



**HAL**  
open science

# A Thesaurus-Guided Method for Smart Manufacturing Diagnostics

Farhad Ameri, Reid Yoder

► **To cite this version:**

Farhad Ameri, Reid Yoder. A Thesaurus-Guided Method for Smart Manufacturing Diagnostics. IFIP International Conference on Advances in Production Management Systems (APMS), Sep 2019, Austin, TX, United States. pp.722-729, 10.1007/978-3-030-30000-5\_88 . hal-02419266

**HAL Id: hal-02419266**

**<https://inria.hal.science/hal-02419266v1>**

Submitted on 19 Dec 2019

**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

# A Thesaurus-guided Method for Smart Manufacturing Diagnostics

Farhad Ameri, Reid Yoder

Engineering Informatics Group, Texas State University, San Marcos, U.S.A  
{ameri, rjy15}@txstate.edu

**Abstract.** The unstructured historical data available in the databases of Computerized Maintenance Management Systems represents a wealth of diagnostic knowledge. In this paper, a methodology for converting the maintenance log data into formal knowledge graphs is presented. The methodology uses text analytics techniques, in combination with human-assisted thesaurus development methods, for generating a formal thesaurus, or knowledge graph, that encodes the semantic relationships between multiple maintenance entities. The knowledge graph proposed in this work uses Simple Knowledge Organization System (SKOS) standard. A java-based tool is developed that uses the generated knowledge graph as the input and conducts light-weight reasoning to support smart maintenance diagnosis

**Keywords:** smart maintenance, knowledge graph, thesaurus, Natural Language Processing

## 1 Introduction

Manufacturing companies often strive to adopt advanced maintenance and control tools and methods to minimize their machine downtime and maximize the availability of their critical assets. In particular, Computerized Maintenance Management Systems (CMMS) are widely used in most industries to manage, plan, and organize preventive and planned maintenance activities [1]. The records of maintenance work orders and activities are often stored in the partially structured database of CMMS packages for archiving, reporting or analysis purposes [2]. The data in CMMS databases can be potentially used as a source of diagnosis knowledge. However, the maintenance log data is often under-used [3]. Enormous collections of historical maintenance logs, representing a wealth of diagnostic knowledge, can be found in most industries. As the number of reports related to maintenance issues in the CMMS databases grows, manual search and analysis of the reports 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. The underlying research challenge that motivates this work is to generate more structured and formal knowledge models, based on the unstructured and informal data available in the maintenance logs that can support smart and automated diagnosis process.

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. However, most often the intervention of human expert is needed for validating the generated models and taxonomies. In general, full automation of taxonomy and ontology creation process is not a feasible approach to follow based on the current state of Artificial Intelligence technology.

The objective of this work is to introduce a hybrid methodology for generating more structured knowledge models from the unstructured maintenance log data. The methodology uses text analytics techniques, in combination with a human-assisted thesaurus development method, for the purpose of generating a formal thesaurus (i.e., knowledge graph). The resulting knowledge graph is intended to encode the semantic and lexical relationships between various entities in the maintenance domain. The proposed methodology uses Simple Knowledge Organization system (SKOS) formalism for thesaurus modeling and representation [4].

A Java-based tool is developed that uses the generated SKOS thesaurus as the input and conducts root cause analysis and diagnosis based on the observed symptoms in a given maintenance artifact (i.e., part or equipment).

## 2 Related Works

The body of work related to analyzing the historical maintenance data is rather sparse. Winker [1] developed a rule-based approach for cleaning the historical work order data available in CMMC databases. The objective was to improve the quality of the data in order to ensure that the data is fit and reliable for further analysis. Sharp et al. [5] used multiple Machine Learning and Natural Language Processing techniques, including unsupervised classification method based on Support Vector Machine (SVM), for automated clustering and tagging of maintenance data. Tagging maintenance logs with appropriate label results in creation of more structured and clean data and reduces the ambiguity in the data. The generated tags can also be used as controlled vocabulary. These works address the necessary steps for pre-processing the data and giving it more structure and semantics. However, the final product cannot be regarded as a formal knowledge graph that can support smart maintenance diagnosis.

## 3 Maintenance Diagnosis Thesaurus

As mentioned before, the *Maintenance Diagnosis (MD) Thesaurus* that is developed in this work is based on Simple Knowledge Organization System. The MD thesaurus provides a formal vocabulary of maintenance terms and at the same time serves as a knowledge graph. In this section, a brief introduction to SKOS is provided and the motivations for using SKOS as the thesaurus representation formalism are discussed. Also, the definitions for some of the key concepts of the thesaurus are provided.

### 3.1 Simple Knowledge Organization System (SKOS)

SKOS is a standard data model, published by, World Wide Web Consortium (W3C), that provides a structured framework for creating different types of controlled vocabulary such as thesauri, concept schemes, and taxonomies to be consumed on the web. A SKOS concept is any unit of thought such as an idea, an object, or an event.

Each concept in SKOS has exactly one *preferred label* (skos:prefLabel) and can have multiple *alternative labels* (skos:altLabel). For example, *Rusting* is the alternative label for *Oxidation* as it is used frequently for referring to the same concept. The broader concept of the *Oxidation* is *Chemical Reaction*, while *Homolytic Oxidation* and *Heterolytic Oxidation* are the narrower concepts; meaning that they are more specialized forms of *Oxidation*. The concepts that are *related* to *Oxidation* include *Galvanization*, *Coating*, and *Rust*.

The concepts are made *related* to one another based on some type of semantic connection between them. For example, the reason *coating* is a related concept to *oxidation* is that coating is the method that is typically used for preventing oxidation. The exact type of relation, however, is not specified in the thesaurus. Each SKOS concept can also have a definition provided in plain English or any other natural language. One major advantage of the SKOS thesauri is that they can be extended, enriched, and validated incrementally by community crowds and shared as linked open data due to their open and standard syntax and semantics. A SKOS thesaurus forms the nucleus of a knowledge graph that can be continuously enriched to support various data-driven and knowledge-intensive application such as semantic search and reasoning, text mining, data integration and alignment, and data analytic.

### 3.2 Corpus

The maintenance data provided by three manufacturing companies was used to build the text corpus associated with the thesaurus. The data across three companies contain a total of approximately 10,000 Work Orders (WO) collected over a 10-year period. The logs are often organized into tabular format and include contextual data, such as the Affected Equipment, Description of Problem, Resolution of Problem, Time/Date Work order was issued, Time/Date Work order was completed, and Maintenance Technician Assigned.

### 3.3 Structure of the MD Thesaurus

Developing a SKOS thesaurus typically starts with creating the top-level categories of terms and further populating the lower-level categories by the terms extracted from the corpus. Currently the MD thesaurus has three main *concept schemes* (collections), namely, *Artifact*, *Maintenance Problem*, and *Maintenance Treatment* as defined in this section. The first level concepts under each category are called *top concepts*. Fig. 1 shows the partial view of the top concepts in the MD thesaurus.



**Fig. 1.** A partial view of the concept hierarchy in MD thesaurus

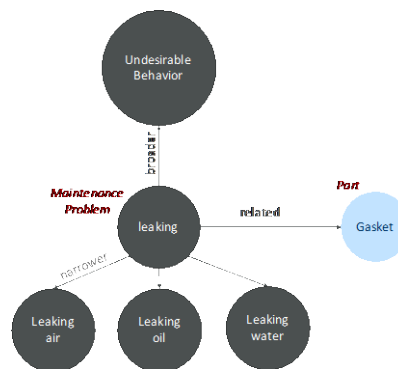
**Artifact** (Concept Scheme) Def. = A physical entity such as individual parts, subsystems, or equipment can be the bearer of a failure

**Maintenance Problem - Failure** (Concept Scheme) Def.= Event in which any part of an equipment or machine does not perform according to its operational specifications or does not possess its desirable qualities

**Functional Maintenance Problem** (Top Concept) Def.= a state in which an asset or system fails to perform a specific function to the desired level of performance.

For example, leaking, as shown in Fig. 2, is considered to be a functional problem as it is considered to be an undesirable behavior of a system

**Non-Functional Maintenance Problem** (Top Concept) Def.= Non-Functional Maintenance Problem occurs when a production equipment (artifact) or one of its parts have one or more undesirable qualities or attributes.



**Fig. 2.** Leaking as an example of a functional problem

**Maintenance Treatment** (Concept Scheme) Def.= A process in which the act is intended to modify or alter some other physical entity with the intention to resolve a maintenance problem. Different types of maintenance treatment include repair, adjustment, replacement, and rebuild.

The top concepts provide the necessary sub-categories under which more specific concepts, extracted from the corpus, can be classified.

### 3.4 ATP Sub-graph

The ATP sub-graph is a sub-graph of the SKOS graph that only contains the *skos:related* relationship between the concepts. The nodes in the ATP sub-graph are of types Artifacts, Problems, or Treatments. The ATP sub-graph describes how artifacts, problems, and treatments are related to each other.

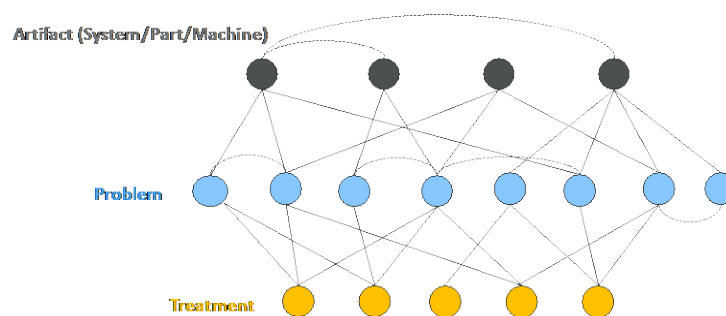


Fig. 3. The structure of the APT sub-graph

### 3.5 Thesaurus Formation and Extension Process

A commercial tool, called PoolParty Taxonomy&Thesaurus Management system [6], was used for creating and extending the MD thesaurus. In PoolParty, each thesaurus can have a document corpus from which the terms can be extracted. In this work, each maintenance record is considered to be a document in the corpus. The MD corpus contained more than 10,000 documents at the time of preparation of this paper. There are two methods for extracting the terms from the corpus and converting them into thesaurus concepts. The first method is based on direct tagging of the relevant terms in the corpus documents. The tagged terms are later added to the collection of *Candidate Concepts*. A candidate concept is formally integrated with the thesaurus when its broader concept is specified by the thesaurus developer. For example, the phrase *broken gear* can be manually tagged in a document by the developer and then placed under *broken part*, as its broader concept. The second method is through automatically extracting a list of n-grams from the corpus and adding the relevant terms to the collection of candidate concepts. In this work, the first method was used.

### 3.6 Thesaurus Validation

The usefulness of the MD thesaurus highly depends on its level of completeness and accuracy. To validate and verify the thesaurus with respect to completeness and accuracy, it was used for *tokenizing* the maintenance records outside the corpus. The

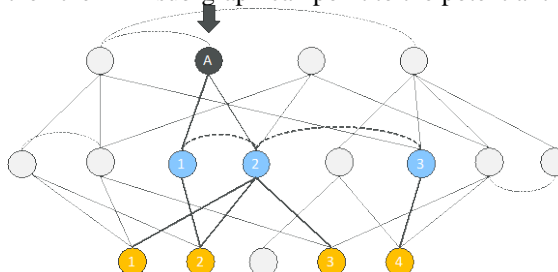
tokenization process entails splitting the text into individual concepts from the thesaurus. A maintenance record is considered to be *adequately tagged* if at least one artifact, one maintenance problem, and one maintenance treatment can be identified in the record. Table 1 shows an example of a maintenance records with complete tags.

**Table 1.** A maintenance record tagged with sufficient tokens

Concept Schema	Top Concept	Concept (preferred label)	Freq.
Artifact	Part	coil	1
Artifact	Part	indexing valve	1
Maintenance Problem	Functional Maintenance Problem	not indexing	1
Maintenance Treatment	Treatment Action	replace	1
Artifact	Part	valve	1
<b>Problem Description: Table will not index.</b>			
<b>Resolution: Table indexing valve coil found bad and replaced</b>			

#### 4 Diagnosis Guided by ATP Sub-graph

The ATP sub-graph can be used for identifying the failures related to a given artifact such as a gear or a pump or a hydraulic system. For example, for artifact A shown in Fig. 4, the related problems are P1, P2, and P3. Since P1 is related to P2 and P2 is related to P3, a loose causality relationship can be inferred. Also, because the links connecting the vertices are bi-directional, it is not possible to determine which problem is the root cause and which problem is the observed effect. But as mentioned before, the goal is to provide a lightweight model with simple semantics to enable some basic reasoning and approximate root cause analysis. For more deterministic and complex reasoning, more expressive ontologies will be needed. If the user picks a specific problem as the probable root cause, then the ATP sub-graph can point to the potential treatments.

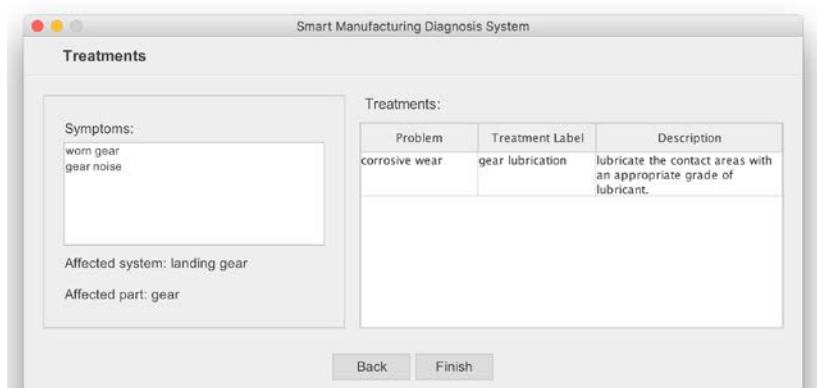


**Fig. 4.** The active portion of the APT sub-graph when a specific artifact is selected.

## 5 Implementation

A Java-based tool, called *Smart Maintenance Diagnosis System (SMDS)* was developed based on the proposed methods. The SMDS tool uses the Apache Jena API for Java. The tool receives the thesaurus in RDF/JSON format as the input. There is a default thesaurus embedded in the tool. However, the user can always upload the most recent version of the thesaurus. Jena is used to parse the thesaurus and to create a map of systems, parts, symptoms, causes, treatments, and the relations between them. The map is then used to analyze and diagnose a selected system or part issue. In the first step, the user provides some contextual information regarding the diagnosis scenario. System type (e.g. landing gear) and part type (e.g. gear) comprise the contextual information in this implementation. Then the user selects the observed symptoms from the drop-down menu. The symptoms are filtered by the tool so that the user can only select from the symptoms (maintenance problems) related to the selected part.

After the symptoms (effects) are selected, then the tool provides the user with a list of potential causes. The potential causes are actually the maintenance problems that are related to the selected symptoms using *skos:related* relationship. Fig. 5 shows the screenshot of the final recommendations of the tool. In the provided example, *corrosive wear* is identified as the potential causes for the selected symptoms, namely, *worn gear* and *gear noise*.



**Fig. 5.** SMDS tool screenshot (4): the tool suggests the treatments based on the selected root causes

As mentioned before, since ATP sub-graph is built based on the observations about the past failures, the resulting conclusions are always approximative based upon the evidence given. The textual description of the recommended treatments are extracted from *skos:definition* property of the concept.



## 6 Conclusion

In this paper, a thesaurus-guided maintenance diagnosis method is proposed. The thesaurus is linked to a text corpus extracted from the CMMS maintenances logs provided by three participating companies. A smart maintenance diagnosis tool was developed and tested based on the proposed method. The test results provided correct diagnosis based on the scope of the thesaurus. The proposed MD thesaurus uses SKOS standard. SKOS knowledge graph provides a solid basis for machine learning and cognitive computing efforts with an organization. SKOS is widely adopted and there exist hundreds of SKOS vocabularies on the web. Therefore, the proposed MD thesaurus can be linked and integrated with other vocabularies to enhance the semantic coverage of the knowledge graph.

One shortcoming of the proposed approach is that it ignores the type of the relations between two concepts and treats all relationships to be the same. This caveat can be countered by superimposing more expressive ontologies on top of the light-weight thesaurus to enable more advanced reasoning. However, developing and extending axiomatic and heavy-weight ontologies can be very costly and time-consuming. Light-weight SKOS thesaurus can be easily developed and extended to support a first-order reasoning process when diagnosing a maintenance problem. In the future, the APT sub-graph will be augmented with probabilistic values to create Bayesian Networks.

### References:

1. Wienker, M., K. Henderson, and J. Volkerts, *The Computerized Maintenance Management System an Essential Tool for World Class Maintenance*. Procedia Engineering, 2016. **138** (Supplement C): p. 413-420.
2. Lopes, I., et al., *Requirements Specification of a Computerized Maintenance Management System – A Case Study*. Procedia CIRP, 2016. **52**(Supplement C): p. 268-273.
3. Melinda, H. and H. Mark Tien-Wei, *Cleaning historical maintenance work order data for reliability analysis*. Journal of Quality in Maintenance Engineering, 2016. **22**(2): p. 146-163.
4. Miles, A. and S. Bechhofer, *SKOS simple knowledge organization system reference*. 2009, W3C.
5. Sharp, M., T. Sexton, and M.P. Brundage, *Toward Semi-autonomous Information*, in *Advances in Production Management Systems. The Path to Intelligent, Collaborative and Sustainable Manufacturing : IFIP WG 5.7 International Conference, APMS 2017, Hamburg, Germany, September 3-7, 2017, Proceedings, Part I*, H. Lödging, et al., Editors. 2017, Springer International Publishing: Cham. p. 425-432.
6. Blumauer, A. *PoolParty Technical White Paper*. 2017; Available from: <https://help.poolparty.biz/doc/white-papers-release-notes/poolparty-technical-white-paper>.