specific case company all maintenance work orders are logged within the CMMS, encompassing crucial information such as work order requests (often initiated by operators in free-text format), work order types, chronological timestamps, spare parts consumption, maintenance costs, and work order reports (submitted in free-text format by repairmen).

Perceived by employees at the company and to a certain extent analyzed [9], the case company is suffering from recurring failures. To identify these recurring failures, maintenance engineers engage in a manual analysis of work orders within the CMMS. This analysis is conducted on a machine-by-machine basis, focusing particularly on the free-text fields within the work order request and subsequent work order reports. As stated by maintenance engineers at the reference company, this process is notably time-intensive and presents challenges in cross-referencing patterns of failures across different machines.

2.2. Data gathering and initial analysis

The case company provided data for this study, which was mixed with both text and numbers. The data consists of extractions from the company CMMS. Even though much more data has been shared (that can be used for validation), one production cell, consisting of five machining centers, that are numbered 02-XXX-61 to 02-XXX-65, have been used in this study, particularly the data dealing with failures. The downloaded data contains all data from when the machines were taken into commission in 2007/2008 until the fall of 2022. In total, the machining cell has had more than 6,000 maintenance work orders reported in the CMMS. Specifically, for the five machining centers, roughly 1,500 work orders are related to failures. A failure (breakdown) is defined as when an operator or assembler cannot run their equipment for whatever reason and must contact the maintenance department to receive support. These reasons can be in the form of availability, safety, quality, and environmental problems.

The information was delivered in CSV format. Excel and MS Power BI were used for the analysis. We used EDA (Exploratory Data Analysis) using Python to perform further analysis of the data to learn more about its patterns, trends, outliers, errors, and inconsistencies, such as missing values and duplicate records. Effective data visualizations, such as histograms, scatter plots, and box plots, helped us make the data simpler to grasp and analyze. Additionally, through interviews, routine meetings with Subject Matter Experts (SMEs), and factory visits, more data and information were obtained.

In addition to the statistical analysis, manual analysis of parts of the data were performed by the SMEs to visualize the potential of free-text field analysis. Also, the SMEs performed a wider data gathering and analysis of CMMS data to validate that recurring failures are a problem at the case company. This was performed by downloading failure data of all equipment at the case company during the window of the first six months of 2023. In order to exemplify how much the worst performing machine equipment affects the overall data and production it was decided to compare the overall results by the top 25 worst performing machine equipment from a number of failure perspective.

2.3. Ontology development

The proposed ontology has been created through a basic data model as described and illustrated in Figure 1 [16].

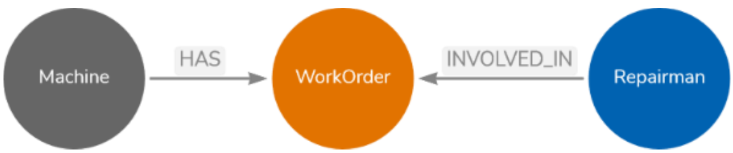


Figure 1. The Basic Data model.

Identification of specific instances of Named Entity Recognition (NER) instances necessitated close collaboration with SMEs. These examples served as training data for the NER model, subsequently enabling the testing of new problem descriptions to identify additional NER entities. A Custom NER solution was created because of this cooperative effort. NER plays a crucial role in Natural Language Processing (NLP) and finds widespread applications in tasks such as information retrieval. It is essential to emphasize that NER is a supervised learning task, relying on meticulous annotation and labeled data.

Custom NER is very crucial for developing domain-specific ontology. The methodology used to develop ontology from Custom NER was *Knowledge Meta Process (KMP): Methodology for Ontology-based Knowledge Management*. The process is a structured approach to developing an ontology. In the beginning, the *kickoff* stage where requirements have been gathered through various methods such as document review, Exploratory Data Analysis (EDA), workshops with SMEs, custom Named Entity Recognition (NER), and factory visits. Which was described in the data gathering stage. In the next stage, the *refinement* stage, the ontology is iteratively defined using the Protégé3 ontology editor. Here, classes, relationships, and instances are added to the ontology, refining it further. The *evaluation* stage follows, where authors checked the quality and consistency of the ontology using The OntOlogy Pitfall Scanner (OOPS!). This was used to evaluate the ontology, identifying any potential issues that need to be addressed. Finally, in the *maintenance* stage, the ontology was published and updated as new knowledge emerged. This ensures that the ontology remains relevant and accurate over time.

During the development of ontology stages, authors worked closely with the SMEs. This approach ensured that the final product was accurate, complete, and useful for its intended purpose. The developed ontology can only be used for AI-based decision systems in manufacturing industries. The method used in the study can be applied to similar cases and may be helpful for future ontology development projects. The entire process is shown in Figure 2.

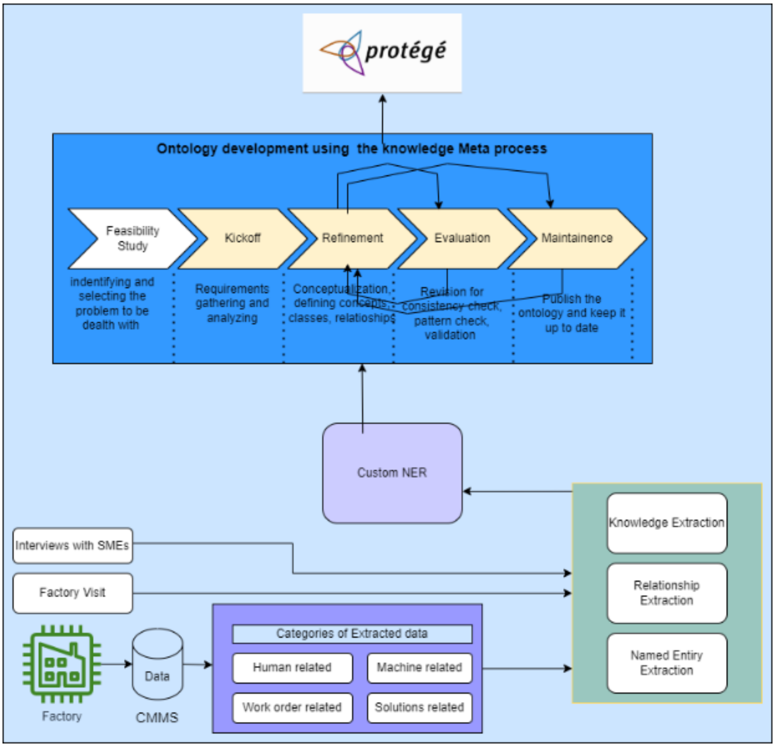


Figure 2. The total process of ontology development.

In Figure 3, the authors used some unstructured data like "*Portalen I maskin läcker så mycket.//XXX-76"* mapped it on the developed ontology.

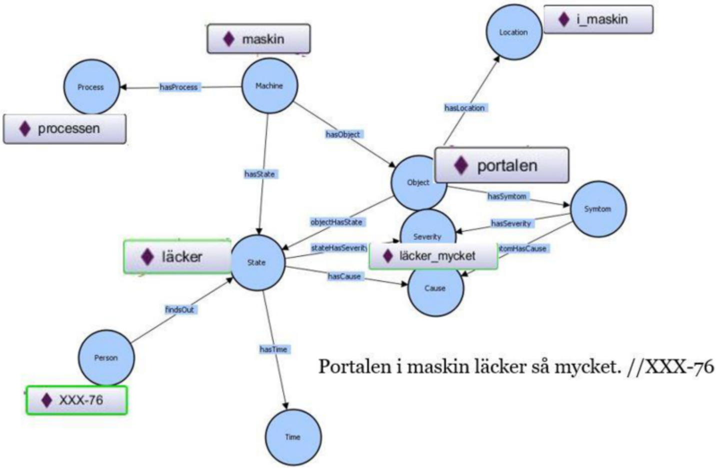


Figure 3. Mapping the data into the Ontology.

3. Potential of free-text field analysis

3.1. Analysis of top 25 worst performing machines

Failure data from the first six months of 2023 was downloaded. It shows that 608 pieces of equipment have suffered failures and that a total of 3,632 failures have occurred. When analyzing the top 25 worst performing machine equipment from a number of failure perspective it shows that they have suffered from 908 failures. That is, the top 25 worst performing machines account for 25% of all failures. Also, the average number of failures for these 25 machines equals 36.32. Considering this is data from six months, it is safe to say that some of these failures are recurring, that the root-cause is not solved on first attempt. Further, if analyzing the timestamps for the failures as well as the total downtime it gives more indications that these 25 machines affect the overall data and production. The top 25 worst machines from a number of failure perspective account for almost 22% of the total timestamps of failures. Lastly, the top 25 worst machines from a number of failure perspective accounts for a little more than 18% of the total downtime caused by failures; see Table 1 for summation.

Table 1. Visualization of the effect of the top 25 worst machines from a failure perspective.

	Failures	Worked hours	Open work order time
Total	3 632	19 160	63 165
Top 25 worst	908	4 160	11 628
%	25.0%	21.7%	18.4%

3.2. Manual analysis of recurring failures and the potential to reduce these

In the manual analysis, we examined the same five machining centers that were also used in developing the digitized solution. These five machining centers have similar design and were all commissioned in 2007/2008. They belong to the same production cell, are all controlled by a storage crane, and manufacture the same type of case material, producing related articles, specifically axles for heavy-duty vehicles.

For the manual analysis, we analyzed failure-related data between the summer of 2021 and 2022. The data includes failure reports (made by operators), technician reports, mean downtime, and the number of worked hours, as well as timestamps related to: when the failure occurred, when the work order was created, and when the work order was closed. 175 failures were recorded for all five machines during this period.

We analyzed the failure reports written by operators, and where similarities were found, the failures were classified as recurring failures. We identified ten unique recurring failures, some of which had occurred on only one individual machine, while others had occurred on several or on all five machines.

3.2.1. Example of a failure report

As mentioned in the introduction, two potential approaches exist for enhancing the analysis of failure data. The first approach entails reducing or circumventing the utilization of free-text entries in failure reports, while the second approach focuses on adopting more sophisticated and advanced methods to process, interpret, and comprehending free-text data. Each approach comes with its own set of benefits, advantages, and challenges. While both approaches should be followed-up and further investigated and refined, this paper will primarily emphasize the second approach within its specific context.

In Figure 4, the results of one of the recurring failures found on the machining centers are visualized. In this example there is a problem with the chip conveyor (in Swedish "spåntransportör"), and it has within the stipulated time frame occurred on four of the machines of a total of 10 instances. If one instead looked at the failure source logged by the technician and performed a fixed drop-down analysis, no less than six unique categories have been used (including uncategorized). This visualizes that relying on dropdown lists is not always effective when analyzing what is a recurring failure and what is not and that both analyses should be carried out.

WO no.	Machine no.	Failure description	Failure source	Reported date	WO created dat
348927	02-XXX63	Får ej signal att spånskyddsdörren är stängd		2022-08-13 06:48	2022-08-13 07
347885	02-XXX65	såpnskyddsdörr går ej ner	Rörliga skydd	2022-06-27 15:48	2022-06-28 07
344893	02-XXX65	Går ej att stänga spånskyddsdörren och porten till maskinen.	Rörliga skydd	2022-04-22 07:48	2022-04-22 08
342312	02-XXX64	larmar för säkrhet relä löst ut på spåntransportör.	Spåntransportörer	2022-02-15 21:08	2022-02-16 06
339689	02-XXX62	spånskyddsdörr vid vertygsväxling går trögt larmar vid varje vertygs byt	Rörliga skydd	2022-01-16 10:52	2022-01-16 11
337176	02-XXX63	såpnskyddsdörr för verktyg	Skyddsutrustning	2021-11-16 06:20	2021-11-16 06
336762	02-XXX64	Larmar på en givare i spåntransportören	Kylvätskeanläggning	2021-11-03 21:20	2021-11-03 22
336759	02-XXX62	spånskyddsdörren smäller ovh fastnar halvvägs	Rörliga skydd	2021-11-03 16:56	2021-11-03 19
336611	02-XXX64	spånkydören fastnad gick inte öppna even manuel	Elutrustning	2021-10-30 21:20	2021-10-31 16
335317	02-XXX64	Behöver återställa säkerhetsrelä i spånskyddstransportören	Spåntransportörer	2021-09-21 22:04	2021-09-21 22

Figure 4. Example of manual analysis of a recurring failure on the machining centers occurring during twelve months of downloaded data, this view is filtered on this recurring failure. Of ten failures with the same problem no less than 6 unique failure sources had been logged.

3.2.2. Equations to approximate potential of free-text field analysis

To approximate potential savings of reducing these recurring failures, some calculations were performed. The equations were created in MS Excel, see Eqs. (1) to (4). The product is PS_{TOT}, which is the total potential savings from working to reduce recurring failures. Many of the inputs were available in the downloaded data or in the analysis, while other data, such as cost for lost production time and spare parts, and efficiency of AI model and maintenance actions were approximated.

- N_{BD} No. of failures
- N_{RBD} No. of recurring failures
- N_{PRRBD} No. of potential reductions of recurring failures
- WH_{BD} Total working hours on failures (h)
- WH_{RBD} Total working hours on recurring failures (h)
- WH_{PRRBD} Potential reduction of working hours on recurring failures (h)
- MDT_{RBD} Mean Down Time of recurring failures (h)
- $C_{Maint.\ hours}$ Cost of maintenance person hours (SEK)
- $C_{Lost\ prod.\ time}$ Approximated cost of lost production time per hour (SEK)
- $C_{Spare\ parts}$ Approximated cost of spare parts per recurring failure (SEK)
- Eff_{AI} Efficiency of AI model (%)
- $Eff_{Maint.\ actions}$ Efficiency of maintenance actions (%)
- $PS_{Maint.\ hour}$ Potential savings in reduced maintenance person hours (SEK)
- $PS_{Lost\ prod.\ time}$ Potential savings in reduced lost production time (SEK)
- $PS_{Spare\ parts}$ Potential savings in reduced spare part cost (SEK)
- PS_{TOT} Total potential savings (SEK)

$$PS_{Maint.\ hours} = C_{Maint.\ hours} \cdot WH_{PRRBD} \cdot Eff_{AI} \cdot Eff_{Maint.\ actions} \quad (1)$$

$$PS_{Lost\ prod.time} = C_{Lost\ prod.\ time} \cdot MDT_{RBD} \cdot N_{PRRBD} \cdot Eff_{AI} \cdot Eff_{Maint.\ actions} \quad (2)$$

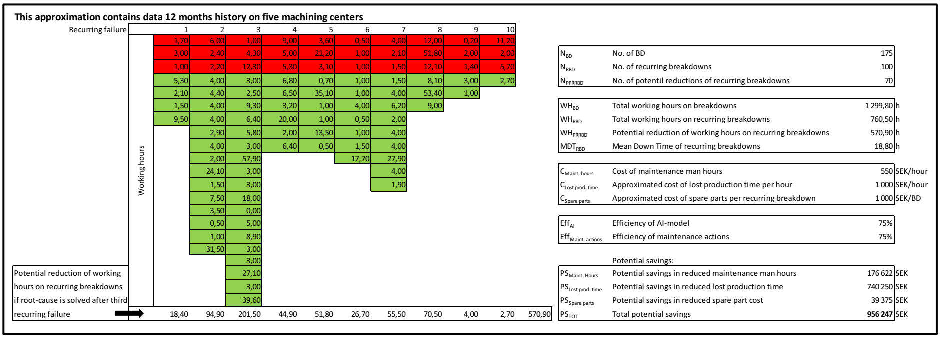
$$PS_{Spare\ parts} = C_{Spare\ parts} \cdot N_{PRRBD} \cdot Eff_{AI} \cdot Eff_{Maint.\ actions} \quad (3)$$

$$PS_{TOT} = PS_{Maint.\ hours} + PS_{Lost\ prod.\ time} + PS_{Spare\ parts} \quad (4)$$

3.2.3. Results of manual analysis

Based on a comprehensive analysis, blending both manual analysis and approximated data, the results unequivocally underscore a substantial potential for effectively mitigating recurring failures. If a warning system could alert the maintenance department that a recurring failure is indeed a recurring one after the third time it occurs and if the maintenance department at that time fixes the root-cause of the problem the potential of reduction of failures is in this case 70 for this particular year. If approximating that the internal cost per maintenance hour is 550 SEK, if the cost for lost production is 1,000 SEK/hour and that the spare part cost per recurring failure is 1,000 SEK at every recurring failure then the potential saving would amount to 1.7 MSEK per year for these five machines. However, it is unlikely that neither the AI-system nor the maintenance department would be 100% efficient in all instances. If reducing the efficiency to 75% for both, the potential saving would, as indicated in Figure 5, be around 950,000 SEK the first year.

In this example, certain uncertainties are evident. The analysis is based on just one year of data, and it is plausible that many of these recurrent failures may have persisted for a more extended period. These machines being analyzed have a history of poor availability, if analyzing other, similar, machines, the results, the potential savings that is, might not be as high. There might be a cost to solve and mitigate the root-cause, this has not been approximated in this model. However, the intention of the example is not firsthand to give an exact business case but to approximate potential savings.



This approximation contains data 12 months history on five machining centers

Recurring failure	1	2	3	4	5	6	7	8	9	10	
Working hours	1,70	6,00	1,00	9,00	3,60	0,50	4,00	12,00	0,20	11,20	
	3,00	2,40	4,30	5,00	21,20	1,00	2,10	51,80	2,00	2,00	
	1,00	2,20	12,30	5,30	3,10	1,00	1,50	12,10	1,40	5,70	
	5,30	4,00	3,00	6,80	0,70	1,00	1,50	8,10	3,00	2,70	
	2,10	4,40	2,50	6,50	35,10	1,00	4,00	53,40	1,00		
	1,50	4,00	9,30	3,20	1,00	4,00	6,20	9,00			
	9,50	4,00	6,40	20,00	1,00	0,50	2,00				
		2,90	5,80	2,00	13,50	1,00	4,00				
		4,00	3,00	6,40	0,50	1,50	4,00				
		2,00	57,90			17,70	27,90				
		24,10	3,00				4,00				
		1,50	3,00				1,90				
		7,50	18,00								
		3,50	0,00								
		0,50	5,00								
		1,00	8,90								
		31,50	3,00								
			3,00								
			27,10								
			3,00								
			39,60								
Potential reduction of working hours on recurring breakdowns if root-cause is solved after third recurring failure	18,40	94,90	201,50	44,90	51,80	26,70	55,50	70,50	4,00	2,70	570,90

N_{BD}	No. of BD	175	
N_{RBD}	No. of recurring breakdowns	100	
N_{PPRRBD}	No. of potentil reductions of recurring breakdowns	70	
WH_{BD}	Total working hours on breakdowns	1 299,80	h
WH_{RBD}	Total working hours on recurring breakdowns	760,50	h
WH_{PRRBD}	Potential reduction of working hours on recurring breakdowns	570,90	h
MDT_{RBD}	Mean Down Time of recurring breakdowns	18,80	h
$C_{Maint. hours}$	Cost of maintenance man hours	550	SEK/hour
$C_{Lost prod. time}$	Approximated cost of lost production time per hour	1 000	SEK/hour
$C_{Spare parts}$	Approximated cost of spare parts per recurring breakdown	1 000	SEK/BD
Eff_{AI}	Efficiency of AI-model	75%	
$Eff_{Maint. actions}$	Efficiency of maintenance actions	75%	
	Potential savings:		
$PS_{Maint. Hours}$	Potential savings in reduced maintenance man hours	176 622	SEK
$PS_{Lost prod. time}$	Potential savings in reduced lost production time	740 250	SEK
$PS_{Spare parts}$	Potential savings in reduced spare part cost	39 375	SEK
PS_{TOT}	Total potential savings	**956 247**	SEK

Figure 5. Screenshot of Excel file with data (12 months history) from the CMMS on the manual analysis of recurring failure on five manufacturing machines as well as approximations on cost of lost production time per hour, cost of spare parts per recurring failure as well as efficiency of the AI-model as well as the resulting maintenance actions.

4. System development

One of the tasks in finding previous similar cases of machine failures involves assessing the semantic similarities between cases. To accomplish this, we aim to utilize the Sentence Transformers framework. It is a Python framework that provides advanced sentences, text, and image embeddings. The initial research behind this framework is explained in [17].

With this framework, we have the capability to create embeddings (vector representations) for sentences or text in more than 100 diverse languages. These embeddings can then be compared using cosine similarity to detect sentences that convey similar meanings. This feature proves invaluable for various tasks, including semantic textual similarity assessment, semantic search, and paraphrase discovery.

There are many pre-trained multilingual models in Sentence Transformers but still, we will use the 'paraphrase-multilingual-mpnet-base-v2', as it is suitable for multilingual semantic search tasks [18]. The model is a Large Language Model (LLM) trained to understand and generate language representations across multiple languages. It is specifically designed for tasks involving semantic understanding, such as paraphrase identification and semantic search. While training the model this was specifically trained on the paraphrase identification which means it is designed to understand the semantic similarity between a given query and existing datasets.

This model is selected in this study due to its multilingual capabilities, as it is a variant of SBERT trained on parallel data from over 50 languages. The model is trained using sentence pairs that are translations of each other in various languages. This method allows the model to create embeddings that can be compared across different languages by identifying a shared semantic space for all these languages [17].

Furthermore, this model is contextual, meaning it generates embeddings that consider the context of the words in a sentence. This feature enables the model to manage misspelled words in a sentence and identify similar sentences within the dataset [17, 19]. Figure 6, below, illustrates the architecture of our system.

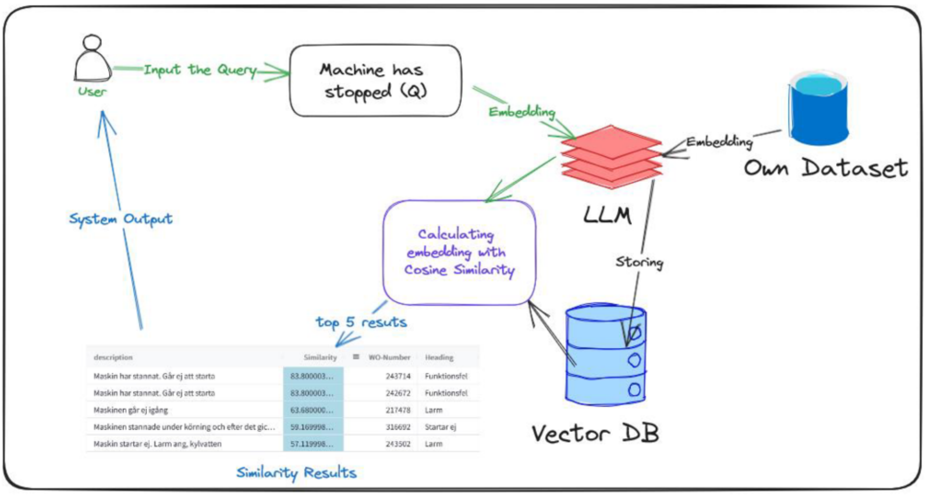
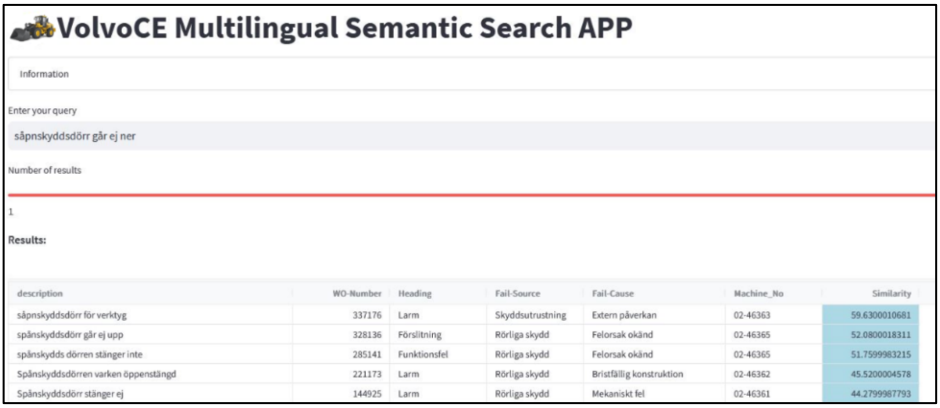


Figure 6. Architecture of the system.

4.1. Early results of the system

In this section, some early results of the system and how it displays the resulting similarities of sentences are shown through screenshots, like searches. The screenshots show that the system can handle misspelled words, see Figure 7 and synonyms, see Figure 8, and show its multilingual capabilities, see Figure 9.



description	WO-Number	Heading	Fail-Source	Fail-Cause	Machine_No	Similarity
såpnskyddsdörr för verktyg	337176	Larm	Skyddsutrustning	Extern påverkan	02-46363	59.6300010681
spånskyddsdörr går ej upp	328136	Förslitning	Rörliga skydd	Felorsak okänd	02-46365	52.0800018311
spånskydds dörren stänger inte	285141	Funktionsfel	Rörliga skydd	Felorsak okänd	02-46365	51.7599983215
Spånskyddsdörren varken öppenstängd	221173	Larm	Rörliga skydd	Bristfällig konstruktion	02-46362	45.5200004578
Spånskyddsdörr stänger ej	144925	Larm	Rörliga skydd	Mekaniskt fel	02-46361	44.2799987793

Figure 7. Screenshot of the system searching the misspelled word "såpnskyddsdörr" which should be spelled "spånskyddsdörr". Even with the misspelled word the system generates hits of similar, recurring, failures with the correct spelling.

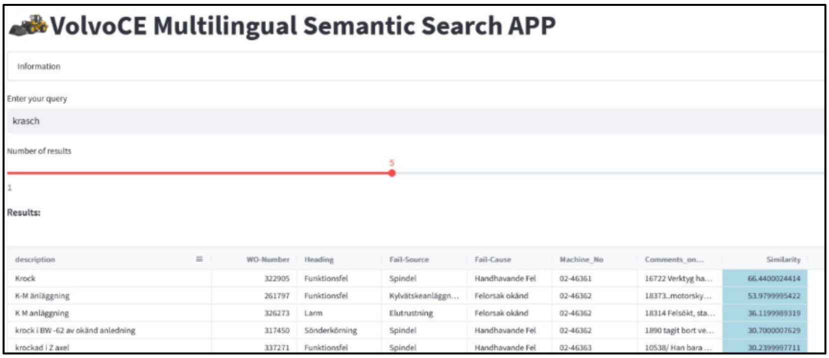


Figure 8. Screenshot of the system searching the word "krash", translated to English "crash" or "collision" generates hits of Swedish word "krock" which is a synonym of "krash", indicating that the system can handle synonyms. The operators and repairmen of the case company rarely use the word "krash" when a machine collision happens, it is more common to use the term "krock".

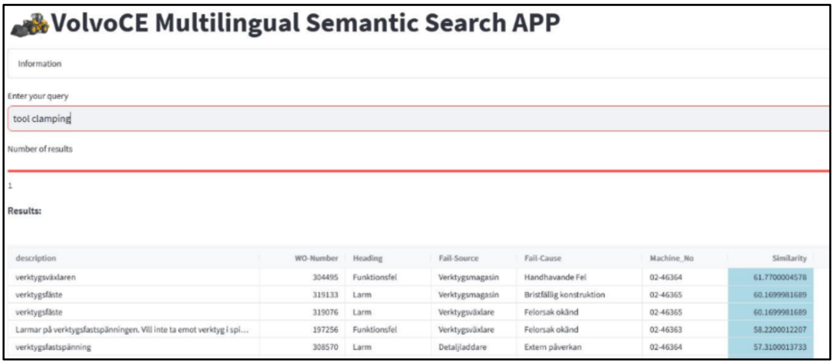


Figure 9. Screenshot of the system visualizing that using the English search words "tool clamping" generates hits of similar, recurring, failures in Swedish.

5. Discussion and conclusions

Analyzing free-text fields through tools found in Artificial Intelligence opens-up for real-time analysis of recurring failures. Through the case study it was visualized that there exists a large potential for both cost savings and reduced environmental impact through early recurring failure detection. As the case company corporation, on a Swedish level, use the same CMMS there could, through such a system be a straightforward way of exchanging and reusing experiences. As all production plants (globally) within the corporation will change to the same CMMS, the possibilities of savings and exchanging and reusing experiences will be even greater. Using LLM removes the language barriers as many CMMS are populated with domestic languages.

In our future research, we aim to further explore other multilingual models. We plan to compare their performance on both benchmark datasets and our proprietary dataset. This comparative analysis will provide valuable insights for this use case in the LLM domain. We would also like to refine the search results by tapping into the failure reporting of the repairmen and which spare parts have been used on previous recurring failures. This would further open the possibility to not only identify recurring failures but to also analyze root-cause in the historical data of CMMSs.

Acknowledgements

This work was supported by the Adapt 2030 project (Adaptive lifecycle design by applying digitalization and AI techniques to production) under Vinnova (Sweden's innovation agency) project grant 2019-05589 within the strategic innovation programme for Production2030 as well as part of the XPRES framework at Mälardalen University.

References

[1] Franciosi C, Voisin A, Miranda S, Riemma S, Iung B. Measuring maintenance impacts on sustainability of manufacturing industries: from a systematic literature review to a framework proposal. Journal of Cleaner Production. 2020;260:121065, doi: 10.1016/j.jclepro.2020.121065

[2] Pires A, Martinho G. Waste hierarchy index for circular economy in waste management. Waste Management. 2019;95:298-305, doi: 10.1016/j.wasman.2019.06.014

[3] Geng Y, Sarkis J, Bleischwitz R. How to build a circular economy for rare-earth elements. Nature. 2023;619:248-251, doi: 10.1038/d41586-023-02153-z

[4] Bengtsson M, Lundström G. On the importance of combining "the new" with "the old" – One important prerequisite for maintenance in Industry 4.0. Procedia Manufacturing. 2018;25:118-125, doi: 10.1016/j.promfg.2018.06.065

[5] Kans M, Campos J, Salonen A, Bengtsson M. The Thinking Industry – An Approach for Gaining Highest Advantage of Digitalisation within Maintenance. Journal of Maintenance Engineering. 2017; 2:147-158.

[6] Rahman H, D'Cruze RS, Ahmed MU, Sohlberg R, Sakao T, Funk P. Artificial Intelligence-Based Life Cycle Engineering in Industrial Production: A Systematic Literature Review. IEEE Access. 2022;10:133001-133015, doi: 10.1109/ACCESS.2022.3230637

[7] Gopalakrishnan M. Data-driven decision support for maintenance prioritisation – Connecting maintenance to productivity. PhD diss., Chalmers University of Technology Sweden: Department of Industrial and Materials Science, 2018.

[8] Chirumalla K, Bengtsson M, Söderlund C. Experience Reuse in Production Maintenance: Practices and Challenges. Proceedings of 22nd European Operation Management Association Conference, EurOMA. 2015, doi: 10.13140/RG.2.1.4939.7924

[9] Salonen A, Bengtsson M, Fridholm V. The possibilities of improving maintenance through CMMS data analysis. In: Säfsten K, Elgh, F, editors. Proceedings of SPS2020. IOS Press; 2020. p. 249-260, doi: 10.3233/ATDE200163

[10] Duffuaa SO, Raouf A. (2015). Planning and Control of Maintenance Systems, Springer Cham, 2015. 348 p, doi: 10.1007/978-3-319-19803-3

[11] Labib A. A decision analysis model for maintenance policy selection using a CMMS. Journal of Quality in Maintenance Engineering. 2004;10(3):191–202, doi: 10.1108/13552510410553244

[12] Rausand, M, Øien K. (1996). The basic concepts of failure analysis. Reliability Engineering and System Safety. 1996;53(1):73-8, doi: 10.1016/0951-8320(96)00010-5

[13] Stenström C, Al-Jumaili M, Parida A. Natural language processing of maintenance records data. International Journal of COMADEM. 2015;18(2):33-37.

[14] Ahmed MU, Bengtsson M, Salonen A, Funk P. Analysis of Breakdown Reports Using Natural Language Processing and Machine Learning. In: Karim R, Ahmadi A, Soleimanmeigouni I, Kour R, Rao R, editors. Proceedings of International Congress and Workshop on Industrial AI 2021; 2021 Lecture Notes in Mechanical Engineering, Springer, Cham. p. 40-52. doi: 10.1007/978-3-030-93639-6_4

[15] Edvinsson L, Sullivan P. Developing a model for managing intellectual capital. European Management Journal. 1996;14(4):356-364, doi: 10.1016/0263-2373(96)00022-9

[16] D'Cruze RS, Ahmed MU, Bengtsson M, Ur Rehman A, Funk P, Sohlberg R. A Case Study on Ontology Development for AI Based Decision Systems in Industry. In: Kumar U, Karim R, Galar D, Kour R. editors. International Congress and Workshop on Industrial AI and eMaintenance 2023. IAI 2023. Lecture Notes in Mechanical Engineering. Springer, Cham, doi: 10.1007/978-3-031-39619-9_51

[17] Reimers, N, Gurevych, I. (2019). Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks. arXiv:1908:10084, doi: 10.48550/arxiv.1908.10084

[18] Sentence Transformers Documentation [Internet]. [cited 2023 Sept 15]; Available from: https://www.sbert.net/index.html

[19] SBERT GitHub page [Internet]. [cited 2023 Sept 15]; Available from: https://github.com/UKPLab/sentence-transformers