



HAL
open science

KNOWO: A Tool for Generation of Semantic Knowledge Graphs from Maintenance Workorders Data

Farhad Ameri, Renita Tahsin

► **To cite this version:**

Farhad Ameri, Renita Tahsin. KNOWO: A Tool for Generation of Semantic Knowledge Graphs from Maintenance Workorders Data. IFIP International Conference on Advances in Production Management Systems (APMS), Sep 2022, Gyeongju, South Korea. pp.188-195, 10.1007/978-3-031-16411-8_24 . hal-04768921

HAL Id: hal-04768921

<https://inria.hal.science/hal-04768921v1>

Submitted on 6 Nov 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License



This document is the original author manuscript of a paper submitted to an IFIP conference proceedings or other IFIP publication by Springer Nature. As such, there may be some differences in the official published version of the paper. Such differences, if any, are usually due to reformatting during preparation for publication or minor corrections made by the author(s) during final proofreading of the publication manuscript.

KNOWO: A Tool for Generation of Semantic Knowledge Graphs from Maintenance Workorders Data

Farhad Ameri, Renita Tahsin

Engineering Informatics Lab, Texas State University, San Marcos, U.S.A
{ameri, r_t298}@txstate.edu

Abstract. A major portion of industrial maintenance data is in unstructured form, which makes its organization, search, and reuse very challenging. For this reason, the knowledge embedded in historical maintenance data is seldom analyzed or reused for purposes such as root cause analysis, failure prevention, and maintenance diagnostics. If the valuable knowledge patterns nested in maintenance data are identified, liberated, and formalized, they can significantly improve the intelligence of maintenance management systems by providing actionable insights. The objective of this research is to help advance the progression from data to information and knowledge through data-driven creation of a public and open-source knowledge graphs built from the textual data available in maintenance workorders. A SKOS-based thesaurus is used to support automated entity extraction from the text. A formal OWL-based ontology provides the semantic schema of the knowledge graph. A software tool (KnoWo) is developed to streamline the text-to-graph translation process. It was observed that the proposed text-to-graph tool chain improves knowledge discovery by analyzing maintenance logs.

Keywords: knowledge graph, maintenance workorder, ontology, thesaurus

1 Introduction

Maintenance is defined as the actions intended to retain an asset or to restore it to a state in which it can perform a required function [1]. In order to minimize machine downtime and maximize the availability of critical assets and equipment uptime, manufacturers use a wide range of maintenance management tools and technologies with varying levels of automation. In particular, Computerized Maintenance Management Systems (CMMS) are widely used in most industries to manage, plan, and execute preventive and planned maintenance activities [2]. Despite widespread adoption of maintenance automation solutions in industry, maintenance management is still a highly human-centric process since efficient acquisition, formalization, and reuse of maintenance knowledge is a major challenge [3].

The data generated during maintenance management activities (including identification, planning, scheduling, and execution) is often stored in Maintenance Work Orders (MWO). MWO records are often generated by maintenance technicians and stored in the database of CMMS packages for archiving, reporting, or analysis purposes

[4]. Enormous collections of historical maintenance logs, representing a wealth of diagnostic knowledge, can be found in most industries. The knowledge embedded in MWOs can be used to inform the maintenance diagnosis process. However, the unstructured data provided by MWOs are often under-used since it is not presented in a computable form [1][5]. Additionally, the data is often flooded with misspellings, jargons, abbreviations, and contradictory statements since they are prepared by multiple technicians with varying levels of experience and background knowledge. There is a need for developing formal methods and models for data cleaning and knowledge extraction and formalization to improve the reusability and findability of the knowledge embedded in MWOs. As the size and number of MWO records increases, manual search and analysis of those records becomes more cumbersome and less efficient. Without proper tools and techniques for analyzing, mining, and contextualizing that knowledge, the usefulness of these maintenance logs is severely limited.

Advanced techniques supported by Natural Language Processing (NLP) and Machine Learning (ML) can be applied to extract useful patterns and rules from the raw text that are otherwise hidden in the historical maintenance work order data [6]. These techniques are particularly useful for cleaning maintenance logs, extracting important terms, classifying, and clustering similar workorders, and identifying associativity relationships between the extracted terms. One difficulty in effective analysis of MWO data using NLP is that the size of MWO data is often smaller than what is needed by most NLP/ML tools for training purposes [7]. A major of the existing data-driven methods is that they take the semantics of data for granted which results in with working with semantically fractured and low-quality data that is not properly contextualized.

An alternative solution for learning from MWO data is to convert the raw text into knowledge graphs. A knowledge graph represents data as a collection of nodes (concepts) connected through edges (relationships). Knowledge graphs are quickly becoming the preferred models for data representation since they promise to bridge the gap between the data and its meaning, and they serve as the natural model for data integration thanks to their flexibility and extensibility. There have been some efforts for automated generation of knowledge graphs from MWO data supported by deep learning and the NLP pipeline [8]. Such works are significant achievements in this field, but they still need to be further developed to address some shortcomings, including a lack of alignment with formal ontologies.

The underlying research challenge that motivates this work is to generate more structured and formal knowledge models, including ontologies and knowledge graphs. Addressing this challenge will require using a hybrid approach that incorporates top-down methods for generation and curation of knowledge models by domain experts as well as bottom-up methods for data-driven extension and validation of knowledge models. This work reports the preliminary results of an ongoing research project.

2 Research Method

The proposed approach for the generation of a knowledge graph uses two semantic models, namely, a thesaurus and an ontology. The thesaurus provides lexical semantics, whereas the ontology provides logical semantics. The thesaurus concepts are mapped to the ontology classes to enable semi-automated text translation into knowledge graphs. Description Logic (DL) is used as the knowledge representation formalism in ontology

modeling. The generated knowledge graph is more available computationally compared to natural language text and can be queried or reasoned over to detect latent or recurring patterns in the data. The proposed approach adopts a human-in-the-loop (HITL) strategy since some contextual and common-sense knowledge is needed for workorder text disambiguation and decomposition into a set of interrelated concepts with well-defined semantics. The level of involvement of human experts gradually diminishes as the thesaurus and the ontology becomes more mature and stable.

3 Maintenance Diagnostics Thesaurus (MDT)

The Maintenance Diagnosis Thesaurus (MDT) is a controlled vocabulary for maintenance terms that uses the Simple Knowledge Organization System (SKOS) for its syntax and semantics. The thesaurus is used to facilitate automatic extraction of key maintenance concepts from the workorder text. SKOS is a standard published by World Wide Web Consortium (W3C) that provides a structured framework for building controlled vocabularies such as thesauri, concept schemes, and taxonomies to be used and understood by both human and machine agents. SKOS models are considered to be lightweight ontologies as they do not have the expressivity of heavyweight, axiomatic ontologies such as OWL models. For this reason, their development cost is relatively low, and they can be quickly expanded in a decentralized fashion by various user communities. The MD Thesaurus provides a set of vocabulary often used by technicians when documenting their observations when diagnosing functional failures and maintenance issues. MD Thesaurus concepts are categorized under seven concept groups or schemes, namely, Action, Artifact, Condition, Event, Function, Material Substance, and Property. The lower-level concepts under each scheme are collected through tagging relevant terms in an experimental dataset. The development and curation of the thesaurus is supported by a web-based tool (SKOS Tool) that enables the user to upload the text, tag and classify the terms under appropriate broader concepts in an interactive fashion. The MD Thesaurus can be exported in the RDF-JSON data interchange standard and shared across multiple platforms that are compatible with Semantic Web standards Fig. 1. Fig. 1 shows examples of the concepts under the Condition concept scheme. These concepts are often used by maintenance technicians to describe the observed condition of a maintainable item (machine or equipment). Another utility of the MD thesaurus is to capture the alternative terms that are often used to refer to the same concept. For example, as shown in Fig. 2, *hydraulic leak*, *lube leak*, *hyd leak*, and *oil leak* are synonym terms that are used for labeling the same concept. Using SKOS annotation property, it is possible to select one preferred label for a given concept and assign multiple alternative labels as needed.

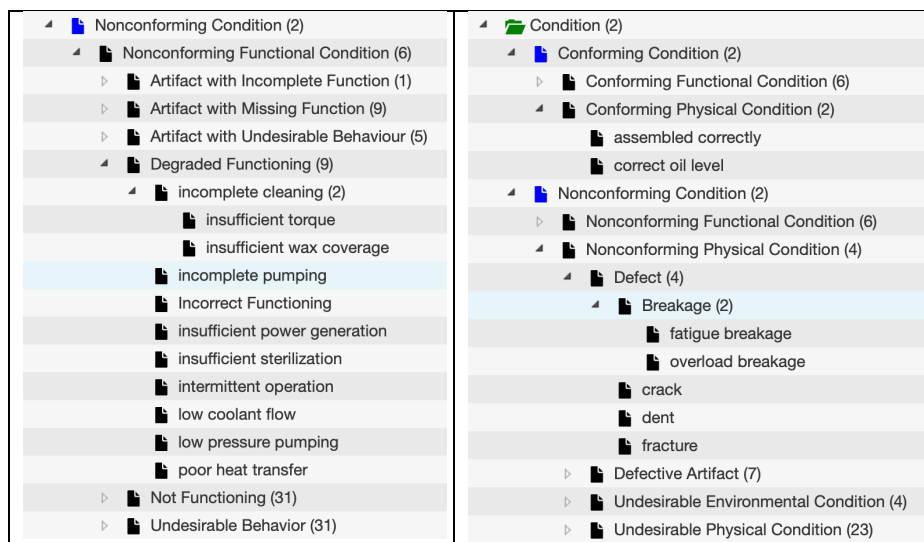


Fig. 1. Examples of concepts under the “Condition” concept scheme.

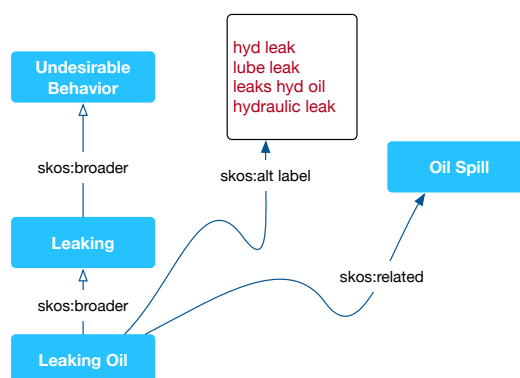


Fig. 2. Broader and Related concepts for Leaking Oil along with its alternative labels

4 Workorder Ontology (WOO)

Workorder Ontology (WOO) is a formal OWL-based ontology that can be used for the formal representation of entities and their relationships in the maintenance domain. WOO uses Basic Formal Ontology (BFO) as its top-level, or foundational, ontology. However, WOO only uses a sub-set of BFO terms and, therefore, it does not import BFO entirely. Using a top-level ontology facilitates ontology reuse and also provides a logical framework for ontology development that is consistent with established philosophical theories. Each class in WOO has a natural language and formal definition to enable unambiguous human-to-human and machine-to-machine communication.

selected workorder (the third table in Fig. 4- left: user-defined concepts). All concepts in the thesaurus are mapped to ontological classes. Therefore, each detected or user-defined concept represents an instantiation of a class in WOO ontology. For example, *blown o-ring* is an instance of *Defective Artifact* class, and *leaking oil* is an instance of *Undesirable Behavior* class. The user can select the individuals that need to be transferred to the next step, which is about connecting the selected individuals using ontological relationships.

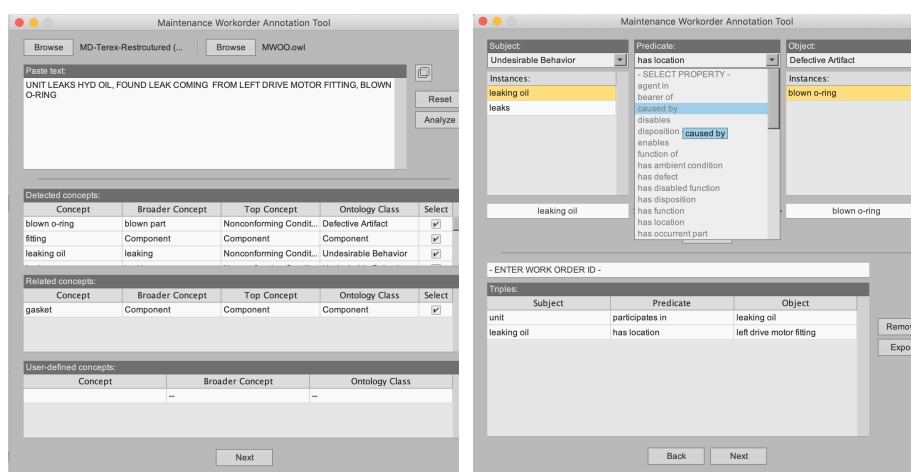


Fig. 4. KNOWO interface for the first step: concept extraction and ontology instantiation (left), KNOWO interface for the second step: connecting the generated individuals using ontological classes (right)

As can be seen in Fig. 4 - right, the user can select the instances in the domain and range (subject and object) of each RDF triples and connect them using the appropriate property.

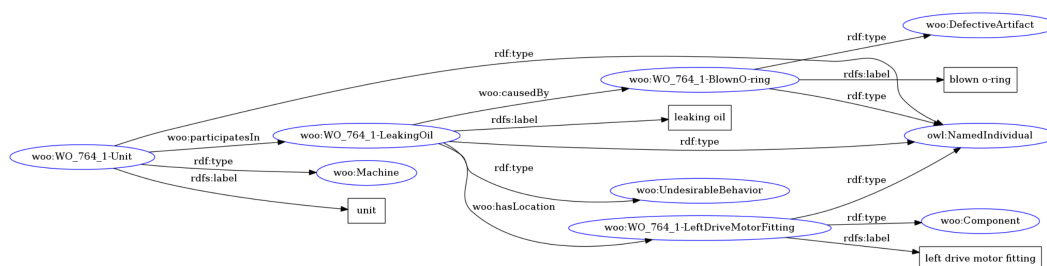


Fig. 5. The final knowledge graph generated automatically based on the exported triples

For example, the user can specify that leaking oil *is caused by* blown o-ring. Once all necessary triples are built, the tool exports the final set of triples in turtle (.ttl) format,

a file format for expressing data in the Resource Description Framework (RDF) data model. The .ttl file can then be visualized using general-purpose RDF visualizers such as RDF Grapher. The output of the RDF Grapher for the selected example is shown in Fig. 5.

6 Validation

The standard method for validating ontologies and knowledge graphs is to test them with respect to their ability in responding to a set of competency questions. These questions are often determined at the early stages of the ontology development process to serve as a set of requirements for the ontology. Some of the competency questions for WOO ontology are listed below:

- What are the causes of different undesirable behaviors of given equipment?
- What are the observed maintenance states of a given machine?
- What is the most frequent type of defect that has caused failure events?
- Which artifacts (machine or equipment) have demonstrated degraded functioning?
- What are the functional units that have been in a defunct state at some time?

A test knowledge graph based on 100 work orders extracted from the CMMS of a construction equipment manufacturer was used during the validation step. All competency questions were formulated as SPARQL queries and executed against the test graph, and it was confirmed that the graph could correctly resolve those queries. The SPARQL query and the retrieved results related to the last query are provided in Fig. 6.

PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>	Functional Unit	Defunct State
PREFIX owl: <http://www.w3.org/2002/07/owl#>	Winch	Not Holding Load
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>	Generator	Not Working
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>	Pump	Not Functioning
PREFIX woo: <http://infoneer.txstate.edu/ontology/MWOO/>	Fuel Pump	Inoperative
SELECT ?item ?state	Engine	Not Functioning
WHERE {	Elect Motor	Not Turning On
?item woo:hasState ?state.	Control Box	Not Operating
?state a woo:DefunctState.	Seat Belt	Not Latching
?item a woo:FunctionalUnit.	ECM	Not Displaying
}		

Fig. 6. The example SPARQL query that returns all Functional Units that have participated in some Defunct State at some time

7 Conclusion

In this paper, a human-assisted method for the creation of RDF knowledge graphs from MWO data was proposed. A java-based tool called KnoWo, was developed to

facilitate the knowledge graph generation process. The main functions of KnoWo include extracting key concepts from MWO text (based on a list of known terms provided by a SKOS thesaurus), instantiating ontology classes according to the extracted concepts, and relating the generated instances using ontological relationships.

This work particularly focused on formalizing the data related to observed symptoms and probable causes at the time of failure. However, a major component of MWO data is related to the action taken to restore an asset. In the future, the ontology and thesaurus will be extended to capture the maintenance treatment and link them to the failures and symptoms. In this way, the graph can be searched to retrieve potential solutions for different types of failures.

References:

- [1] Melinda, H., and Mark Tien-Wei, H., 2016, "Cleaning Historical Maintenance Work Order Data for Reliability Analysis," *Journal of Quality in Maintenance Engineering*, **22**(2), pp. 146–163.
- [2] Wienker, M., Henderson, K., and Volkerts, J., 2016, "The Computerized Maintenance Management System an Essential Tool for World Class Maintenance," *Procedia Engineering*, **138**(Supplement C), pp. 413–420.
- [3] Sharp, M., Sexton, T., and Brundage, M. P., 2017, "Toward Semi-Autonomous Information Extraction for Unstructured Maintenance Data in Root Cause Analysis," *Advances in Production Management Systems. IFIP WG 5.7 International Conference, APMS 2017, Hamburg, Germany, September 3-7, 2017, Proceedings, Part I*, H. Lödding, R. Riedel, K.-D. Thoben, G. von Cieminski, and D. Kiritsis, eds., Springer International Publishing, Cham, pp. 425–432.
- [4] Lopes, I., Senra, P., Vilarinho, S., Sá, V., Teixeira, C., Lopes, J., Alves, A., Oliveira, J. A., and Figueiredo, M., 2016, "Requirements Specification of a Computerized Maintenance Management System – A Case Study," *Procedia CIRP*, **52**(Supplement C), pp. 268–273.
- [5] "Data-Driven Methods for Predictive Maintenance of Industrial Equipment: A Survey | IEEE Journals & Magazine | IEEE Xplore" [Online]. Available: https://ieeexplore.ieee.org/abstract/document/8707108?casa_token=_Z-I6jlA8M8AAAAA:sebswuj-64yjhVqWv4yLEWgJHpJJZkd9kcCwOhh0r_z4Y6A2tjrCTOjajJDr8KkQZJLcFIntkkA. [Accessed: 26-Jun-2022].
- [6] Lukens, S., Naik, M., Saetia, K., and Hu, X., 2019, "Best Practices Framework for Improving Maintenance Data Quality to Enable Asset Performance Analytics," *Annual Conference of the Phm Society*.
- [7] Sexton, T., Hodkiewicz, M., Brundage, M. P., and Smoker, T., 2018, "Benchmarking for Keyword Extraction Methodologies in Maintenance Work Orders."
- [8] Stewart, M., and Liu, W., 2020, "Seq2KG: An End-to-End Neural Model for Domain Agnostic Knowledge Graph (Not Text Graph) Construction from Text," *KR*, **17**(1), pp. 748–757.