NLP for Enterprise Asset Management: An Emerging Paradigm

Pedro Santos*†, Nuno Datia*†, Matilde Pato*‡, José Sobral*§, Nuno Gomes*†, Noel Leitão¶ and Manuel R. Ferreira¶
**Lisbon School of Engineering (ISEL) Politécnico de Lisboa*, Lisbon, Portugal
†*NOVA LINCS, NOVA School of Science and Technology*, Monte da Caparica, Portugal
‡*LASIGE, FCUL, Universidade de Lisboa*, Lisbon, Portugal
§ *CENTEC, Technical University of Lisbon (IST)*, Universidade de Lisboa, Lisbon, Portugal
¶*TDGI – Property Management Technology*, Porto Salvo, Portugal
Email: *pedro.santos@deetc.isel.ipl.pt, *†datia@isel.ipl.pt,

***Abstract*—In the field of asset management, a Work Order refers to a document that outlines the necessary steps to carry out a maintenance operation on a specific physical asset. The text on this Work orders providing details about the problem and the actions required are open-ended, not normalized, and Technician' dependant, presenting challenges for automating asset management Work Order processing. To address the issue of automating the analysis of Work Orders, Natural Language Processing techniques are employed to process the content of these documents. The aim is to identify and extract relevant information related to actions and components within the sentences. This paper presents the Reliability Centred Maintenance for Assets solution, which utilizes a semi-automatic, human-in-the-loop approach to determine a standardised and condensed set of actions and components. The results indicate a significant increase in the number of annotations, reaching a ratio of 1:14. By implementing this solution, the manual workload associated with analysing Work Orders can be reduced, thereby improving decision support and analytical processing of the data contained within these documents.**

***Index Terms*—NLP, Asset Management, KB curation, Interactive interface, Industry 4.0**

I. Introduction

Asset management focuses on the operation management and maintenance of assets, considering strategic optimisations and decisions such as maximising each asset's working time, minimising downtime costs and predicting when a failure might occur. The primary purpose of asset management is to get the maximum value from assets, making companies more productive and with less associated costs [1, 2]. Physical assets are any item that has value or produces a value that is under the domain of a given organisation, such as machines, equipment, and tools. The maintenance management during an asset's life cycle is achieved through WOs. Every WO has a field that allows the technician to describe the job accomplished and other relevant details that characterise it. The created sentences are considered "free text", left to the technician's criteria. Consequently, this freedom can lead to spelling errors, non-existent words, or a lack of consistency in the terms used. Hence, different terms are used to describe the same situation in WO that are alike.

Inconsistent terminology can cause problems in various fields of society, including asset management [3]. We identified three of the most relevant situations from the problem domain expertise contacts: (i) *Miscommunication*: When different people use different terms to describe the same concept, it can lead to confusion, misunderstandings, and misinterpretation of information. This can result in errors, delays, and other problems; (ii) *Reduced efficiency*: Inconsistent terminology can slow down information sharing and collaboration, especially when multiple teams or departments are involved. This can lead to delays and productivity loss; (iii) *Reduced accuracy*: Inconsistent terminology can also result in inaccurate data, making it difficult to track and compare information across different systems, databases, or reports.

The data used in this research is gathered from real WOs, created in the context of asset management activities in a hospital by a long-standing company in the field. The analysis of the *corpus* in the WO descriptions sheds light on the variety of descriptions for the same problem [4]. The company had already implemented a taxonomy-like solution, resulting in significant mismatches between the WO description and the problem solved. The root cause of this mismatch is the extra time the technician needs to fill a WO by searching for the right words on the taxonomy. They opt to put something that allows them to finish the WO and continue working. One way to address this problem and get a better set of words describing what has been done in WOs is to post-annotate them. Annotation can bring several benefits to asset management: (i) *Increased accuracy*: Annotations can improve the accuracy of asset management by providing additional information about an asset and the corrective actions that have been made; (ii) *Improved organisation*: Annotations can help categorise and organise assets. Adding descriptive labels or tags makes it easier to locate and retrieve assets when needed.

This work was supported by the REV@CONSTRUCTION mobiliser project, under the grant LISBOA-01-0247-FEDER-046123 from ANI - National Innovation Agency, and by NOVA LINCS (UIDB / 04516/2020) and LASIGE (UIDB/00408/2020 and UIDP/00408/2020) with financial support from FCT— Fundação para a Ciência e a Tecnologia, through national funds. This work contributes to the Strategic Research Plan of the Centre for Marine Technology and Ocean Engineering (CENTEC), which is financed by FCT under contract UIDB/UIDP/00134/2020.

2375-0138/23/$31.00 ©2023 IEEE
DOI 10.1109/IV60283.2023.00049

This can save time and effort when searching for assets in an extensive collection; (iii) *Improved searchability*: Annotations can make assets more searchable by providing additional context and information. This can improve asset management efficiency and save time when searching for specific assets. This work proposes a solution with a human-in-the-loop that uses Natural Language Processing (NLP) to process the WOs and derive a type of annotation — the action done on the assets and the asset component that was intervened. With the identified annotations, WOs can be automatically processed by identifying action-component pairs. Since some WOs may include unknown terms, a manager may need to enter some data that allows the algorithm to continue processing the WO [2]. The developed solution can extract valuable knowledge from historical WO. The user interface helps a domain expert curate the Knowledge Base (KB) that will support future WO processing and also contribute to better decision support when analysing problems reported in WOs.

The contributions of this work are the following: 1) integrating learning methods, modelling uncertainty through probability distributions, including conditional randomisation for context-specific variation, and incorporating human expertise and feedback;2) WEB application designed to reduce the manual workload of processing WOs, by allowing domain experts to interact with the learning algorithm and the KB. This paper is structured as follows: Section II presents the research background before this research, and Section III describes the solution, including the data used, the modelling and algorithms used. Section IV evaluates RCM4Assets, and Section V concludes and points out future directions for work.

II. Related Work

For years, inconsistent terminology has been a research topic in many fields [5]–[7]. Ontologies and taxonomies can help with vocabulary consistency, but they are difficult for technicians to use when they face tight deadlines. More recently, the usage of NLP solutions has increased, in particular, to address some problems in management [8]–[10]. Many approaches use NLP to tackle this problem; some use more classical Machine Learning (ML) approaches [10, 11], while others use deep learning-like solutions [12, 13]. Payette et al. [12] use NLP to improve the quality of historical maintenance data for Hydro-Québec TransÉnergie's power transmission assets. In particular, the authors have used Long Short-Term Memory (LSTM) with Encoder Representations from Transformers (BERT), to create a classifier for the assets (Corrective, Preventive or no maintenance). Alternatively, Sexton et al. [3] proposed a method to prepare maintenance logs, i.e. WO, for statistical processing. They employed a hybrid solution that combined ML techniques with human guidance. To establish a controlled vocabulary (referred to as "folksonomies"), they utilised NLP to identify relevant tags. The domain experts' input was crucial for deciding which tags were essential, separating them from the non-informative ones. A linear-kernel Support Vector Machines (SVM) was used to classify logs in a One-vs-Rest multi-label scheme. In a different domain, Obeid et al. [14] employed a human-in-the-loop approach to annotate corpora using part-of-speech (POS) tagging. The annotation interface allowed users to correct misclassified terms. The annotation of WO involved the use of NLP, specifically POS tagging and ML. WO containing unknown words were placed in a queue, allowing domain experts to edit and select terms related to actions. This process enabled the automation of processing for other WO.

III. RCM4Assets Solution

These maintenance logs consist of WOs with several fields (70). Technicians fill in the information to best describe the intervention and the repair made (if any). The dataset is collected from real WO from a healthcare facility, totalling $38\,445$ WOs. Every WO has the unique identification of the asset (an id, a family descriptor, among others) and has a description field containing the action made. This field is processed using NLP to extract relevant terms that describe the WO with a normalised vocabulary. However, some WO has a different writing style, unknown words, or errors, making it impossible to derive a concise description automatically. The Reliability Centred Maintenance for Assets (RCM4Assets) system was developed to address this problem. It leverages a large heterogeneous set of historical maintenance logs and is designed to have domain experts aid in solving ambiguous situations. Fig. 1 presents an overview of the system, where two significant components are displayed: (i) The RCM4Assets AI Core (V); and (ii) the RCM4Assets UI (U). Raw WO (A), the ones taken directly from the asset management system, are processed to extract a set of relevant terms, identifying (B) the action(s) made upon which component(s) of the asset. When the terms are not in the KB, WO is put into a waiting queue (D) until a domain expert identifies the action and the component or edits the WO, correcting any eventual mistakes (C). The new terms are inserted into the KB, and the domain expert can trigger automatic processing on all unprocessed WO (E). This is a two-fold action. On the one hand, it updates the ML model with the recently provided information about previously unknown activities and components; on the other hand, it leverages the new knowledge to process the WO on the waiting queue automatically. RCM4Assets aligns with Industry 4.0 by incorporating AI into industrial processes, aiming to develop procedures that are more productive, flexible, efficient, and profitable.

A. Formalisation Of The Problem

Let us consider a WO $w_i \in W$. Each w_i contains a free text describing the actions done (and in what components) to solve the reported problem, represented by X — a set of sentences. Let us consider the pair (a, c) where a is an action done over a component c for a given asset. For the sake of simplicity, the asset is associated with w_i since each WO is done over one and only one asset. Let us consider a KB to be a consistent set of assertions. Let (a, c) be considered valid if $(a, c) \in$ KB. Notice that c can represent more than one word interpreted as an atomic term in the domain (e.g. air filter). We want to

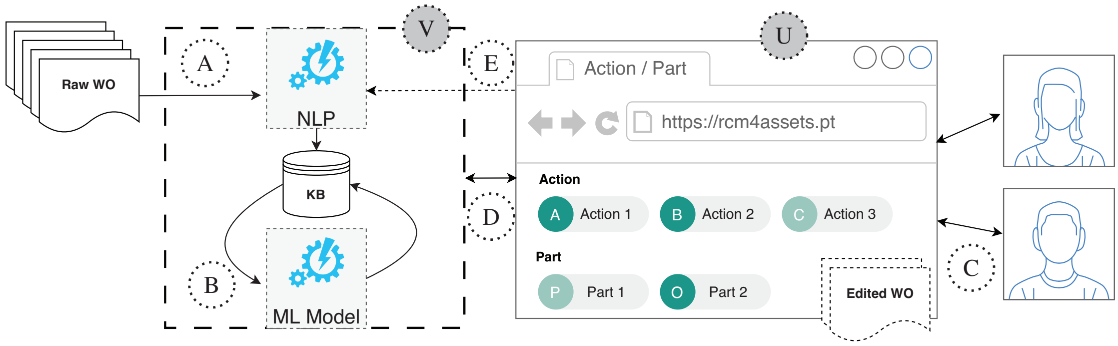


Fig. 1: Overview of the RCM4Assets system. Raw WOs are automatically processed, based on the existing knowledge stored in the KB. The WO that need attention from the users are signalled. After a human analysis, they are re-processed automatically.

derive a function $\mathscr{F}(X)$ that maps each w_i into a set of (a, c), such

$$\mathscr{F} : x_i \rightarrow \{(a_1, c_1), (a_2, c_2), \ldots, (a_n, c_n)\} \tag{1}$$

where $x_i \in X$ and $(a_1, c_1), (a_2, c_2), \ldots, (a_n, c_n)$ is named A_i — the actions in the WO w_i. Now let us consider G, to be the set of WO, such

$$\forall x_i \in G, \forall_i^{|A_i|}(a_i, c_i) \in \text{KB} \tag{2}$$

and R, to be the set of WO, such

$$\forall x_i \in R, \exists_i^{|A_i|}(a_i, c_i) \notin \text{KB} \tag{3}$$

B. RCM4Assets AI Core

RCM4Assets core implements the function $\mathscr{F}$, thus converting each work order w_i into set A_i. Using some domain terminology, each $(a, c) \in A_i$ is a service catalogue. Do notice the KB is updated from the interaction with the domain expert user, adapting to the changing scenarios regarding asset management. It consists of three modules: (i) NLP, (ii) ML model, and (iii) a KB. First, a subset of about $19,000$ WO is considered, each going through a NLP pipeline [4]. The following steps are applied: tokenisation, lowercase, stopword removal, multi-word token expansion, standard POS tagging and lemmatisation. The ML model is initially trained using this subset of work orders to learn the words and their corresponding standard tags in the *corpus*; this subset contains unique descriptions. That is, no exact work orders are considered. Furthermore, the ML model consists of a custom-made POS tagger that can recognise infinitive verbs, concrete nouns, and other standard tags (e.g. pronouns). The infinitive verbs identify each sentence's actions (a), and concrete nouns (c) identify components. Thus, we can compute the pairs (a, c) according to the state of the world stored in the KB, which is bootstrapped with 31 actions and 46 components. If the ML model does not recognise a given the word as an action (the infinitive verb), a standard "verb" tag is attributed because the initial training is only done on standard tags. Otherwise, the "infverb" tag is used. In other words, the methodology begins with an NLP pipeline to process WOs and assign standard POS tags to words. These standard tags are used to train the model and serve as a foundation for generating responses when the model lacks sufficient information.

Hence, by combining the common POS tags assigned to each word in the NLP pipeline with the known actions and components present in the KB, it is possible to replace "verb" and "noun" tags with "infverb" and "concnoun", respectively, for cases where the KB has identified the word as an action or component. Thus, the model is based on a Conditional Random [15] and is trained using the words and associated tags present in this initial set of WOs alongside the data current in the KB. In cases where the model does not know a given action or component, the WO belongs to the R set, and it is signalled to be analysed by a human in the user interface (see Fig. 2.U). When automatically processing a WO, there are rules to decide if it belongs to the G or R set.

These rules have been carefully discussed with domain experts so that the algorithm converges to automatically process well-formed WO and filter out those with some inconsistency. They are applied after ML classification. The first rule is whether or not the model is aware of all the actions and components that exist in a given WO description, that is if a "verb" or "noun" tag is found after model prediction instead of the "infverb" or "concnoun" tag, the WO is considered to belong to the R set (Eq. 3). Then, it is analysed whether at least one action and one component are known, one "infverb" and one "concnoun" tag, respectively. Finally, the remaining rules correspond to cases where the technician performed a set of actions in many components, so the WO is evaluated for consistency in (a, c) pairs. If the WO passes all the rules, the action component (a, c) pairs are identified and converted to service catalogues. The WO is considered to be automatically closed. For the cases where the WO does not pass one of the rules, it is considered to belong to the R set and set to be analysed by the domain expert.

After user interaction, R is transformed into G and uses the same modules and procedure. Since the user interaction defines a new set A_i, accordingly stored in the KB, the automatic

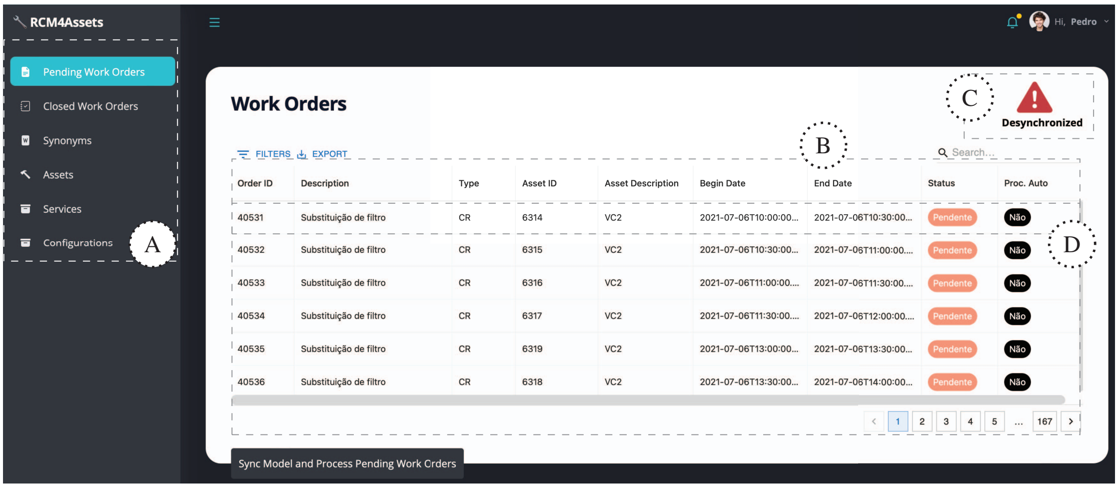


Fig. 2: The interface of the RCM4Assets. The depicted screenshot implements the visualisation of the R set of WO. The highlighted areas represent key components for the interaction and to get visual feedback based on the exchange, namely: A the menu area; B a place to show detailed information upon action on the menu; C a status indicator indicating the need to sync the G and R sets; D a WO representation where the domain expert user needs to take action.

processing of R after one interaction round will generate a new set, G' such

$$|G'| \leq |R| \tag{4}$$

For most cases,

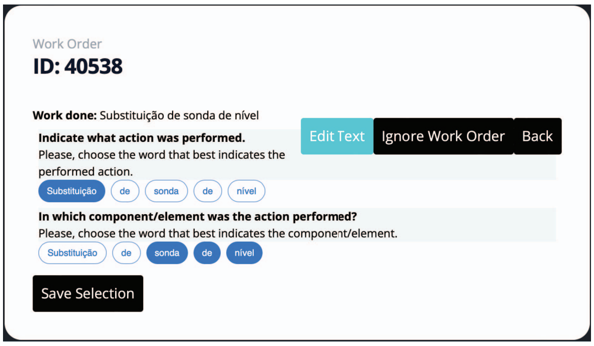
$$|W^t| \ll |R| - |G'| \tag{5}$$

where $|W^t| \in |R|$ and represents the WOs that have been tweaked by the user. The exceptions are some very specific WO, as shown in our previous work [2].

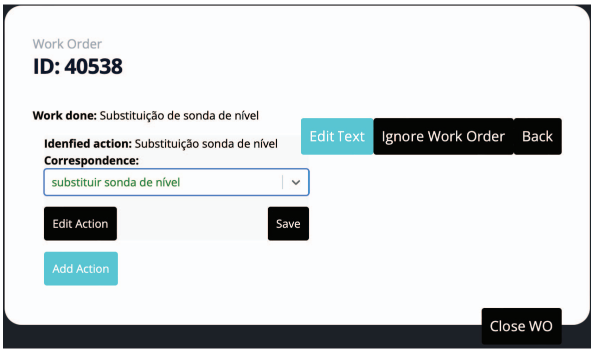
C. RCM4Assets Interactive UI

To gather continuous feedback from domain experts, we chose to conduct week-long sprints. This is important because user feedback is crucial for interface design, as it provides valuable insights into how users engage with a product and what they expect from its features. Continuous user feedback allows designers to identify pain points, usability issues and areas for improvement [16]. By gathering feedback throughout the design process, designers can make informed decisions and create user-centred designs. In addition, user feedback helps designers to prioritise features and make design decisions based on what users need RCM4Assets system implements $\mathscr{F}(x)$ in the AI Core (Fig. 1.V) and transposes elements from R into G implemented a human-in-the-loop, semi-automatic (Fig. 1.U) approach. The design of the RCM4Assets UI tries to simplify the information presented to the domain expert while presenting a powerful functionality [14]. Among other actions, a domain expert user can override the data inserted by the technician. The main UI is depicted in Fig. 2. Although the design is not polished at this time, we do our best to follow some of the desired features from Human Computer Interaction (HCI), namely, it: *(i)* has affordances (e.g. Fig. 2.D); *(ii)* provides feedback on the user's action (e.g. clicking on a line changes the interface, Fig. 2.B). The interactive UI is designed to be used by domain experts. For the sake of simplicity, from this point forward, term user will refer to the domain experts. The selection menu (Fig. 2.A) presents a set of options, namely: (i) the WOs that need attention, (ii) the closed WOs, (iii) the list of similar actions (synonyms), (iv) information about the assets, (v) the existing services (the pairs), and (vi) the application configuration. Only the first one allows interaction between the user, the algorithms, and the KB.

The selection of the first option on the menu will show the list of WOs that need human intervention (Fig. 2.B) — R set. It should be worth mentioning that annotating a couple of WO can already lead to the automatic processing of many other WO, as indicated in Eq. (5). Clicking in one line shows the interface to define the set A_i for a given WO, as depicted in Fig. 3a. This first step allows the user to enter one action and one component associated to the WO. Confirming the pair leads us to the next screen (Fig. 3b), where we can select a synonym from the KB. When you select the word that indicates action in the previous screen (Fig. 3a), the KB only stores verbs in their infinitive form, which makes it possible to establish a correspondence (synonym) between the technician's word and the KB's infinitive verb. This synonym identification is made by querying the KB for infinitive verbs that match the first four letters with the word selected as the action; if no correspondence is found, the user is asked to type the given action as an infinitive verb. Conjugated verbs and abstract nouns, which indicate action, are mapped to infinitive verbs in the KB, establishing synonyms. These



(a) Interface to define the pair (a, c) for a given WO. The component can be a compound word. This means that more than one chip can be selected.



(b) Select a synonymous from the list, if any, on enter a new pair in the KB.

Fig. 3: Screenshots of the steps for entering the pair (a, c) for one WO. 3a it is shown after the user click on a WO (Fig. 2.D) and 3b appears after the user selects an action and a part.

synonyms are then used to convert a given conjugated verb or abstract noun into an infinitive verb, which the ML model can identify. The example shows a transformation from an abstract noun ("substituição", meaning "replacement") into an infinitive verb ("substituir", meaning "replace"); if action or component do not exist, the pair (a, c) is added to the KB. This manual processing is essential since it allows for knowledge discovery regarding what words constitute actions and components, which is then reflected in the KB and, consequently, in the ML model.

Do notice the presence of prepositions ("de"). In the initial prototypes, they were omitted. However, based on continuous user feedback, the inclusion of prepositions was requested to enhance understanding of the actions taken, as they provide additional context. The WO can be closed when all the actions and components are covered. That action triggers a warning in the interface, telling the user that new information is available. We may want to sync the RCM4Assets AI core with the newly entered information and try to automatically reprocess the R set (Fig. 2.C). You can see from the content of Fig. 2.B that using the first WO to enhance the KB, we can process automatically, at least, all the WO depicted in the first page.

IV. Evaluation

User interface evaluation, which ensures that interface design meets the needs and expectations of the users, is an essential aspect of software development. Evaluating the interface helps identify potential design flaws and usability issues and provides insight into user behaviour and preferences. However, most of the interface problems, as far as the users are concerned, were addressed, mitigated or postponed since the UI development was made using a continuous feedback approach. Thus, users already gave feedback in the early stages of development, although a qualitative one. Therefore, it is essential to assess another aspect of the RCM4Assets solution — the ability to amplify each user interaction — lowering the burden of manual annotation of many WO [17]. Therefore, the evaluation process focuses on understanding the relationship between each WO changed by a user and the number of WOs that can be processed automatically because of the KB update. The evaluation procedure is the following. First, the most recent $1,000$ WO set was selected, and the KB was populated with the most common actions and components. The KB is initiated with 31 actions, 34 synonyms and 46 components. Then, in the first round, the WO were processed automatically by the RCM4Assets AI core, which produces two sets, R_1 and G_1. 15 WOs from R_1 were tweaked and annotated by the users, and the KB was updated. The next round starts with the updated KB and given input of $R_1 \cup G_1$ that, after passing RCM4Assets AI core will produce R_2 and G_2. This process is repeated five times. The number of edited WO was more significant at the beginning and end, ranging in the interval $]10,\ 15]$. The results are presented in TABLE I.

As we can see, in many cases, a slight increase in manual work translates into a significant increase in automatically processed work orders. In fact, $|W^t|$ and $|R|$ are negatively correlated (-0.877), indicating that. Fig. 4 depicts such a relation, where it becomes clear the initial annotations have a higher impact than later annotations.

A. Discussion

From the data in TABLE I and Fig. 4, it becomes clear that the initial annotation effort is the most effective way to unblock the automatic processing of many WO. The ratio is the highest — approximately one annotation enables 14 WO to be automatically processed. After two rounds, the system seems unable to properly amplify the annotation effort,

TABLE I: Evaluation of the value of each WO annotation for automatic service classification.

Round	$\lvert W^t \rvert$	$\lvert R \rvert$	$\lvert G \rvert$	**Ratio**
Round_0	0	663	337	—
Round_1	15	459	541	1:14
Round_3	25	422	578	1:3
Round_4	35	412	588	1:1
Round_5	50	383	617	1:2

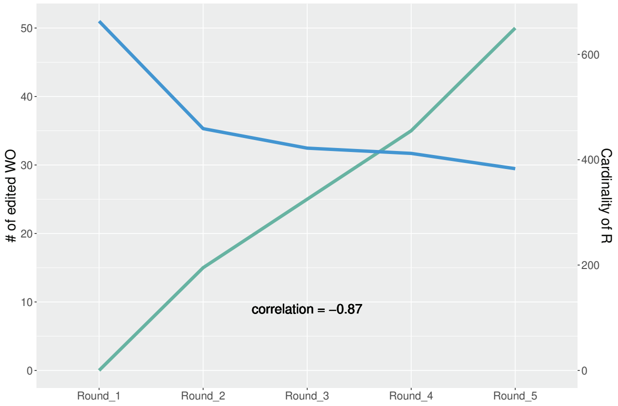


Fig. 4: Relation between the number of WOs edited (left Y-axis, Green line) and the size of set R (right Y-axis, blue line). The correlation between the two is -0.87

achieving a 1:1 ratio, which is undesirable. A closer look at these apparent complicated WO let us conclude: (i) 10% of the WOs (110) cannot be processed automatically since they are from regular maintenance or inspections tasks that need manual processing; (ii) around 20% of the WOs suffer from different problems, including the absence of actions and/or component information, not related information to the WO (e.g. location); (iii) the order of appearance of the WO in the interface may influence the introduction of less functional annotations since the users do not know which are the most effective ones. Together, these groups of WO represent around 30% of the WO $(273$ of $1\ 000)$ that will not be automatically processed using the current algorithm.

Nevertheless, there are still 10% of the remainder of WO that, after five rounds, are not automatically processed. It is challenging to generalise those WOs because they are precise. Despite the stated problems, the evaluation shows the proposed solution can be effective on the large subset of WO that can be automated. The identified issues can be addressed by: (i) changing the interface of the technician's apps, giving them feedback on the "correctness" of the terms they use to describe the WO, and providing them instant feedback on the automation that is achievable by the set of words used; (ii) change the order in which the WOs appear to the domain experts (Fig. 2.B).

V. Conclusion

The use of free text in WOs poses a challenge to asset management automation. However, NLP techniques can process WOs and identify relevant sentence parts for action and component identification. This paper has presented the RCM4Assets solution, which uses a semi-automatic, human-in-the-loop process to produce a normalised and reduced set of actions and components that resumes WO. The solution can reduce the manual work involved in WO analysis, increasing the efficiency of WO processing. The results show an amplification of the manual annotations in the first couple of interactions, up to 1:14 ratio, reducing the number of WO that need manual processing. RCM4Assets represents a significant step forward in automating asset management and can potentially improve the efficiency and effectiveness of maintenance operations. For future work, when presenting users, the WO should consider the ones that have the potential to unblock most of the automation. Also, WO with equal pairs should not appear together, as it would force users to move to the next page seeking different ones. Furthermore, the UI used by technicians to fulfil WOs should include NLP solutions, especially using the RCM4Assets AI Core.

References

[1] A. Crespo Marquez, J. F. Gomez Fernandez, P. Martínez-Galán Fernández, and A. Guillen Lopez, "Maintenance management through intelligent asset management platforms (iamp). emerging factors, key impact areas and data models," *Energies*, vol. 13, no. 15, p. 3762, 2020.

[2] P. Santos, M. Pato, N. Datia, J. Sobral, N. Leitão, M. Ramos Ferreira, and N. Gomes, "Recommending words using a bayesian network," *Electronics*, vol. 12, no. 10, p. 2218, 2023.

[3] T. Sexton, M. P. Brundage, M. Hoffman, and K. C. Morris, "Hybrid datafication of maintenance logs from ai-assisted human tags," in *2017 ieee international conference on big data (big data)*, pp. 1769–1777, IEEE, 2017.

[4] P. Santos, N. Datia, M. Pato, and J. Sobral, "Comparing word embeddings through visualisation," in *2022 26th International Conference Information Visualisation (IV)*, pp. 91–97, 2022.

[5] T. Le and H. David Jeong, "Nlp-based approach to semantic classification of heterogeneous transportation asset data terminology," *Journal of Computing in Civil Engineering*, vol. 31, no. 6, p. 04017057, 2017.

[6] J. Thorne and A. Vlachos, "Automated fact checking: Task formulations, methods and future directions," *arXiv preprint arXiv:1806.07687*, 2018.

[7] N. Melluso, I. Grangel-González, and G. Fantoni, "Enhancing industry 4.0 standards interoperability via knowledge graphs with natural language processing," *Computers in Industry*, vol. 140, p. 103676, 2022.

[8] Y. Kang, Z. Cai, C.-W. Tan, Q. Huang, and H. Liu, "Natural language processing (nlp) in management research: A literature review," *Journal of Management Analytics*, vol. 7, no. 2, pp. 139–172, 2020.

[9] R. Sivarethinamohan, S. Sujatha, and P. Biswas, "Envisioning the potential of natural language processing (nlp) in health care management," in *2021 7th International Engineering Conference "Research & Innovation amid Global Pandemic"(IEC)*, pp. 189–193, IEEE, 2021.

[10] M. C. May, J. Neidhöfer, T. Körner, L. Schäfer, and G. Lanza, "Applying natural language processing in manufacturing," *Procedia CIRP*, vol. 115, pp. 184–189, 2022.

[11] P. Santos, N. Datia, M. Pato, and J. Sobral, "Suggesting words using a bayesian network," in *2022 Inforum*, 2022.

[12] M. Payette, G. Abdul-Nour, T. J.-M. Meango, and A. Côté, "Improving maintenance data quality: Application of natural language processing to asset management," in *16th WCEAM Proceedings*, pp. 582–589, Springer, 2023.

[13] Q. Zhang, Q. Yuan, P. Lv, M. Zhang, and L. Lv, "Research on medical text classification based on improved capsule network," *Electronics*, vol. 11, no. 14, p. 2229, 2022.

[14] O. Obeid, S. Khalifa, N. Habash, H. Bouamor, W. Zaghouani, and K. Oflazer, "Madari: A web interface for joint arabic morphological annotation and spelling correction," *arXiv preprint arXiv:1808.08392*, 2018.

[15] J. Lafferty, A. McCallum, and F. C. Pereira, "Conditional random fields: Probabilistic models for segmenting and labeling sequence data," 2001.

[16] H. H. Olsson and J. Bosch, "Towards continuous customer validation: A conceptual model for combining qualitative customer feedback with quantitative customer observation," in *Software Business: 6th International Conference, ICSOB 2015, Braga, Portugal, June 10-12, 2015, Proceedings 6*, pp. 154–166, Springer, 2015.

[17] N. Datia, J. Moura Pires, and N. Correia, "Time and space for segmenting personal photo sets," *Multimedia Tools and Applications*, vol. 76, pp. 7141–7173, 2017.