PO.001

# Value from free-text maintenance records: converting wind farm work orders into quantifiable, actionable information using text mining

**Erik Salo, Dr David McMillan, Prof Richard Connor**
University of Strathclyde

## Abstract

The aim of this project is to demonstrate how text mining can help wind farm operators extract unique, quantifiable maintenance information from historic work orders. A good overview of past maintenance efforts can help develop an reliability-centred maintenance strategy for the future in terms of labour intensity, budgeting and spare parts logistics [1, 2]. However, work orders - where significant information is entered by a human in the form of free text – do not provide any straightforward means for automated analysis [3, 4].

Our approach introduces a novel combination of machine learning techniques supported by expert judgement. Significant focus is on the vocabulary - spelling error correction, semantic matching of synonyms and abbreviations. This allows tasks to be grouped by their underlying meaning, not only the characters they contain. The principal output is a frequency distribution of all groups of equivalent tasks. Further categorical analysis allows to focus on specific plant systems or components, as well as failure modes.

Data from an industrial partner's major onshore wind farms in Scotland was used to test our approach against manual analysis. Potential savings were identified in weeks of effort, or £2-9k in labour cost per site, in addition to an improved maintenance strategy. The remaining challenges mainly lie in increasing accuracy and reducing operator input. These are being addressed by our continued research, but also highlight opportunities for collaboration and standardisation across the industry to maximise the value of data.

## Objectives

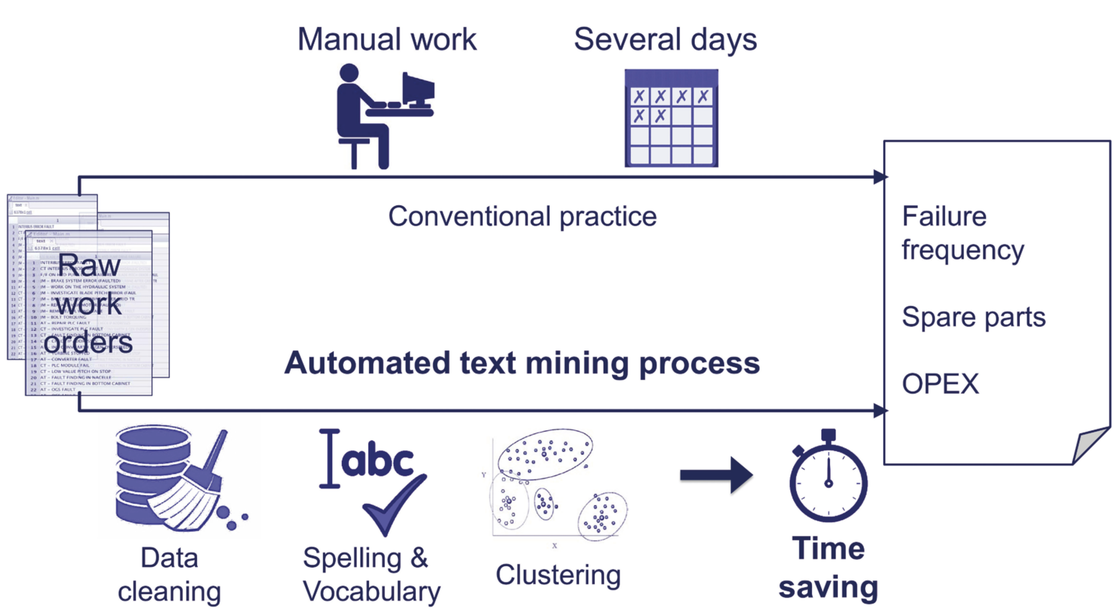


Figure 1. Automation helps to save operator time in work order analysis

This project aimed to develop a proof-of-concept text mining methodology which would simplify the analysis of historical work orders, as illustrated in Figure 1. Two main goals were set in cooperation with the industrial partner:

1) Actionable outputs: work orders grouped according to maintenance task, frequency provided for each group, top 40 presented on Pareto chart.
2) Labour time reduction so that a fleet of hundreds of turbines could be analysed with reasonable effort.

## Methods

Input data was obtained in the form of Excel tables from the SAP platform and imported into Matlab without prior treatment. Corrective maintenance tasks were then filtered for further analysis. Scripts for importing and filtering data were adapted from [3].

The main stages of the methodology shown in Figure 2 and candidate methods for each stage - such as specific clustering algorithms - were established based on literature, including [5, 6]. The final combination of methods was determined heuristically during extensive testing. The program was written in Matlab.

## Results

Figure 2 presents the final methodology. As in [3], two main factors limited the choice and effectiveness of text mining methods: relatively small dataset size for statistical inference, and a high proportion of noisy data. As a result, some steps were left aside after testing, including synonym recognition. Manual synonym mapping was a viable alternative at the given dataset size; in the longer term, improving data collection to achieve higher data quality would be recommended [4].

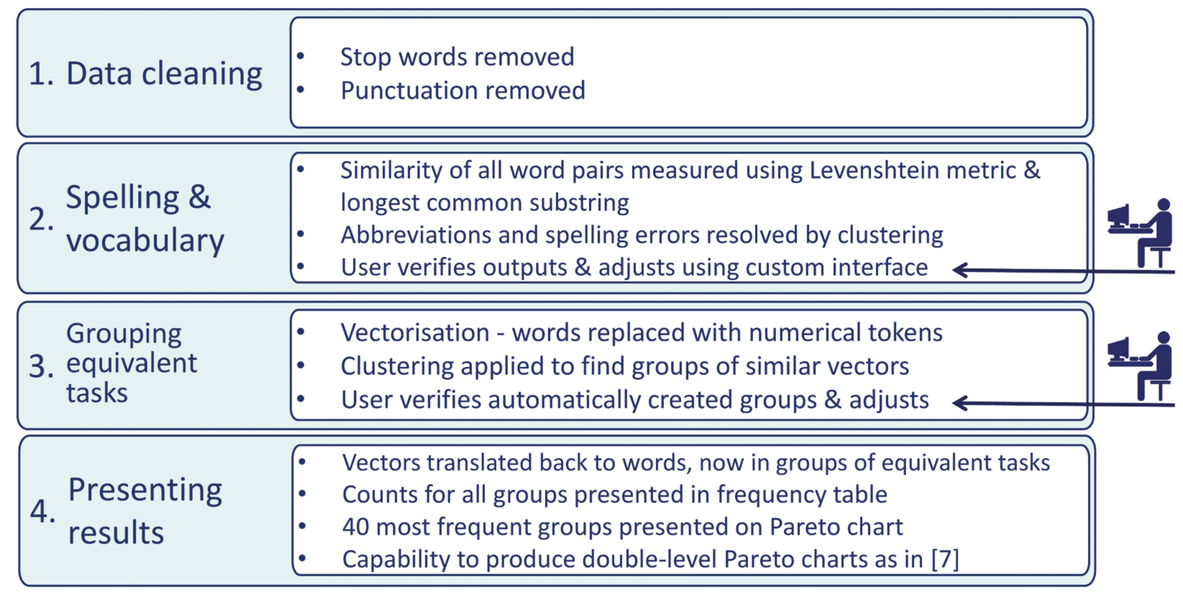


Figure 2. Final methodology (The operator symbol indicates the remaining manual inputs)

Figure 3 demonstrates a time reduction in a dataset of 3400 records from 10 working days to around one day. Any remaining adjustments were greatly simplified owing to dimensionality reduction. The industrial partner valued such analysis at roughly £2000; larger datasets may offer more savings [3]. The effort could be further reduced in the future by standardising the terminology and turbine taxonomy [1, 4], such as in [7].

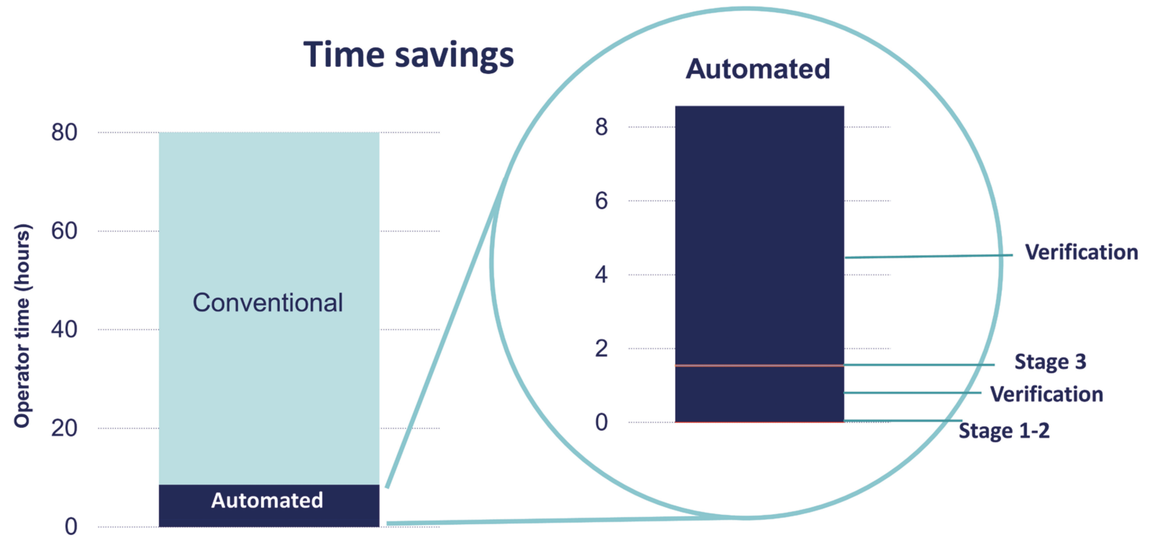


Figure 3. Time savings achieved on a test dataset

## Conclusions

The methodology developed in this project was able to automate the most mundane parts of manual analysis, and was therefore considered a success by the industrial partner. The remaining level of user input was found acceptable.

Input in the form of expert judgement is an important part of analysing maintenance records. The use of supervised machine learning for more efficient use of operator input is proposed as an opportunity for further research.

The development process highlighted a cooperation opportunity to standardise task nomenclature and data collection across the wind energy industry, with benefits ranging from reliability-centred maintenance to life extension.

## References

1. Mcmillan, David and Ault, Graham (2012). Towards reliability centred maintenance of wind turbines. *PMAPS 2012*
2. Hodkiewicz, Melinda and Tien-Wei Ho, Mark (2016). Cleaning historical maintenance work order data for reliability analysis. *Journal of Quality in Maintenance Engineering, Vol. 22 Issue: 2, pp.146-163*
3. Salo, Erik (2017). Analysis of SAP work order data by turbine technology type for onshore wind. *MSc thesis, University of Strathclyde*
4. Letcher, Trevor M. (2017). *Wind Energy Engineering. A Handbook for Onshore and Offshore Wind Turbines*
5. Connor, Richard et al. (2011). A bounded distance metric for comparing tree structure. *Information Systems, Vol. 36 Issue 4, pp: 748-764*
6. Banchs, Rafael E. (2013). Text mining with MATLAB
7. Bueno Gayo, Juan (2011). Final Publishable Summary of Results of Project ReliaWind