Proceedings of the 35th European Safety and Reliability & the 33rd Society for Risk Analysis Europe Conference
Edited by Eirik Bjorheim Abrahamsen, Terje Aven, Frederic Bouder, Roger Flage, Marja Ylönen

doi: 10.3850/978-981-94-3281-3_ESREL-SRA-E2025-P4491-cd

Extracting Reliability and Maintenance Knowledge from Maintenance Reports of Freight Transport Trains: a Methodology for Annotation based on Ontology and SpERT

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

Enrico Zio

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

Abstract: We consider the problem of extracting information from repositories of maintenance reports of freight transport trains, aiming to identify factors influencing malfunctions and failures, and assess the effectiveness of maintenance activities. We propose a methodology for automatically annotating maintenance reports, which involves assigning semantic labels to the words of the reports and identifying the relations between them. The conciseness of the texts and the extensive use of technical language pose significant challenges, which are overcome by combining an industrial maintenance ontology with the Span-based Entity and Relation Transformer (SpERT) method. Specifically, SpERT is fine-tuned in two stages: initially on a large dataset of maintenance reports from other industrial sectors, and, then, on a limited number of manually annotated maintenance reports of electrical freight transport trains. The obtained results show that the proposed methodology successfully identifies entities and relations in maintenance reports of freight transport trains.

Keywords: Maintenance Reports, Ontology, Annotation, Entity and Relation Extraction, Natural Language Processing

1. Introduction

Maintenance reports typically combine structured data, such as multiple-choice fields to record the involved components, the event severity and the intervention date, with unstructured Maintenance Short Texts (MSTs) written by the operators and containing relevant information regarding the malfunction or failure occurred and the maintenance activities performed (Bikaun et al. 2024).

Recently, there has been growing interest in developing methods to automatically extract information from these MSTs, aiming to identify factors influencing the occurrence and severity of malfunctions and failures, and assessing the effectiveness of maintenance activities.

A first task to extract information from MSTs is annotation. This involves classifying the words of the texts in predefined classes (entities) (e.g., type of components and maintenance actions) and identifying the relations between them (Bikaun et al. 2024). A major challenge in the annotation process lies in the inherent ambiguity and context-dependency of language, which complicates labelling words and recognizing relations among them. This is particularly pronounced in the context of industrial maintenance due to:

- the conciseness of the MSTs, which may hinder accurate annotation, as the limited information may lead to ambiguous interpretations (Conte et al. 2021);

- the complexity of the language, which uses technical words, domain-related acronyms, abbreviations and codes (Brundage et al. 2021).

Consequently, the same word or phrase can have different semantic meaning depending on its context, leading to inconsistencies in annotation. For example, the word *valve* can be associated to different entities based on the contextual information: in the text "*emergency safety valve*" it is best classified as a protective object, whereas in the text "*gate valve*" it is best classified as a controlling object.

This work addresses the challenge of automating the annotation of maintenance reports of freight transport trains by combining an expert-based ontology and an annotation method. An ontology is a formal representation of knowledge relevant to a particular domain or task. It is typically structured using hierarchical levels of entities and relations among them. Specifically, the ontology developed in (Bikaun et al. 2024) for the mineral processing and infrastructure industries is adapted to the context of freight transport trains. Then, the annotation task is performed using the Span-based Entity and Relation Transformer (SpERT) method (Eberts and Ulges 2020). SpERT is an NLP method that allows jointly classifying words or spans (contiguous sequences of adjacent words) in entities and identifying the relations between them. In this work, SpERT is fine-tuned in two stages: initially on a large dataset of labelled MSTs from the mineral processing and infrastructure industries, taken from (Bikaun et al. 2024), and, subsequentially, on a limited number of manually annotated maintenance reports of freight transport trains. This double fine-tuning process allows learning domain-specific and case-specific semantics, improving ability of SpERT to accurately classify entities and relations in new maintenance reports.

The remainder of this work is organized as follows: Section 2 states and formulates the problem; Section 3 illustrates the developed ontology; Section 4 describes the annotation method based on the SpERT method; Section 5 presents the case study. Section 6 reports the results obtained; Section 7 discusses the work conclusions.

2. Problem statement and formulation

We consider a repository of D maintenance intervention reports $\{d_i, i = 1, \dots, D\}$ collected from a fleet of freight transport trains.

Each report d_i, $i = 1, \dots, D$, is a free-text description of the accident or malfunction that occurred, and the maintenance intervention performed. A generic report d_i is composed of P_i words $\{w_i^1, \dots, w_i^p, \dots, w_i^{P_i}\}$.

The problem of annotating the reports is addressed by defining an ontology composed of N entities $\{E_1, \dots, E_n, \dots, E_N\}$ and M relations $\{R_1, \dots, R_m, \dots, R_M\}$. Then, Span-based Entity and Relation Transformer (SpERT) is used to annotate each report d_i, $i = 1, \dots, D$, by: 1) assigning each word w_i^p or span (e.g. $[w_i^{p-1}, w_i^p]$) to an entity E_n, and 2) assigning a relation R_m between two identified entities of the report. Figure 1 shows an example of annotation of a report.

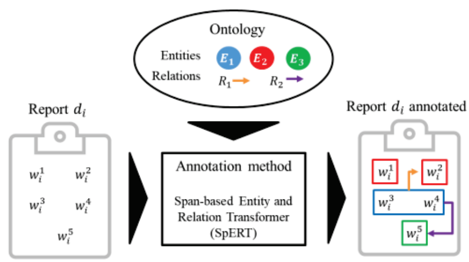

Fig.1: Annotation of a report

3. Ontology

We consider the ontology developed in (Bikaun et al. 2024). It features 224 distinct entities hierarchically organized in three levels. The first level contains five primary entities:

- *Activity*, indicating activities related to maintenance and support actions performed on physical objects (e.g., *refill*, *replace*, *clean*);
- *PhysicalObject*, indicating an object in the system (e.g., *pump*, *engine*, *valve*). The second and third levels classify the

objects based on their function (e.g., *HoldingObject*, *ConnectingObject*);

- *Process*, indicating accidental events occurring to physical objects (e.g., *leakage*, *stuck open*);
- *Property*, indicating the attributes of a physical object (e.g., *crack*, *isolated*);
- *State*, indicating the conditions of physical objects (e.g., *broken, degraded*).

and 6 relations:

- *contains*, used to denote the containment of physical objects (e.g., engine *contains* oil);
- *isA*, used to denote a subtype relationship between entities (e.g., diesel engine *isA* engine);
- *hasPart*, used to denote part-whole relationships (e.g., engine *has part* radiator);
- *has Agent*, used to denote entities actively involved or initiating an action or event (e.g., repair *has agent* operator);
- *hasPatient*, used to denote entities or undergoing an action or event (e.g., leakage *has patient* pipe);
- *hasProperty*, used to denote the possession of a particular characteristic by an entity (e.g., pipe *hasProperty* degraded).

In the present work, the ontology developed for the mineral processing and infrastructure industries is adapted for an application to freight transport trains with the objective of streamlining the annotation process. Specifically, the following modifications are made:

1) The structure of the ontology is reduced from three to two levels, by considering only the entities in level 2 of the original ontology;

2) Some level 2 entities not related to maintenance of transport trains (e.g. *PresentingObject*) are not considered.

3) The relation *contains* is merged with the relation *hasPart* due to their semantic similarity.

The use of an ontology with a limited number of levels and entities is expected to reduce the effort of the expert in the annotation process, and the possibility of inconsistent assignments of entities and relations.

Table 1 reports the entities and relations of the ontology considered in this work.

4. Annotation method

The annotation of the reports is performed using the Span-based Entity and Relation Transformer (SpERT) method (Eberts and Ulges 2020). SpERT is a state-of-the-art method for identifying entities and relations in texts. It is based on the Bidirectional Encoder Representation from Transformers (BERT) model. Unlike traditional annotation methods that process entities and relations in separate stages, SpERT enables simultaneous annotation of entities and their corresponding relations, which has been proven to improve contextual understanding and reduce errors (Eberts and Ulges 2020).

The MST is first tokenized, i.e. split into words and sub-words, and, then, transformed into a sequence of numerical vectors, which represent the contextualized semantic meaning of the tokens in a multidimensional feature space, by BERT. For entity extraction, SpERT employs a span-based approach, where multiple spans of text are generated and independently classified into a set of predefined entities using fully connected layers. This allows SpERT identifying entities of varying lengths, avoiding the limitations of token-level classification.

SpERT classifies also relations between pairs of extracted entities using other fully connected layers. These layers receive as input a combination of the contextual embeddings of the two entities and global features extracted by the BERT encoder at the report-level. As a result, SpERT can assess the semantic compatibility of entity pairs and connect them using relations defined in the ontology.

The two tasks of identifying entities and relations are performed independently, allowing the optimization of both objectives simultaneously without mutual constraints.

Table 1: Ontology used in this work.

	Level 1	Level 2	Examples
Entity	MaintenanceActivity	Adjust	*refill, clean*
		Replace	*replace, change*
		Repair	*refit, tighten*
		Reset	*restart*
	SupportingActivity	Isolate	*isolate*
		Observe	*notice, observe*
		Communicate	*say, announce*
	PhysicalObject	Controlling	*contactor*
		CoveringObject	*panel*
		DrivingObject	*engine*
		EmittingObject	*alarm*
		GeneratingObject	*battery, fan*
		GuidingObject	*cable*
		HoldingObject	*bolt, screw*
		HumanInteractionObject	*screen*
		InformationProcessingObject	*relay*
		InterfacingObject	*terminal box*
		MatterProcessingObject	*filter*
		ProtectingObject	*diode,*
		RestrictingObject	*resistor, brake*
		SensingObject	*sensor*
		StoringObject	*tank*
		TransformingObject	*inverter, pump*
	Substance	Gaseous	*air*
		Liquid	*oil*
		Solid	*dust*
	Personnel	Personnel	*operator*
	Process	DesirableProcess	*no operational impact*
		UndesirableProcess	*leakage*
	Property	DesirableProperty	*active*
		UndesirableProperty	*crack*, *stuck open*
	State	NormalState	*normal*
		DegradedState	*degraded*
		FailedState	*broken*
Relation	contains	contains	engine *contains* oil
	isA	isA	diesel engine *is a* engine
	hasAgent	hasAgent	operator *has agent* repair
	hasPatient	hasPatient	leakage *has patient* pipe
	hasProperty	hasProperty	pipe *hasProperty* broken

In this work, the annotation of the reports is performed considering the $N = 34$ entities and $M = 5$ relations in the second level of the ontology of Table 1.

The SpERT method is fine-tuned in two stages. In the first stage, the pre-trained BERT model in (Devlin et al. 2019) is fine-tuned using the dataset of 1067 maintenance reports taken from (Bikaun et al. 2024). This dataset, which had been previously annotated considering the ontology developed in (Bikaun et al. 2024), has been reannotated for the purpose of the present work considering the ontology of Table 1. This first fine-tuning stage serves as a foundation for adapting the SpERT to the domain-specific language and structure of maintenance texts. Then, a second fine-tuning stage is carried out using a smaller subset of reports from the case study considered, which has been manually annotated.

This two-stage fine-tuning process helps the method to learn both domain-specific semantics and case-specific annotations, improving its ability to accurately classify entities and relations in new maintenance reports.

5. Case study

We consider a repository of D MSTs of electrical freight transport trains. The exact value of D, which is in the order of hundreds, is not reported for confidentiality reasons. These MSTs were written by operators after maintenance interventions on electric and mechanical components of the trains. Two examples of MSTs are reported in Table 2.

Table 2: Two examples of MSTs.

Maintenance report
lp reported ios loss of ed braking. tcu was showing isolated
aux converter isolated normalized vcb open contactor faulty

Fifty randomly sampled MSTs have been manually annotated considering the entities and relations of the ontology of Table 1. Thirty of these MSTs are used for the second fine-tuning step of SpERT (Section 4), ten as validation set for setting the method hyperparameters and ten for evaluating the performance of the proposed annotation method.

6. Results

Fig. 2 shows a comparison between the manual annotations and the predicted annotations for a report of the test set. It can be noticed that the method correctly identifies most of the entities and relations, except the span *not solved*, which occurs only few times in the dataset compared to other spans such as *tripped*. Also, the method identifies the relation *hasPatient* between *normalization* and *isolation*. Although this relation is absent from the manually annotated report to avoid redundancy, it could still be considered valid as it connects a maintenance action to the occurred accidental event.

Tables 3 and 4 report the performance metrics of precision, recall and f1-score, which have been computed individually for the specific classes (entity and relation) and globally, as average across all classes.

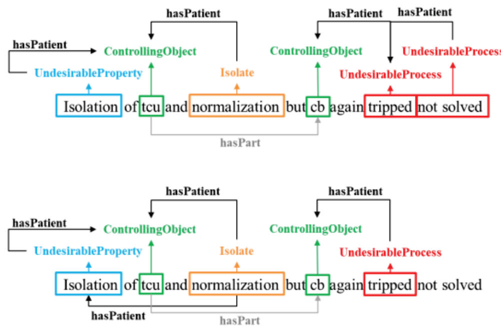

Fig.2: Comparison of the manual annotation (top) and the predicted annotation (bottom) for a MST of the test set

Precision is defined as the fraction of spans assigned to a specific class that correctly belong to that class, recall as the fraction of spans belonging to a specific class that are correctly assigned to that class, and f1-score as the harmonic mean of precision and recall. A span is considered correctly classified if it has been assigned to the same entity, whereas a relation is considered correctly classified if the relation and the connected spans are correctly identified, without considering the correctness of the classification of the two entities.

Concerning the classification of the entities, the results obtained on the test set are consistent with those reported in (Bikaun et al. 2024), where a global f1-score of 0.716 is obtained for the classification of the entities. As expected, the entities that are most present in the dataset are those that have been more accurately classified. For example, all the *ControlingObject* entities are correctly identified since most of the maintenance interventions involve issues with the electrical components related to the motion and control of the system.

Concerning the classification of relations, the overall performances are less satisfactory than those observed for the classification of the entities. This is in agreement with the findings of (Bikaun et al. 2024), where a global f1-

Table 3: Accuracy in the classification of the entities (sorted by number of occurrences in the test dataset).

Entity	Precision	Recall	F1-score	Occurrences in the test dataset
ControllingObject	1.00	1.00	1	6
UndesirableProcess	1.00	0.83	0.91	6
UndesirableProperty	0.80	1.00	0.89	4
Adjust	1.00	0.50	0.67	4
Isolate	1.00	1.00	1	4
StoringObject	0.00	0.00	0	3
Liquid	0.75	1.00	0.86	3
FailedState	0.67	1.00	0.80	2
NormalState	0.00	0.00	0.00	2
Observe	1.00	0.50	0.67	2
GuidingObject	0.50	0.50	0.50	2
Personnel	0.00	0.00	0.00	1
Replace	1.00	1.00	1.00	1
Average	0.85	0.73	0.78	40

Table 4: Accuracy in the classification of the relations (sorted by number of occurrences in the test dataset).

Relation	Precision	Recall	F1-score	Support
hasPatient	0.56	0.5385	0.5490	26
hasPart	1	0.5	0.6667	2
hasAgent	0	0	0	1
Average	0.5769	0.5172	0.5455	29

Table 5: Distribution of assigned entities in the *D* reports sorted by the number of occurrences.

Entity	Frequency
ControllingObject	0.51
TransformingObject	0.35
GuidingObject	0.08
StoringObject	0.01
RestrictingObject	0.01
InterfacingObject	0.01
SensingObject	0.01
DrivingObject	0.01
HoldingObject	0.01
CoveringObject	0.00

score of 0.655 is obtained for the classification of the relations.

The classification of relations is more difficult than that of entities because it requires to correctly identify not only the type of relation but also the two spans involved in the relation.

Notice that the support of the relations *isA* and *hasAgent* is relatively smaller than that of the relation *hasPatient*. This is due to the conciseness of MSTs, which tend to contain only spans related to the affected objects, their issues and maintenance actions, often omitting information (entities) of the subsystems to which they belong and the technician performing the action.

The method is then applied to automatically annotate all *D* reports. Table 5 shows the distribution of entities assigned to the identified spans and some examples. Notably, the most frequently assigned entities are *ControllingObject* (e.g., *contactor*) and *TransformingObject* (e.g., *oil pump*) as most of the maintenance interventions in the case study involve issues to electrical (e.g. *dynamic no closing*, *stuck open*) and hydraulic (e.g. *leakage*, *level low*) components.

7. Conclusions

In this work, we have developed a methodology for annotating maintenance reports of freight transport trains. The methodology builds upon the approach proposed in (Bikaun et al. 2024). It combines a streamlined ontology with the SpERT method for annotation. SpERT is fine-tuned first on a literature dataset and, then, on a small number of manually annotated reports from freight transport trains.

The accuracy obtained in the classification of the entities is satisfactory, indicating reliable identification of physical objects, their accident and maintenance actions performed in the reports. However, the accuracy obtained in the classification of the relations is lower, suggesting that challenges in correctly capturing the connections between entities are still present.

Future work will explore the use of Large Language Models (LLMs) to perform the annotation with the objective of improving the classification performances.

Extracting reliability and maintenance knowledge from maintenance and safety reports can provide valuable information. The automated annotation of reports is a first step towards the development of methodologies to support maintenance decision-making. Systematically identifying and counting the occurrences of entities and relations is at the basis of the quantitative assessment of the frequency of occurrence and severity of malfunctions and failures.

Acknowledgements

The authors gratefully acknowledge Alstom for providing the data and for their support in the development of the proposed methodology and in the analysis of the results.

The participation of Dario Valcamonico and Enrico Zio to this work is supported by the SAFEPOWER project under the HORIZON-CL5-2024-D3-01 call of the European Union, Grant Agreement 101172940. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting authority can be held responsible for them.

The participation of Piero Baraldi to this work is supported by FAIR (Future Artificial Intelligence Research) project, funded by the NextGenerationEU program within the PNRR-PE-AI scheme (M4C2, Investment 1.3, Line on Artificial Intelligence).

References

Bikaun, Tyler K., Tim French, Michael Stewart, Wei Liu, and Melinda Hodkiewicz. 2024. “MaintIE: A Fine-Grained Annotation Schema and Benchmark for Information Extraction from Maintenance Short Texts.” In *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*, 10939–10951. https://aclanthology.org/2024.lrec-main.954.

Brundage, Michael P., Thurston Sexton, Melinda Hodkiewicz, Alden Dima, and Sarah Lukens. 2021. “Technical Language Processing: Unlocking Maintenance Knowledge.” *Manufacturing Letters* 27: 42–46. https://doi.org/10.1016/j.mfglet.2020.11.001.

Conte, Anna, Coline Bolland, Lynn Phan, Michael P. Brundage, and Thurston Sxton. 2021. “The Impact of Data Quality on Maintenance Work Order Analysis: A Case Study in Historical HVAC Maintenance Work Orders.” In *Proceedings of the European Conference of the PHM Society*. https://doi.org/10.36001/phme.2021.v6i1.2814.

Devlin, Jacob, Ming Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. “BERT: Pre-Training of Deep Bidirectional Transformers for Language Understanding.” In *NAACL HLT 2019 - 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies - Proceedings of the Conference*, 1:4171–86. https://doi.org/10.18653/v1/N19-1423.

Eberts, Markus, and Adrian Ulges. 2020. “Span-Based Joint Entity and Relation Extraction with Transformer Pre-Training.” In *24th European Conference on Artificial Intelligence (ECAI 2020)*. https://doi.org/10.3233/FAIA200321.