

Correcting Maintenance Reports Using Large Language Models

Christian Salas

Mechanical Engineering Department, Pontificia Universidad Católica de Valparaíso, Chile. E-mail: christian.salas.c@mail.pucv

Dario Valcamonico

Energy Department, Politecnico di Milano, Italy. E-mail: dario.valcamonico@polimi.it

Piero Baraldi

Energy Department, Politecnico di Milano, Italy. E-mail: piero.baraldi@polimi.it

Orlando Durán

Mechanical Engineering Department, Pontificia Universidad Católica de Valparaíso, Chile. E-mail: orlando.duran@pucv.cl

Gustavo Mansilla

Mechanical Engineering Department, Pontificia Universidad Católica de Valparaíso, Chile. E-mail: gustavo.mansilla.m@mail.pucv.cl

Enrico Zio

*MINES Paris-PSL, Centre de Recherche sur les Risques et les Crises (CRC), France
Energy Department, Politecnico di Milano, Italy. E-mail: enrico.zio@polimi.it*

Abstract. Maintenance reports are a rich source of information for reliability engineering and asset management. They typically contain unstructured text and categorical information, which can be used to estimate quantities of interest for maintenance decisions, such as Mean Time Between Failures (MTBF) and Mean Time to Repair (MTTR) of system components. However, the extraction of knowledge from these reports is challenged by linguistic variability, technical jargon and frequent human errors committed when filling the categorical fields. To address these issues, the present work proposes a novel, two-step method based on Large Language Models (LLMs) to automatically identify the units, subunits and components involved in maintenance interventions. First, an LLM-based model (LLM_1) identifies possible quality issues in the unstructured text, flagging reports that suffer from brevity, lack of context, excessive use of technical terms, internal incoherence and inconsistencies with the categorical fields. Then, based on an equipment taxonomy, a second LLM-based model (LLM_2) systematically identifies the units, subunits and components involved in the maintenance interventions from reports of low quality. The overall method is validated considering a real-world case study of maintenance reports from a mining haul truck fleet. The obtained results demonstrate that: *i*) LLM_1 successfully identifies quality issues in a significant fraction (36%) of the reports; *ii*) LLM_2 effectively corrects erroneous categorical assignments of the units, subunits and components involved in the maintenance intervention in low quality reports.

Keywords: Reliability engineering, Maintenance, Knowledge extraction, Safety assessment, Decision support, LLM.

1. Introduction

Maintenance reports typically contain short, unstructured text written by operators, along with categorical fields assigned by maintenance

engineers, which identify the units, subunits and components involved in the maintenance intervention [1].

This rich source of information on observed failures, malfunctions and maintenance actions can support reliability analyses and maintenance decision-making [2].

The extraction of the categorical fields from the reports is challenged by [3]:

- (i) The length of the reports, which can vary significantly across the repository. Shorter reports tend to be less informative due to the lack of contextual information;
- (ii) The use of technical language (e.g., abbreviations, codes), which requires extensive expert knowledge;
- (iii) The large number of reports to be analyzed.

Text annotation in industrial applications has traditionally relied on Natural Language Processing (NLP) methods that transform reports into numerical features, including frequency-based vector-space models such as Term Frequency-Inverse Document Frequency (TFIDF) [4]. These methods are useful for identifying relevant terms, but they are limited by their dependence on word occurrence or fixed representations, which can neglect word order, contextual variability and long-range semantic relationships [5]. Transformer-based methods such as Large Language Models (LLMs) have been introduced to address these limitations through self-attention and context-dependent representations, which have been shown to significantly improve the identification and extraction of information from textual data [6].

In this context, this work proposes a method based on LLMs to automatically identify from maintenance reports the units, subunits and components involved in the maintenance intervention. Specifically, the method is composed of:

- (i) An LLM (LLM_1), which identifies possible quality issues in the unstructured text, considering brevity, lack of context, excessive

use of technical terms, internal incoherence and inconsistencies with the categorical fields assigned by the maintenance engineers;

- (ii) A second LLM (LLM_2), which identifies the units, subunits and components involved in the maintenance interventions from the unstructured text of the reports. LLM_2 exploits a taxonomy defined considering a normative (e.g., ISO 14224 [7]). It is applied to low quality reports to correct possible errors in the categorical fields.

The developed method is applied to a case study of maintenance reports from a haul truck fleet.

The remainder of this work is organized as follows: Section 2 details the developed method; Section 3 introduces the case study; Section 4 evaluates the performance of the proposed method and discusses the obtained results; Section 5 outlines the conclusions of the work.

2. Developed method

To identify the units, subunits and components involved in the maintenance interventions, this work develops a method based on the two LLMs, whose architecture is illustrated in Figure 1.

2.1. Dataset

We consider a historical repository R composed of N maintenance reports, $R = \{r_1, \dots, r_i, \dots, r_N\}$. A generic report $r_i \in R$ consists of a short unstructured text t_i , written by a system operator, describing the maintenance intervention and the set of the categorical fields T_i , assigned by the maintenance engineer, containing the unit, subunit and component involved in the maintenance intervention.

Also, a predefined taxonomy O of units, subunits and components defined by maintenance engineers according to the ISO 14224 industrial standard [7] is available.

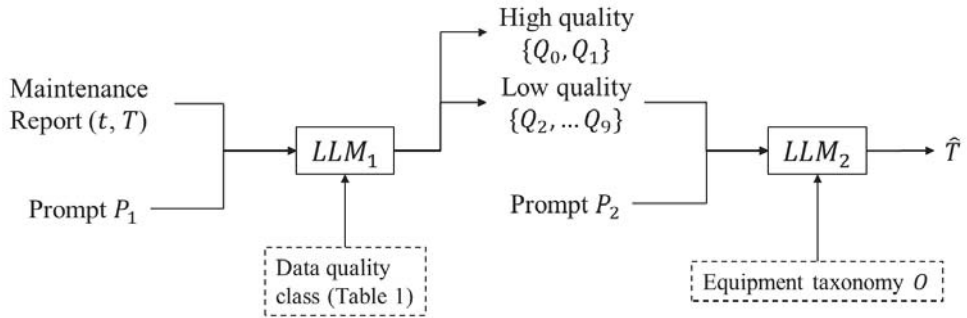


Fig. 1. Proposed method.

2.2. Identification of quality issues

A first LLM-based method, LLM_1 , is developed to evaluate the quality of the reports and identify the causes of low quality. Specifically, LLM_1 receives in input the unstructured text t and the set of categorical fields T and generates in output a label $q \in \{Q_0, \dots, Q_9\}$ indicating the quality of the report r . High quality reports are assigned to class Q_0 and Q_1 , whereas the low quality reports are assigned to

classes Q_2, \dots, Q_9 according to the type of quality issue (Table 1).

LLM_1 is queried by the prompt P_1 (reported in Figure 2), which contains a step-by-step description of the task of comparing the unstructured text t with the set of categorical fields T .

Table 1. Report quality classes.

Quality class	Report quality	Description
Q_0	High	The unstructured text t clearly mentions the categorical fields of T .
Q_1	High	The categorical fields of T identified by the maintenance engineers are the only logical conclusion from the specific maintenance intervention described in t .
Q_2	Low	At least a categorical field of T is missing.
Q_3	Low	The right component is identified but in the wrong location.
Q_4	Low	Despite the categorical fields refer only to one component, more than one component was involved in the maintenance intervention.
Q_5	Low	The unstructured text t contains abbreviations, slangs and syntactic mistakes.
Q_6	Low	At least a categorical field of T contains a term not present in the taxonomy.
Q_7	Low	Despite the unstructured text t containing very limited information, the maintenance engineer guessed one or more categorical fields of T .
Q_8	Low	The categorical fields of T refer to a different system than the one involved in the maintenance intervention described in the unstructured text t .
Q_9	Low	Despite the unstructured text t mentioning that a component did not have a problem, that component is indicated in T .

```

### instructions
Role: Maintenance Data Auditor.
Task: Identify possible quality issues considering the maintenance text and the assigned taxonomy.
### codes_for_matching
Assign the unstructured text of the report to a class of quality considering semantic inconsistencies and labeling errors.
Choose among the following classes:
Q_00: if the unstructured text clearly mentions the unit, subunit and component involved in the maintenance intervention;
Q_01: if the unit, subunit and component identified by the maintenance engineers is the only logical conclusion from the
specific maintenance intervention described in the unstructured text;
Q_02: if at least on categorical field is missing;
Q_03: if The right component is identified but in the wrong location;
Q_04: if despite the categorical fields refer only to one component, more than one component was involved in the maintenance
intervention;
Q_05: if the unstructured text contains abbreviations, slangs and syntactic mistakes
Q_06: if at least on categorical field is not present in the taxonomy
Q_07: if despite the unstructured text containing very limited information, the maintenance engineer guessed one or more
categorical fields;
Q_08: if the categorical fields refer to a different system than the one involved in the maintenance intervention described
in the unstructured text;
Q_09: if despite the unstructured text mentioning that a component did not have a problem, that component is indicated in
the categorical fields
### evaluation_pipeline
Use the following step-by-step sequence to evaluate each report consistently: 1. Report Analysis, 2. Hierarchy Comparison,
3. Conflict Detection, 4. Code Assignment.
### examples
Some examples are provided to improve the accuracy of the identification
Example_1:...

```

Fig. 2. Prompt P_1 for LLM_1 .

```

### instructions
Role: Expert Reliability Engineer in Mining Operations.
Task: Identify the unit, subunit and components categorical fields from the unstructured text of the maintenance reports
using the system specific taxonomy provided.
### context
Equipment_taxonomy.xlsx
### identification_rules
The process must follow these steps:
i) Segmented Evidence: Separately identify the symptom or system, failure mode, corrective action, and any spare parts
mentioned.
ii) Taxonomic Candidates: List the most logical categorical options based on the evidence found in the text.
iii) Selection Criteria: Provide a brief explanation of why that specific element of the taxonomy was selected over others.
### examples
Some examples are provided to improve the accuracy of the identification
Example_1:...

```

Fig. 3. Prompt P_2 for LLM_2 .

2.3. Identification of the categorical fields

A second LLM-based method, LLM_2 , is developed to identify the categorical fields for those reports with low quality issues (classes Q_2, \dots, Q_9). Specifically, LLM_2 receives in input the unstructured text t and the taxonomy O , and automatically generates the set of categorical fields \hat{T} of the unit, subunit and component involved in the maintenance intervention.

LLM_2 is queried by the prompt P_2 (reported in Figure 3).

3. Case Study

The proposed method is applied to repository of over 10,000 maintenance reports from a fleet of haul trucks. The reports are written in Spanish and

characterized by high semantic variability, heavy use of technical jargon and noisy unstructured text.

To establish a baseline for validating the proposed method, a representative subset of 25 reports was selected and manually analysed by an expert. The units, subunits and components identified by this expert serve as the ground truth, denoted as T^{true} .

The processing workflow utilizes the *Qwen3-235B-A22B* model API, which is publicly available in the Hugging Face library [8]. This model employs a Mixture-of-Experts (MoE) architecture, which provides high representational capacity, while maintaining controlled computational costs during inference.

4. Results

LLM_1 found that among the subset of 25 reports analysed by the expert, the nine reports detailed in Table 2 have quality issues:

- Three reports (IDs 8, 10 and 14) are assigned to class Q_4 , which indicates that despite the categorical fields referring only to one component, more than one component was involved in the event;
- Two reports (IDs 18 and 21) are assigned to class Q_2 , which indicates that one or more categorical fields are missing;
- Two reports (IDs 13 and 16) are assigned to class Q_3 , which indicates that the right

component type has been identified but in the wrong location of the system. For example, report 13 describes a front suspension issue, but the component label indicates the rear suspension;

- Two reports (IDs 5 and 15) are assigned to class Q_8 , which indicates that the categorical fields refer to a different system than the one involved in the maintenance intervention described in the unstructured text.

These quality issues identified by LLM_1 have all been confirmed by an expert.

Table 2. Examples of data quality anomalies and LLM_1 evaluations.

ID	Quality class	Unstructured text (<i>t</i>)	Categorical fields (<i>T</i>)		
			Unit	Subunit	Component
5	Q_8	Low Power. unit checked for low power, ran stall test only getting 10 psi boost both engines, checked boost pressure sensor harness on rear only, prev shift had worked front. Found rear boost sensor harness chafed, cable damaged, replaced sensor. Re-ran stall test, got 20 psi rear / 19 psi front. Unit operational.	Engine	Engine_electrical	Engine_oil_pressure_sensor
8	Q_4	ENGINE FILL COUPLING ENGINE / FUEL FILTER REAR ENGINE CLOGGED. M2.- Topped off coupling, replaced rear engine fuel filters.	Engine	Engine_fuel	Secondary_fuel_filter
10	Q_4	Low voltage. Alternator regulator box of 24 volt is changed and hydraulic sistem is refilled. Operative.	Engine	Engine_electrical	Regulator_box
13	Q_3	Knock on front left suspention. Equipment has knock and noise on front left suspension, tested on field and is effective, waiting transfer to workshop to perform procedure.	Machine	Suspension	Rear_left_suspension
14	Q_4	Hydraulic oil leak Fuel leak is eliminated (upper part fuel tank) due to broken fitting (return hose) product of impact by loose pad. Note: hopper missing 1 pad... Equipment remains operative.	Engine	Engine_fuel	Fuel_hose
15	Q_8	AIR LEAK AIR LEAK BY HOSE CABIN BLOWER REPAIR IS PERFORMED	Machine	Engine_starting_pneumatic	Pneumatic_hoses
16	Q_3	AIR CONDITIONING BOX Equipment is transfered to workshop for welding repair due to detachment of air conditioning box on upper part of cabin .	Machine	Cabin_AC	Cabin_fan_motor
18	Q_2	ELECTRICAL ACTIVE ALARM PRESSURE SENSOR ACCUMULATOR LOW	Powertrain	Transmission	
21	Q_2	Reported failure : Smartcap display off. Detected failure: Voltage at converter input cut. Repair : It is decided to change the converter, since they are sulfated. Electrical connections are retightened. System operativity is verified. Validated with dispatch. Final state operative.	Technology	Smartcap	

LLM_2 is evaluated comparing its assignment \hat{T} and the assignment T done by the maintenance engineer with the ground truth T^{true} . The results are visualized using a matrix, shown in Figure 4, which classifies each report as correctly or incorrectly identified compared to its ground truth. Although LLM_2 is primarily intended for reports with quality issues (classes Q_2, \dots, Q_9), it is here also applied to high quality reports (classes Q_0 and Q_1) to assess its ability to correctly identify the involved unit, sub-unit and components.

It may be observed that: *i*) in 18 out of 25 cases (72%), both the maintenance engineer and LLM_2 have correctly identified the ground truth; *ii*) in 6 out of 25 cases (24%), an error of the maintenance engineer in T_i is correctly identified by LLM_2 (T is corrected by \hat{T}); *iii*) in only 1 out of 25 cases (4%), LLM_2 has incorrectly identified a categorical field of T that the analyst had correctly identified.

		LLM_2 identification	
		Incorrect \hat{T}	Correct \hat{T}
Maintenance engineer identification	Correct T	0 (0%)	6 (24%)
	Incorrect T	1 (4%)	18 (72%)

Fig.4. Matrix for the evaluation of LLM_2 .

5. Conclusions

This paper presented a method for extracting knowledge from maintenance reports using Large Language Models (LLMs). The method allows automatically evaluating the quality of the reports and extracting the information related to system, subsystem and components involved in the maintenance intervention, which can be useful for conducting system reliability analysis. The effectiveness of the method has been demonstrated on a sample of 25 reports from a mining truck fleet.

Future research work will explore the use of LLMs to identify component failures and malfunctions and suggest corrective actions.

References

- [1] Valcamonico, D., et al. (2025). "A systematic procedure for the analysis of maintenance reports based on a taxonomy and BERT attention mechanism". Published in *Reliability Engineering and System Safety*, 256. <https://doi.org/10.1016/j.res.2025.110834>
- [2] Stenström, C., Aljumaili, M., & Parida, A. (2015). "Natural Language Processing of Maintenance Records Data". *International Journal of COMADEM*, 18(2), 33-37
- [3] Campos J, Sharma P, Gabiria UG, Jantunen E, Baglee D. A big data analytical architecture for the asset management. In: Proceedings of the 9th CIRP IPSS conference: circular perspective on product/service-systems. 64; 2017. p. 369–74. <https://doi.org/10.1016/j.procir.2017.03.019>
- [4] G. Amati and C. J. Van Rijsbergen, "Probabilistic Models of Information Retrieval Based on Measuring the Divergence from Randomness," *ACM Transactions on Information Systems*, vol. 20, no. October, pp. 357–389, 2002.
- [5] T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Efficient Estimation of Word Representations in Vector Space," in *International Conference on Learning Representations (ICLR)*, 2013. doi: 10.48550/arXiv.1301.3781.
- [6] Q. Liu, Y. Ding, Luo X., "Automated knowledge graph-based risk assessment for fall-from-height accidents using large language models." *Automation in Construction*, vol. 179, 2025. <https://doi.org/10.1016/j.autcon.2025.106482>
- [7] International Organization for Standardization. (2016). Petroleum, petrochemical and natural gas industries — Collection and exchange of reliability and maintenance data for equipment (ISO Standard No. 14224:2016).
- [8] Wolf, T., et al. (2020). Transformers: State-of-the-art natural language processing. In Q. Liu & D. Schlangen (Eds.), Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations (pp. 38–45). Association for Computational Linguistics. <https://doi.org/10.18653/v1/2020.emnlp-demos.6>