Asset Problem, Cause and Remedy Extraction

Jordan Makins

Wei Liu, and Michael Stewart
Department of Computer Science And Software Engineering
University of Western Australia

Melinda Hodkiewicz
Department of Mechincal Engineering
University of Western Australia

Melissa Aspinall
CEED Client: Water Corporation

Abstract

Historical maintenance work order (MWO) records contain unstructured data about loss of asset function and the maintenance actions required to restore or retain function. The Computerised Maintenance Management database also provides structured fields for the maintainer to identify, using codes, information of interest to reliability engineers such as the 'problem', 'cause', and 'remedy' for each MWO. However, these codes are often recorded inconsistently by the operators, maintainers, and planners generating the MWOs leading to issues when they are used by reliability engineers for analysis and to assess effectiveness of maintenance strategy. The aim of this project is to apply Technical Language Processing (TLP) to the unstructured MWO texts to infer 'problem', 'cause', and 'remedy' codes in a consistent way, removing reliance on the human data entry. TLP is computer-driven manipulation, understanding and interpretation of human language on engineering texts. The TLP pipeline involves annotation of the 449 unstructured texts to map words to named entities related to semantic concepts in the texts. Classification accuracy as measured by F1-score is 0.925 for item, 0.75 for problem and 0.875 for remedy. There is limited causal information in the unstructured text field and further work using additional data such as from Failure Mode & Effects Analysis will be required. The value proposition of this work is to improve the productivity of reliability engineers who currently check these codes manually and to correct codes automatically and at scale on historical work orders to aid analysis.

1. Introduction

The Water Corporation (WC) manages a large asset base of over $38 billion dollars (replacement value). To balance the costs and risks associated with managing their assets, the Water Corporation has developed maintenance strategies for each asset. Strategies utilize Failure Mode, Effects and Criticality Analysis (FMECA), operation and maintenance manual recommendations from Original Equipment Manufacturers, and existing personnel expertise. A MWO is generated each time a task, either preventative or corrective, is required. Between 1999 and 2021, over 30,000 MWOs were generated for the Borefields A, B, C and D. A subset of 449 MWOs from the locations have been annotated for this project as annotations require a team of domain experts and are time intensive. Bore fields have been selected as they are a high-cost area. Bore field data can serve as a proof of concept for application of TLP across broader asset locations. Bore fields contain the following items: pumps, motors, valves, switchboards, bores, pipes, alarms, and communication systems. Examples of problems with bore field items are leaks and noise; remedies are replace, refit and repair; and causes are wear, overheat and looseness. Only around 30% of data colleced are useful in current form as structured codes for ‘problem’, ‘cause’, and ‘remedy’ are often missing or noniformative. With the exception of ‘cause’, this information is frequently and explicity available for data mining in the unstructured long text from MWOs. However, manually processing these to extract information in the form of codes for analysis is a time-consuming task with challenges in maintaining consistency when multiple people are involved. The use of Natural Language Processing (NLP) with machine learning (ML) for improving productivity and automating mundane tasks has seamlessly become integral to everyday life. Consider the use of digital assistants such as Siri and Amazon Alexa. Digital assistants enable end-users with the ability to quickly search information on-demand to prevent the need to laboriously type queries into a search engine. Search engines such as Google Search predict popular and relevant searches based on similar searches. Gmail auto categorises emails into either primary, social, or promotions based on their contents. These platforms are driven by NLP through the interpretation and understanding of human natural language.

Furthermore, ML models are easily maintainable and improve over time as they are exposed to more data. An industry-wide problem with unstructured maintenance text data is that it is often incomplete, inconsistent, and laden with technical jargon and organizational abbreviations. Off the shelf NLP performs poorly on technical texts like work orders and as a result TLP pipelines focussed on maintenance texts are being developed (Brundage, Sexton, Hodkiewicz, Dima & Lukens, 2021). The primary objective of this research is to assess the performance of a TLP-based pipeline written in Python for extracting ‘problem’, ‘cause’ and ‘remedy’ codes from unstructured text. TLP will save time and provide an additional means of categorising MWO and structuring data to make more information readily available to reliability engineers.

2. Process

The pipeline involves generating insights from the ML model by constructing a graph database and web-based visualisation interface as illustrated in Figure 1. Firstly, the Python code normalises the maintenance text to improve annotations and the downstream task of ML. The ML model is trained and then deployed on unseen maintenance free text to predict entities. Postprocessing is executed to relate items in the same work order, relate specific problems to items and the remedies performed. Lastly, the work orders are visusalised as a knowledge graph for analysis by reliability engineers.

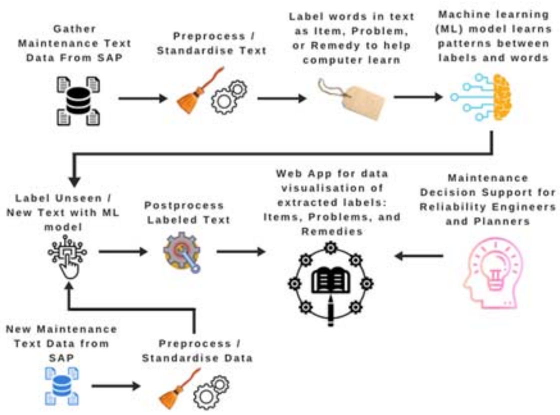

Figure 1 Diagram of the Dataflow to generate insights for Reliability Engineers.

2.1 Data Preprocessing

The objective of the preprocessing stage is to make the text more consistent, so the ML algorithms can better learn contextual word representations (Babanejad, Agrawal, An & Papagelis, 2020). The pre-processing script is written in Python and aims to standardise text with the procedure shown in Figure 2. The preprocessing script focuses on correcting english words and is designed to ignore organizational and maintenance specific words and abbreviations, which will be addressed in the postprocessing stage.

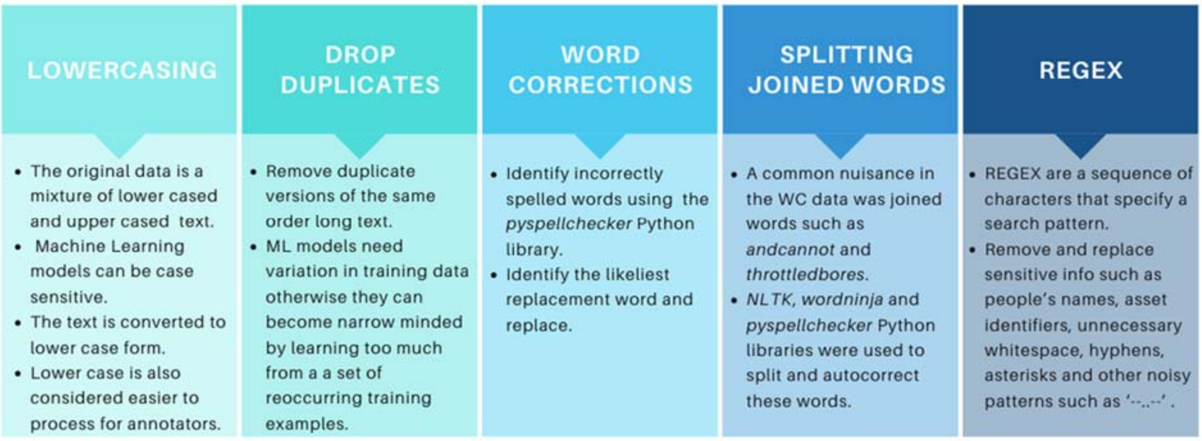

Figure 2 Fundamental procedures of the pre-processing script developed for WC.

2.2 Annotating Data for Machine Learning

Supervised ML requires labelled datasets so that computers can learn patterns and predict or classify new data accurately. Annotating is the process of assigning labels to unlabelled data. Annotations were performed on Redcoat, a web-based annotation tool (Stewart, Liu & Cardell-Oliver 2019). A team featuring several subject matter experts assigned labels such as *Item*, *Remedy*, *Problem*, *Function*, *Malfunction*, and *Negation* to words in the maintenance text. *Problem* is extended to group instances from the text where a *Malfunction* is present, or a *Function* tag is preceded by a *Negation* tag. A total of 449 documents representing maintenance records from Borefields A, B, and C were annotated to achieve an average of 53% annotator agreement. The annotations for each maintenance record are assigned an annotator agreement measure.

2.3 Machine Learning

ML requires data to be split into training, validation, and test sets. Best results were achieved with a 80:10:10 split, i.e. eighty percent of the data was assigned to the training set, ten percent to the validation and the remaining ten percent to the testing dataset. A state-of-the-art language model called Bidirectional Encoder Representations from Transformers (BERT) was selected to learn sentence context based numerical representations of words from the body of words present in WC data (Devlin, Chang, Lee & Toutanova, 2018). A sequence of words from maintenance records are the inputs as shown in Figure 3. BERT aims to model the semantic meaning of a word in numeric form. The outputs of the first machine learning model are the context based numerical representations of words, which are then passed into a final machine model for automatically classifying words as *items*, *problems* and *remedies*.

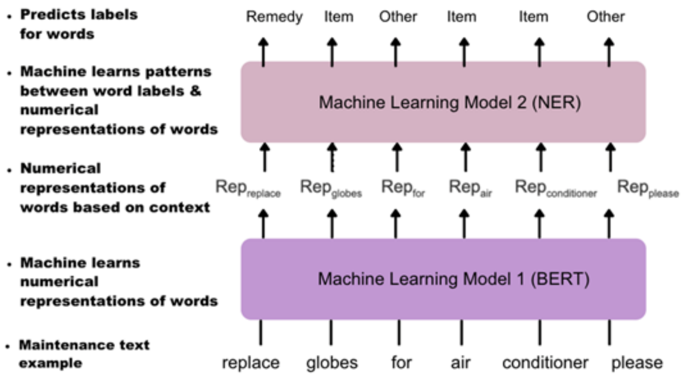

Figure 3 Machine learning process for extracting Item, Problem and Remedy from maintenance text.

2.4 Data Postprocessing

The postprocessing script adds a list of categories from ISO15926 to group the predicted word-classes. For example: the parent category *rotating equipment* is applied to the *item* words *motor*, *pump* and *compressor*. This facilitates simple and efficient querying of asset classes when interacting with the data visualisation. The code normalises maintenance and organization specific terms. For example, as *clean* and *cln* are commonly present in the text, the variant *cln* is replaced by *clean* to keep insights concise and consistent. Similarly, *magflo* is replaced by *magnetic flow meter*. The corrections dictionary is easy to maintain as new key-value pairs only need to be entered once.

2.5 Data Visualisation for Decision Support

A graph database platform stores the structured data to interactively support querying and visualization of the MWOs. The graph is filled with the outputs from the ML model. *Items*, *problems*, and *remedies* are represented by the coloured nodes green, red and blue respectively. The graph database can be queried by WC personnel for answers to questions such as how many motors have failed in the past month, are the motors failing in ways that are in line with expectations or not, which bore sites have the most reactive maintenance and planned maintenance, and what are the most frequent problems occuring in Bore Fields.

3. Results and Discussion

The results show that the ML model is capable of accurately and automatically extracting words that are *items*, *problems*, *remedies* from substantial amounts of maintenance text. Key performance metrics such as precision, recall, and F1-score were used to evaluate the accuracy of the ML model at accurately extracting which words are classed as *problems*, *remedies* and *items*. Precision measures the proportion of words that are predicted as *items* and are actually *items*. Recall measures the proportion of total *items* that are correctly predicted as *items*. Table 1 displays the classification results on WC data. F1-score combines precision and recall for an overall measure of class-weighted model accuracy. It is less prone to biasing classes with higher relative frequency than the traditional accuracy measure.

By Class:	Precision	Recall	F1-Score
Item	0.8889	0.9333	0.9256
Problem	0.7200	0.8077	0.7500
Remedy	0.8810	0.9024	0.8750

Table 1 The best model classification results from the machine learning model achieved an overall F1-Score (micro) of 0.8338.

The problem 'leak' was identified as a frequently occuring problem over the past five years with specific regard to valves. There were a total of at least nine valve leaks as shown in Figure 4. Reliability engineers can now unlock value by finding solutions to the following questions such as: Are we using a run-to-failure strategy with regard to valves and is the current strategy the best option? Is it necessary to put preventative maintenance in place or do we continue along with current strategy?

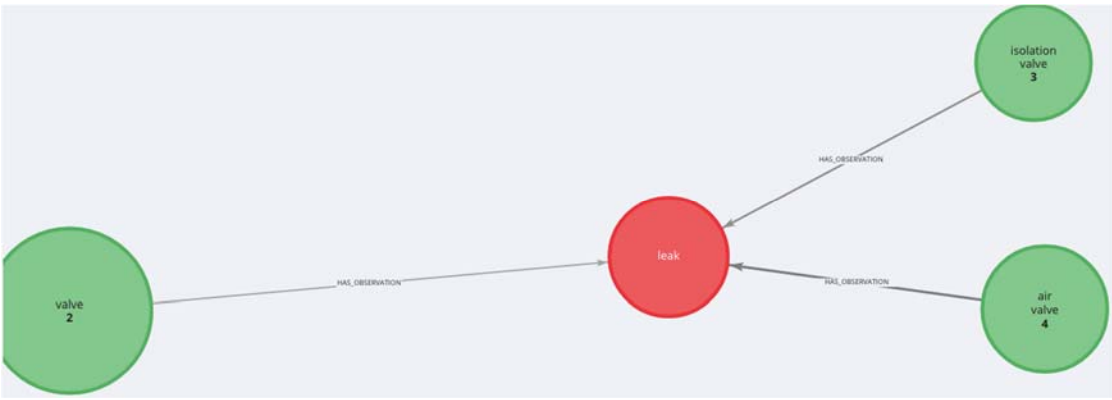

Figure 4 Leak was identified as a frequent problem with valves.

A specific bore site was identified as requiring a higher level of maintenance activity relative to other bore sites as shown in Figure 5. This bore site can be investigated to inspect the granularity of the specific maintenance tasks that were performed on the bore site over the five year period as illustratred in Figure 6. From the data-driven insight, reliability engineers can now perform a thorough investigation as to why this bore site needs excessive attention, and gain value by answering the questions: Are we aware of this intuitively or is this a surprise, and how can we adjust strategy accordingly?.

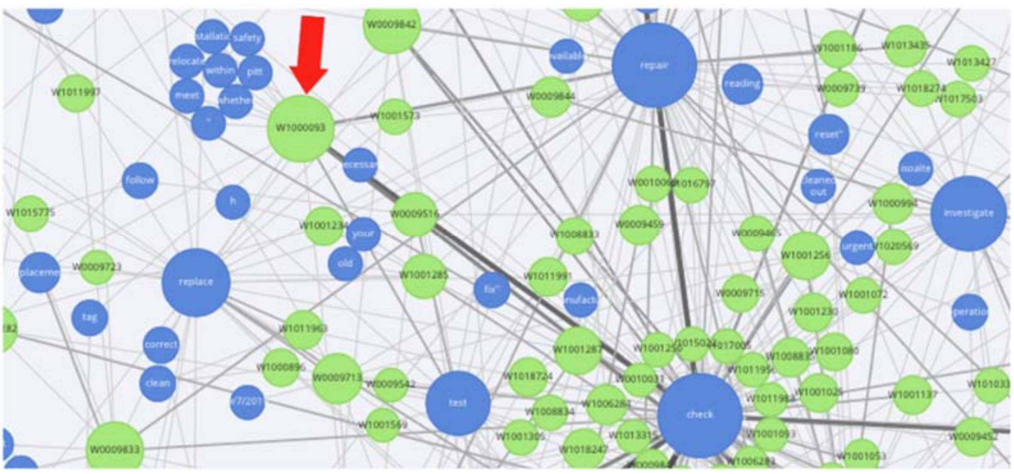

Figure 5 Survey and identify which bore field site has the most maintenance activity based on node size.

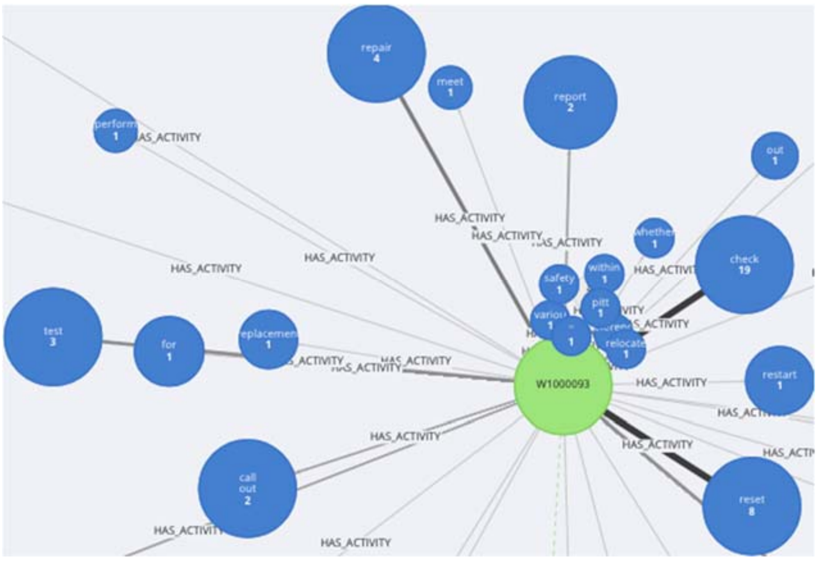

Figure 6 Maintenance activities performed at a specific bore field site with a higher relative level of activity in the past five years.

4. Conclusions and Future Work

The use of a structured data pipeline approach ensures the results are repeatable. The graph database serves as an effective decision support tool for reliability engineers and planners. Significant cost savings could be achieved by adjusting asset management strategy. Further work to be undertaken includes methods to automate the currently human-driven annotations of the maintenance text as this step is both technically challenging and time demanding. Another area of improvement could be exploring lexical normalisation methods for automating the standardising and cleaning of organization specific maintenance data. This should improve annotation consistency and downstream ML tasks such as NER.

5. Acknowledgements

I would like to acknowledge and express gratitude to my academic supervisors: Melinda Hodkiewicz, Michael Stewart, Wei Liu, and my Water Corporation mentor Melissa Aspinall for their ongoing support and positive encouragement. I would also to thank Water Corporation, CEED Director Jeremy Leggoe and Amanda Bolt for providing this research opportunity.

6. References

Babanejad, N., Agrawal, A., An, A., & Papagelis, M. (2020, July). A comprehensive analysis of preprocessing for word representation learning in affective tasks. In Proceedings of the 58th annual meeting of the association for computational linguistics pp. 5799-5810.

Brundage, M. P., Sexton, T., Hodkiewicz, M., Dima, A., & Lukens, S. (2021). Technical language processing: Unlocking maintenance knowledge. Manufacturing Letters, 27, 42-46.

Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805.

Stewart, M., Liu, W., & Cardell-Oliver, R. (2019, November). Redcoat: a collaborative annotation tool for hierarchical entity typing. In Proceedings of the 2019 Conference on empirical methods in natural language processing and the 9th international joint conference on natural language processing (EMNLP-IJCNLP): system demonstrations pp. 193-198.