Development of an Ontology for Aerospace Engine Components Degradation in Service

C. Okoh[1], R. Roy[1], J. Mehnen[1], L. Redding[1] and A. Harrison[2]
[1]*EPSRC Centre for Innovative Manufacturing in Through-Life Engineering Services*
Department of Manufacturing and Materials, Cranfield University, MK43 0AL, Cranfield, U.K.
[2]*Rolls Royce, Derby, U.K.*

Keywords: Ontology, Taxonomy, Degradation Mechanisms, Aerospace Engine, Component Degradation, Information Extraction, Knowledge Management, Service Knowledge, Through-life Engineering Service.

Abstract: This paper presents the development of an ontology for component service degradation. In this paper, degradation mechanisms in gas turbine metallic components are used for a case study to explain how a taxonomy within an ontology can be validated. The validation method used in this paper uses an iterative process and sanity checks. Data extracted from on-demand textual information are filtered and grouped into classes of degradation mechanisms. Various concepts are systematically and hierarchically arranged for use in the service maintenance ontology. The allocation of the mechanisms to the AS-IS ontology presents a robust data collection hub. Data integrity is guaranteed when the TO-BE ontology is introduced to analyse processes relative to various failure events. The initial evaluation reveals improvement in the performance of the TO-BE domain ontology based on iterations and updates with recognised mechanisms. The information extracted and collected is required to improve service knowledge and performance feedback which are important for service engineers. Existing research areas such as natural language processing, knowledge management, and information extraction were also examined.

1 INTRODUCTION

In some specific domains events recorded for knowledge capture, sharing and reuse are usually represented in text formats. Information extraction (Wang et al., 2006) is employed to data when seeking to identify and capture the required degradation mechanisms for service knowledge (Doultsinou et al., 2009). The concept of information extraction is essential in the respective domains of health care, energy, power, and aerospace where various events are encountered in the maintenance of machines.

An understanding of knowledge management (KM) (Dadzie et al., 2009; del-Rey-Chamorro et al., 2003), natural language processing (NLP) techniques (Dale et al., 2000), information extraction (IE), taxonomy (Saleem and Bellahsene, 2008), degradation mechanisms (DM) (Okoh et al., 2014) and an ontology (Ahmad and Colomb, 2007; Serra et al., 2013) is needed to improve validated results for better decision making.

The domain corpus is a repository of unstructured and semi-structured information. The task to identify, extract and retrieve the relevant data lies in the domain of natural language processing. The extraction of specific information from natural language is compared with the expected data. Information extraction is part of NLP with the task of extracting entities such as names of persons, locations, and organisations. In this case, Named Entities (NE), Cause and Effect causality ordering approaches are implemented by using the verb cue phrase (Kim et al., 2009). The data are then structured in the ontology.

A taxonomy is a structured arrangement of terms and concepts (Ryu and Choi, 2006). This presents a representation of knowledge with domain specific concepts. In populating the ontology with terms, duplicate words are avoided to eliminate redundancy.

In this paper, the case study focuses on evaluating the validity of the taxonomy of the degradation mechanisms for an existing ontology. Sanity checks were used to manually observe and count the number of identified and captured

Okoh C., Roy R., Mehnen J., Redding L. and Harrison A..
Development of an Ontology for Aerospace Engine Components Degradation in Service.
DOI: 10.5220/0005090201080119
In *Proceedings of the International Conference on Knowledge Engineering and Ontology Development* (KEOD-2014), pages 108-119
ISBN: 978-989-758-049-9

mechanisms. The degradation mechanisms and keywords within issues reported by service engineers describe defects observed during maintenance, repair and overhaul from a through-life engineering services perspective (Roy et al., 2013). This understanding is required to establish the synonyms of the words to be extracted from the corpus (Ryu and Choi, 2006). This work is based on a case study carried out within the aerospace maintenance domain. The contribution is the practical use of the 'AS-IS' and the 'TO-BE' framework to validate and develop an ontology within a service maintenance domain. Sanity checks ascertain the effectiveness of the extractor and show improvement in the performance of the 'TO-BE' ontology.

The remainder of this paper is organised as follows. The background of related fields is discussed in Section 2. The research methodology is described in Section 3. Section 4 presents and discusses the evaluation of the results and Section 5 presents the research conclusions with regard to the benefits of the validation and identification process.

2 BACKGROUND

2.1 Knowledge Management

Knowledge Management (KM) is increasingly being implemented in global engineering and service organisations. Knowledge acquisition, storage, retrieval and interaction are part of KM (Dadzie et al., 2009). It ensures information is secure and well managed (del-Rey-Chamorro et al., 2003) with the purpose of information reuse and sharing. KM is an information system strategy based on insights and experiences of domain experts to gain competitive advantage. KM can be used to develop taxonomies in order to produce and manage ontologies.

2.2 Ontology

An Ontology is an explicit specification of conceptualisation (Gruber, 1993). Recent work has shown the importance of ontology as a problem solving tool of conceptualisation of entities (Maedche and Volz, 2001). The concepts and relations are used to reason and describe domain knowledge. This is a hierarchical arrangement which represents a more natural means of information management in a unique domain.

Ontology assists in developing models of a domain based on reality, perception, conceptualisation, communication and interpretation (see Figure 1). Axioms (reasoning about the meaning) are described by means of asymmetric and intransitive can be related to symmetric, irreflexive and intransitive. It is aimed at capturing specific intended and excluding non-required concepts by reason of conceptualisation (Guarino et al., 2009). It defines terms and relationships inside the domain. Types of ontology include domain, representational, application and generic.

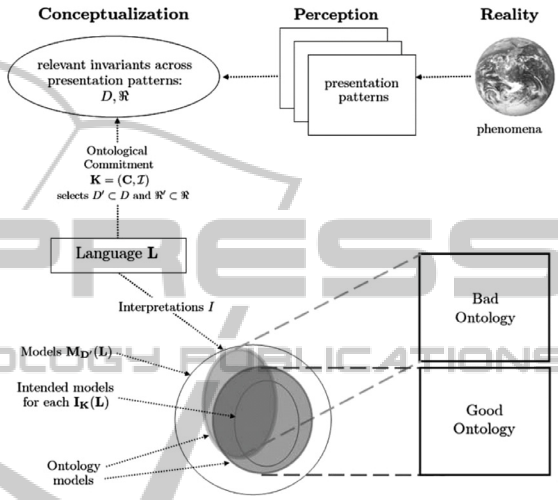


Figure 1: The relationships between perception, conceptualisation and language for communication (Adapted from Source: Guarino et al., 2009).

A primary focus of most ontologies is the taxonomy of classes and subclasses (also called concepts and synonyms) related to different properties (roles) (see Figure 2) which describes the attributes with the role's restrictions defined (Uschold and Gruninger, 2004).

A development process to identify and extract relevant terms or keywords is proposed in this paper. These terms are considered as taxonomy. Taxonomy represents knowledge acquisition of similar words. The taxonomy extracted from a service maintenance context for different degradation mechanisms include Fracture with synonyms as crack, tear and, break. An ontology is a knowledge repository of the taxonomy with inter-relationship of a conceptualisation of terms as illustrated in Figure 2. Both taxonomy and ontology are sometimes used interchangeably. However, taxonomy is often used in industry and ontology is commonly used by academics.

Figure 2 is an extension of the ontology tree showing a simplistic knowledge representation with levels relating to the mechanical component. It also

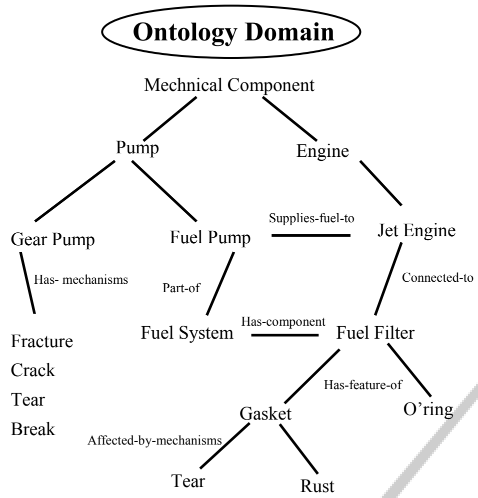


Figure 2: Example of an Ontology (Adapted from Source: (Uschold and Gruninger, 2004).

links the component to the feature and mechanisms with examples of defects which can affect an asset.

Identification of image and shape feature bird classification is based on real-world objects and semantic-based retrieval (Liu et al., 2007). There are critical unchanging properties such as physical bird shapes and characteristics. These are used to manage and model bird classification for knowledge reuse. The ontology provides a similar shared understanding of a specific domain in both humans and computers. It provides a semantic starting point for meaningful definitions. The UNAS, (2000) describes criteria for the design of ontologies relating to the common approaches and visualisation used in ontological engineering. Ahmad and Colomb, (2007) argue that a server development should be determined by considering what the ontology is designed for (e.g. ontologies for business and engineering applications should vary slightly). Jasper and Uschold, (1999) present a framework for understanding and classifying the application of ontologies. Scenarios have been categorised into (1) neutral authoring of a single language, (2) common access to information by more than one person or computer applications and (3) indexing – ontology as a tool for indexing information.

The applicability of ontology structure is the conceptualisation of lexical entries. It represents concepts, the hierarchy and lexical signs for relations and non-taxonomic relations (Maedche and Volz, 2001).

In ontology, maintenance is related to the pruning of the information (Sabou et al., 2005). Unwanted keywords are eliminated in a given domain, whilst refinement, the recognition of the relevant keywords which are not resident in the ontology are removed (Maedche and Volz, 2001). In pruning, the domain relevance is determined by the comparison of the concepts retrieved from a specific domain with the rate of occurrence acquired from a general domain. In refinement, the learning phase enhances the functional capability of the algorithm, so that, unrecognised words can be identified as concepts and conceptual behaviours.

2.3 Information Extraction

Information extraction is one of the important sub-areas of NLP. Named Entities Recognition (NER) recognises named entities from within a phrase or clause or group of sentences. The named entities can be classified based on pre-defined terms such as organisation, person and location. The NER is context-dependent and the extraction is accomplished by string matching if the sentence is incomplete (Wang et al., 2006).

Pre-processing is required in information extraction to accurately retrieve a more complex structure which contains events and relations. The detection and characterisation of semantic relations between entities in the text is applicable to information extraction of relations (Jiang, 2012). Lanfranchi et al., (2007) proposes an extraction and search knowledge for the aerospace industry. Correia et al., (2011) illustrates extracting ontology hierarchies from text by tagging, extraction of candidate classes, identification of hyponyms and synonyms as well as identification and representation of taxonomic relationships.

2.4 Degradation Mechanisms

In the aerospace service domain, a number of factors lead to deterioration of mechanical components. These components become permanently damaged when the threshold is exceeded. Degradation mechanisms relate the causes to the resulting damages (see Figures 3 and 4). For instance, wear caused by poor lubrication, fracture caused by induced vibration of the engine vanes and crack as a result of oxidation. The focus of the case study was to validate the identified synonyms of the taxonomy of wear, corrosion, fracture, and deformation, and to develop an ontology.

Wear is the loss of material as a result of usage over time (Ameen et al., 2011). Lack of lubrication in a rotating engine can lead to wear. Erosion is a form of wear, while cavitation and rub are the causes of wear. Wear may lead to corrosion depending on the operating environments.

Corrosion is a chemical deterioration process leading to material loss. Oxidation and sulphidation can cause components degradation under high thermal stress (Pomeroy, 2005). An example of corrosion is rust caused by oxidation (Figure 3). Oxidation can also result in creep which leads to deformation of the material and eventually causes crack or spallation. Corrosion can be uncontrollable and irreparable.

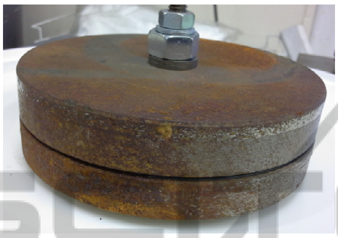


Figure 3: Surface corrosion of metal discs.

Fracture is the result of a separation of material due to cracking or disintegration (see Figure 4). It reduces the functionality of a component. Fractures may occur as a result of chemical effects, shock or stress and increases as strain rate increases. However, deformation happens by reason of Creep which is a slow growth caused by an applied stress. Other types of fracture include crack, tear, burst, peel and split which can either be micro-crack or macro-crack (Medjaher et al., 2012).

(a) (b)

Figure 4: Bearing with fracture (a) outer ring failure and (b) inner ring failure (Source: (Medjaher et al., 2012).

Deformation is the effect of a change in the geometry or shape of a component such as shrinking, stretching, bending, and twisting due to cumulative strain on a component when force is applied. Deformation is either time dependent or time independent mechanisms (Norman, 2013). In Creep deformation the component gradually accumulates over time with the presence of high temperature and thermal cycles stress until the product fails. Elastic deformation results from applied stress on an asset which returns to its original condition when the stress is removed. Plastic deformation occurs when a component exceeds its elastic limit (threshold) and results in a permanent change to the physical structure of the item even when stress is removed.

Degradation mechanisms result from a combination of mechanical, electro-chemical, operational, and environmental conditions. In grouping the identified concepts, an understanding of failure modes and effects analysis (FMEA) as a measure for qualitative analysis is required (Rausand and Høyland, 2004). The FMEA is a procedural method to identify possible failures in a design, an assembly or a manufacturing process, a product or a service. Failure modes are means by which things fail or defects occur and can be potential or actual. Effects analysis examines and helps to understand the consequences of the failures. The aim of the FMEA is to take actions to reduce failures, beginning with the highest-priority failures. Failures can be prioritised by analysing the severity of the consequences, the frequency of occurrence and ease of detection. In this paper, classification of failures is presented as classes and subclasses of knowledge in the domain ontology (Figure 2).

This case study focuses on mechanisms reported by service engineers in the form of text in MS word, Excel, etc. The goal of this validation process is to ensure the mechanisms or damage recorded in the event reports are recognised by the information extraction tool.

Current and future research activities in these areas include evaluation of these applications in large scale datasets assuming an increased requirement for KM (Dadzie et al., 2009) which include the study of different methods of concept mapping and analysis to identify differences between feature combination and integration (Zhang et al., 2011). This includes investigative approaches for automatic mapping (Liu et al., 2007).

3 METHODOLOGY

This paper is based on the validation of taxonomy of the degradation mechanisms which is a list of terms. This is essential in deciding allocation of the mechanisms to create a robust data collection hub. The data collected are required to enrich Service Knowledge and performance feedback to policy makers. The methods used in addressing this work include literature research, observations, and interview with domain experts.

This study shows how to extract concepts for ontology development to aid knowledge sharing and reuse. The information will be pre-processed and filtered from the raw 'on-demand' data sets of textual information. These data contain various keywords (terms and concepts) which will be systematically and hierarchically arranged for use in the engineering service domain.

The data will be grouped into classes of degradation mechanisms which include fracture, wear, corrosion, deformation and causes. The 'AS-IS' and 'TO-BE' framework using an iterative process before, during, after updates and sanity check technique will be implemented.

The case study is essential to enhance Service Knowledge by equipping service engineers with tested, trusted and approved ontology whilst analysing a vast amount of textual failure data. This is important because the relevant failure data will be validated against history information to ascertain through-life performance of the components. This is useful for service engineers when reporting failures encountered in engine maintenance.

This work examines events associated with engine component testing and the means by which the records are processed. A series of trials and processing updates were introduced to a framework to deliver a proposed solution to capture concepts from observed failure. The failure information examined contains the engine name and type, events or issues encountered by the component, year of manufacture and date, and mechanism stating the type of degradation experienced during the test.

An acquisition of a large amount of data from engine information recorded various types of failure modes, mechanisms and types of component feature. The mechanisms and causes of degradation were assessed and analysed in order to understand how the data would be extracted. The FMEA will be used to gain an understanding and assessment of the type of damage present in the textual information. The dataset is a collection of recorded issues observed, evaluated, decided, tracked and closed by service engineers in the maintenance domain. Mechanisms such as bent, shrink, and crack are considered keywords (synonyms). The synonyms of mechanisms are manually identified, extracted from the text and entered onto the ontology.

The taxonomy refers to the synonyms of degradation mechanisms for existing ontology. The procedure to modify the ontology requires an understanding of the process, meaning of the types of damage and under what class it should be. The keywords are uploaded onto the system for a rerun and re-analysis with an embedded recognition tool. This paper shows the method to identify and collect concepts and synonyms using the recognition tool.

3.1 Keywords Grouping

This case study illustrates the process of keyword identification and grouping. The identification process includes:

- Define and seek specific meaning to a degradation mechanism to ensure better understanding of the taxonomy of the degradation and causes (Table 1)
- Attempt to ask and answer questions to ascertain whether the identified keyword is relative to a specified category of degradation mechanisms (Table 3)
- Identify, assess and filter degradation mechanisms based on material loss, separation, geometry change and property maintained in order to predict or determine (using a flowchart) whether the material under investigation is affected by either corrosion, wear, deformation or fracture. The meaning of the mechanisms must be understood (Figure 5)

Table 1: Sample concepts and meanings.

Class	Subclass	Definitions
Deformation Change geometry	Bent	Altered from an originally straight or even condition
Fracture Material separation	Cracked	Damage showing lines on the surface of having split without coming apart.
Wear Material loss	Abraded	Scrape or wear away by friction or erosion
Corrosion Material loss	Blistered	A raised bubble, as on a painted or laminated surface

Table 2: Sample of concepts and questions.

Class	Subclass	Questions
Deformation	Bent	Is the material altered from its original condition?
Fracture	Cracked	Is there a separation within the material?
Wear	Abraded	Is there a scrape on the material?
Corrosion	Blistered	Are there raised bubbles on the material?

As shown in Figure 5, the process starts with observing the issue with the material in terms of loss, change in shape and properties, and questions to classify the defects.

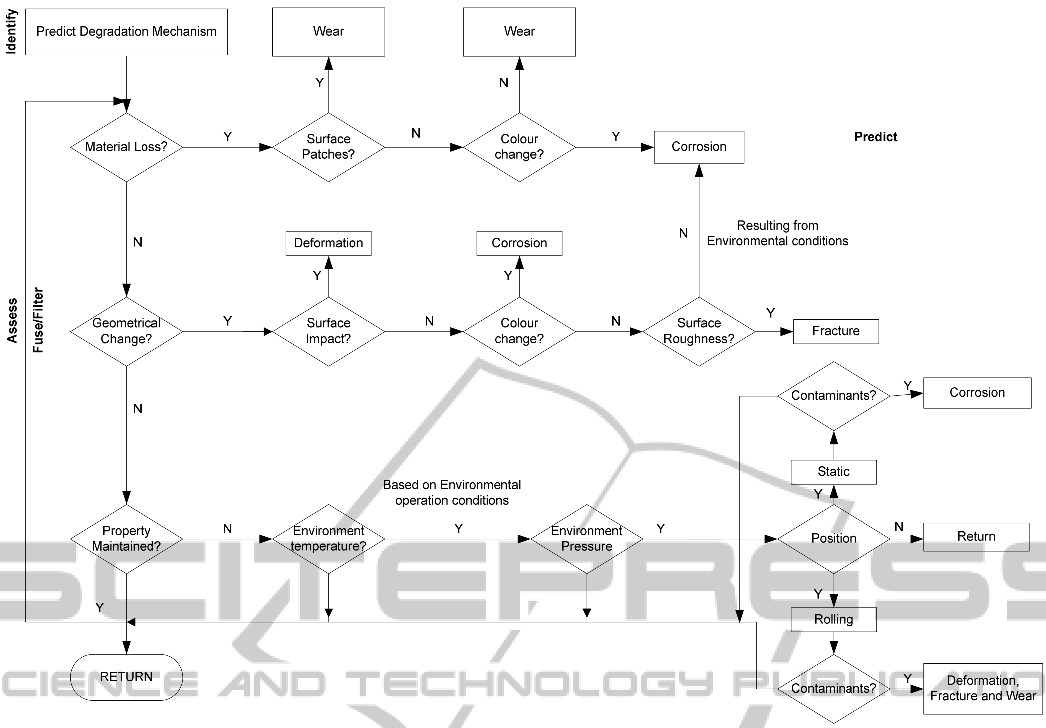


Figure 5: Flowchart to determine degradation mechanisms.

3.2 Risk Matrix with Failure Modes

In order to build a robust ontology that can deliver better performance, various probable states of events and their consequences should be considered. These events are described within the FMEA. Performance of materials is affected by some event which are termed critical. The critical events are differentiated by the keywords used. The keywords are identified and extracted from the FMEA.

To achieve an increased understanding of the concepts and classification within the ontology, risk factors applied to the identified keywords are based on the level of seriousness of the damage and severity of consequences of failure. The severity of the mechanisms results from the understanding of the FMEA procedure in analysing failures. Hence, a relationship was created between the taxonomy of degradation mechanisms and the severity of the failure modes to generate an ontology for problem-solving and decision making.

The identification of potential failure modes on the lowest level of damage and upward hierarchy is a representation of the ontology. The severity of failure modes is classified into minor, major, critical and catastrophic. Minor failure does not degrade the overall performance beyond acceptable limits. Major failure will degrade the system beyond acceptable limits, but can be adequately controlled by alternate means. Critical failure will degrade the component further than the acceptable limits and create a safety threat. A catastrophic failure could result in preventing performance of the intended operations (Rausand and Høyland, 2004).

Physical or chemical processes can lead to events which cause the lowest level of arrangement of failure mechanisms such as deformation, fracture, corrosion and wear represented in Figure 6. Failure rates for each failure mode are recorded. The failure rates are further classified according to frequency of occurrence to give a better understanding of causality for informed decision making (Table 3).

Table 3: Failure rate categorisation (adapted from source: (Rausand and Høyland, 2004).

No	Failure Rate	Occurrence
1	Very unlikely	Once per 1000 years / more
2	Remote	Once per 100 years
3	Occasional	Once per 10 years
4	Probable	Once per year
5	Frequent	Once per month / more often

The Failure rate is different for various operational domains with respect to a failure mode.

The FMEA contains information useful for operation and maintenance. The risk is the severity of the consequences mapped against failure rate.

Table 4: Risk matrix showing different failure modes (adapted from source: (Rausand and Høyland, 2004).

	Severity Categories (Consequences)				
Failure rate	Very unlikely	Remote	Occas-ional	Probable	Frequent
Catastrophic	X				
Critical		X			
Major			X		
Minor					X

As indicated in Table 4, however, a minor damage (e.g a Spall) to a critical component (bearing) can be catastrophic, in that it hinders the performance of the entire system. The failure rate and the severity categories show that the catastrophic failure is very unlikely to occur because the relationship within the ontology is properly developed and utilised by reason of the system, subsystem and component levels Figure 6.

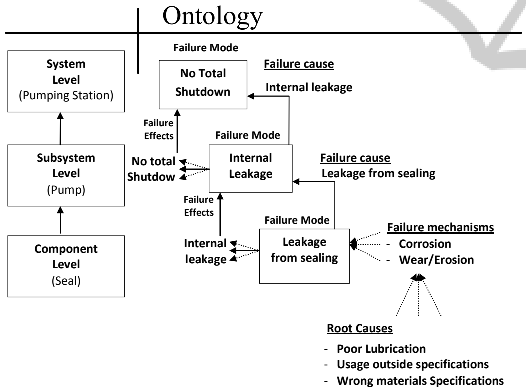


Figure 6: The relationship between failure cause, failure mode and failure effect. (Adapted from Source: (Rausand and Høyland, 2004).

This analysis informs decision making when seeking to consider the choice to either scrap or continue to use the component under investigation. This link is created to assist in detecting failure mechanisms easily based on the approved and agreed threshold. It relates to the use of the monitored operating and maintenance information as inputs to determine through-life performance in terms of remaining useful life of the component under investigation by observing geometry, property loss and material loss (Okoh et al., 2014).

3.3 AS-IS and TO-BE Framework

The 'AS-IS' and 'TO-BE' state is a business process model adapted from (Q-BPM, 2014). It serves as a guide to help understand where we are, where we need to be and how to get there. Applying the framework of this research – it is the current state of the ontology, the desired robust ontology and what to do to get the ontology to the desired robust state. This systematic process is iteratively executed.

The findings are feasible using the proposed 'AS-IS' and 'TO-BE' process model (Figure 7). The model was chosen based on the knowledge of the proposed and agreed solution.

This model is implemented to bridge the gap between 'AS-IS' and 'TO-BE' by way of process improvement. Advantages of the framework include planning, continuous improvement, knowledge retention and learning, process visualization, training, audit and compliance (Q-BPM, 2014).

The original data set (knowledge representation) is the 'AS-IS' which needs to be updated and maintained. The proposed knowledge representation is the 'TO-BE'. The 'AS-IS' model will be updated with the identified entries (new additional concepts) and the results presented - the amount of concepts returned in terms of success rate.

As indicated in Figure 7, each stage addresses a task. The process is planned, to know exactly what to extract and how it should be addressed, check by comparing both current and future states for the taxonomy of degradation mechanisms, then act by agreeing and implementing the results.

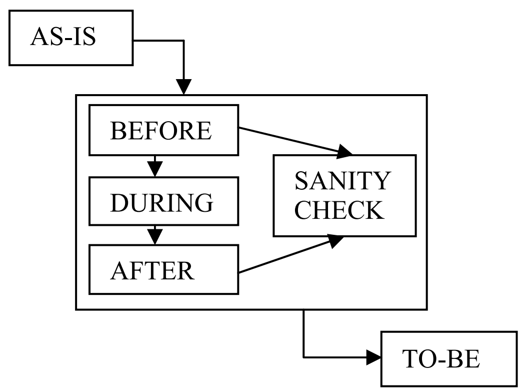


Figure 7: The AS-IS and TO-BE Framework.

3.3.1 Procedure for Analysis

The existing records are event reports presented in Excel. In order to analyse the records, the concepts must be identified. The procedures to analyse information in line with the architecture in Figure 8 are as follows:-

1. *Before Update* - when results and degradation process are initially processed to capture mechanisms (see Figure 9)
2. *During Update* - current state when the results and the degradation process are manually checked to find the number of precise and accurate mechanisms captured.
3. *After Update* – when the results and the degradation processes are checked against event information to identify mechanisms in 'during update'. The concepts which the recognition tool failed to capture are updated within the ontology and then uploaded to take effect for the next 'trial run'. *Note*: The Excel file should be closed and reopened. The recognition tool automatically runs in the background to effectively update changes.

As illustrated in the architecture in Figure 8, the *Corpus* is the application domain in Excel. The *metadata* (information about the identified data), *Concept* (similar or alternate keyword (synonym) in the metadata) to *feature* (the specific data) and *Message* (the selected information to extract from) are represented in the event information section with the related mechanisms / defect types. The mining of data with the recognition tool is done in the *Data Extraction* section. The section returns the results, while the *update* is when the ontology is amended with any newly found concepts.

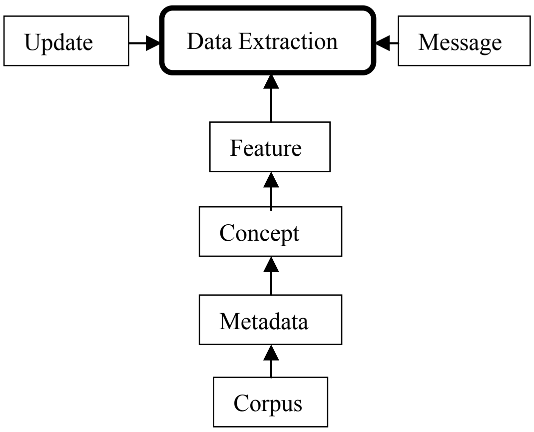


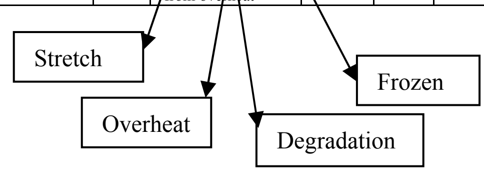
Figure 8: The architecture to extract and analyse data.

3.3.2 Sanity Check Procedure

This sanity check ensures data integrity (Boritz, 2005). The sanity check technique in this context is the manual count of concepts identified, captured and stored as a taxonomy in the ontology. The data extracted from the event information should be accurate and consistent irrespective the number of times the tool is implemented as long as the ontology is updated accordingly with captured concepts. The audit is done on the update section as presented within the architecture in Figure 8.

The sanity check is physically counting the concepts and by running the embedded Recognition tool. The procedures for the technique are as follows:-
1. *TR Right* - when the extracted concepts from the event information are correct. An example is 'fret'.
2. *TR Miss* - refers to the concepts not captured by the recognition tool but are correct. The concepts are identified from failure events and fixed by adding the same onto the ontology e.g 'frozen'.
3. *False Positive* (FP) - when there is an extraction of an incorrect concept in the event information. The fix for the FP is removal of the concept in the taxonomy e.g 'close'.
4. *Human error* - when there is a misspelling of actual concepts. The recognition tool will not identify and capture it. For instance 'luse' instead of 'lose', to fix this, the word 'luse' is added in the ontology. The reason for this is because service representatives report events from different locations in the world and typing mistakes are bound to occur, but it is advisable to train the tool to extract 'luse'.

Mechanism	DP	Event Information	TR Right	TR Miss	FP	HE
	DP	Speed validation, shutdown made causes a frozen of the component A degradation of the rail due to stretch resulting from overheat		4		



* FP: False Positive; HE Human; DP Deterioration process

Figure 9: Manually identified mechanisms.

Concepts TR Missed: The number of concepts which the recognition tool missed frozen, wetting, degradation and overheat. As indicated in Figure 9, the concepts missed are manually identified and captured onto the ontology, while Figure 8 illustrate the identification process.

Concepts False Positive (FP): The number of concepts which returned as FP is four (4), e.g 'mark, markings, mark and another mark'. These words are not concepts, hence, should not be captured by the recognition tool.

The ontology is updated with the newly identified concepts. The ontology is uploaded and the analysis is repeated. The manually identified concepts were captured by the recognition tool as indicated in Figure 9. The number of mechanisms which appeared in the TR right column means the sanity check rightly identifies the concepts captured by the tool as shown in the first and third columns.

Mechanism	DP	Event Information	TR Right	TR Miss	FP	HE
Frozen Degradation Stretch Overheat	DP	Speed validation, shutdown made causes a frozen of the component A degradation of the rail due to stretch resulting from overheat	4			

Concepts found

* FP: False Positive; HE Human; DP Deterioration process

Figure 10: Analysed data with identified mechanisms.

As indicated in Figure 10, the analysed data returned the expected concepts. This is feasible because the tool had been trained to identify and capture the new concepts. That is, after update and run of the ontology, the concepts found.

4 RESULTS

A case study in the area of taxonomy of degradation mechanisms allows for initial evaluation of the effectiveness of the process. Maintenance event information was used in this research. The recognition tool developed in Java enhanced the extraction process.

In the first experience, one hundred rows of records were selected and examined. The rows were manually analysed by the researchers, who manually identified the mechanisms. There was a manual comparison with the results found on the information extraction tool.

The results show the current and future states ('AS-IS' and 'TO-BE') of the ontology. While the 'AS-IS' is an ill-structured presentation of keywords anywhere in the ontology module, the 'TO-BE' is a well-structured representation of the taxonomy of the degradation mechanisms within the ontology module. The 'TO-BE' is a proposed and agreed structural arrangement by policy-makers. A high level illustration of the 'TO-BE' is presented in Figure 11 as deformation, wear, fracture and corrosion. Table 5 shows a comprehensive final taxonomy of the 'TO-BE' ontology.

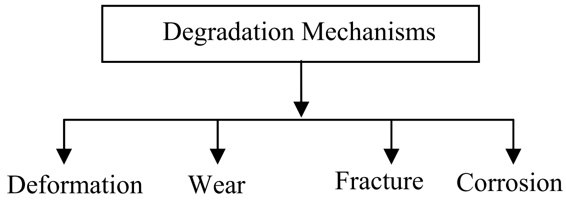


Figure 11: Final Taxonomy of Degradation Mechanisms for the TO-BE Ontology.

The presence of exclusion is needed in certain elimination process. The exclusion is implemented with the symbol '!!'. Two approaches were attempted to fix FP issues: the first is the exclusion of concepts in the ontology module and secondly, classify in another ontology module. Both are good, but in term of clear-cut representation and performance, the latter is better.

A total 2420 event records were analysed. First, the number of concepts 'contained' and 'NOT contained' in the event information is based on the existing 'AS-IS' structure. The event information was interrogated with the extraction tool. The outcome of the number of concepts and the percentage is presented.

At the start of the process, the values were 499 as 'contained' concepts and 1921 as 'NOT contained'. That is a 21 to 79 percentage of the total records. At this stage, the ontology was refined and updated with concepts found in the event information.

To confirm the validity of the process, the first 97 records were selected. 31 rows 'contained' concepts while the 66 rows did 'NOT contained' concepts. The 97 records were 'sanity checked' to ensure concepts were correctly recognised. The outcome was 12 rows were blank with null keywords. Amongst the 85 records analysed, the domain expert identified some concepts which the recognition tool missed and added.

The records were continually iteratively processed by identifying new concepts and updating the ontology. While the amended ontology was uploaded and run, the outcome shows a massive improvement when compared to the results of the initial process. The outcome is based on the initial startup of the system and software. Whereas 892 mechanisms 'contained' were filtered, which is 37.9%, the 1528 mechanisms 'not contained' were observed, which is 63.1%. The total records of rows analysed were 2420.

The researchers used the last 66 records to 'double-check' the validity of the techniques and the effectiveness of the ontology. The number of concepts identified is counted manually before the first run of the ontology. However, based on the 66 records analysed by manually checking and counting, 67 concepts were identified instead of 63, including 4 false positives and 7 false negatives missed by the recognition tool. After the update of the AS-IS ontology, the 7 false negative concepts which resulted in 70 identified concepts.

The results were also compared with adopting a precision, recall and F measurement for performance evaluation of the field of information retrieval (Liu et al., 2007; Dellschaft and Staab, 2006). Also, in ascertaining the extraction performance for learning based on the manual identification regarding Precision and Recall (Sabou et al., 2005).

Precision = mechanisms_found/All_expected
Recall = mechanisms_found/All_mechanisms_found
F = (2 x Precision x Recall)/(Precision + Recall)

where, *mechanisms_found* is the number of keywords returned, *All_expected* is the total expected keywords returned; *All_mechanisms_found* total keywords found. The precision of 94%, recall of 90% and F measurement of 92%.

Ontology pruning and refinement (Maedche and Volz, 2001) are introduced at this stage. Pruning (Sabou et al., 2005), is removing irrelevant concepts in the ontology, that is 34 assumed old concepts which are classified into a different ontology module. The refinement is effective by the upload and run of the TO-BE ontology with the relevant concepts based on the new classification. This refinement accurately and precisely retains the newly identified and the existing concepts regarding the application domain. A total of 42 concepts returned. The success rate is based on the 'After Update' of the agreed 'TO-BE' ontology.

In using the Ontology building support (Maedche and Volz, 2001), the ontology precision model would be

(Returned + found) / (Returned + found + old)

Where *Returned* is the concepts based on agreed 'TO-BE' ontology by the domain expert, *found* is the concepts not considered during the building of the ontology and *old* are the irrelevant concepts

Table 5: Final Taxonomy of the TO-BE Ontology.

	Deterioration process		Deterioration	Degradation				
Chemical deterioration	**Mechanical deterioration**							
Corrosion	**Deformation**	**Fracture**	**Wear**	**Cosmetic deterioration**	**Deterioration cause**	**Location**	**Containment**	**Material property change**
Anti-corrosion	Deformed	Burst	Worn	Blueing	Friction	Fouls	Leak	Brittled
Burnt	Bent	Cracked	Abraded	Polished	Fatigue	De-bond	Drip	Embrittled
Galvanic corrosion	Builtup	Cut	Brinelled	Tarnished	Oscillated	Clashed	Weep	Hardened
Oxidised	Deposited	Perforated	Cavitated	Bruised	Resonated	Contacted	Lost	Softened
Pitted - corrosion	Bulged	Disintegrated	Chaffed	Burnished	Hit	Debonded	Misfilled	Coked
Stress corrosion	Collapsed	Ruptured	Eroded	Stained	Bumped	Ratcheted	Breakout	Glazed
Rusted	Shingled	Snapped	Scrape	Streaked	Banged	Released	Spill	Dealloyed
Sulphidated	Compressed	Divided	Frayed	Discoloured	Wiggled	Separated		Melted
Microbial corrosion	Elongated	Split	Fretted	Discolored	Vibrated	Delaminate		Laquered
Scorched	Extruded	Flaked	Lumped		Strained	Slipped	**Transmission**	Weakened
Thermal erosion	Distorted	Punctured	Galled		Stressed	Displaced	Blocked	Thermal deterioration
	Flattened	Spalled	Picked up		Fire	Pooled	Clogged	Creeped
	Shrunk	Blistered	Roughened		Ingested	Hide	Starvation	Frozen
	Twisted	Peeled	Plucked		Injestion	Dislocated	Short circuit	
	Stretched	Wrecked	Scalloped		Ingress	Misaligned	Jammed	
	Burred	Sheared	Material transfer			Misassembled	Seized	
	Battered	Lifted	Plowed		Damaged	Misclocked		
	Dented	Broken	Exfoliation		Overloaded	Mismatched		
	Depressed	Fragmented	Scuffed		Overspeed	Misfitted		
	Dimpled	Chipped	Rubbed		Overpressure			
	Lapped	Creviced			Overfill			
	Indented	Torn			Contaminated			
	Nicked				Iced			
	Grooved				Overheat			
	Gouged				Bruise			
	Scratched				Corrosion fatigue			
	Scored				Unbalanced			
	Skewed				Foreign object damage (fod)			

removed from the ontology module. From this exercise, it is observed that the higher ratio results in better support for ontology development. The precision reveals 55% of relevant instances retrieved. The result shows a significant pass rate compare with ‘AS-IS’ ontology of concepts. This is dependent on the application domain and the relevant concepts in the corpus is training the software to learn.

5 CONCLUSIONS

This paper demonstrates the verification and validation of the taxonomy of the degradation mechanisms based on the sanity check technique. The analysis of the framework is validated by manual identification, capture and counting of individual concept. The relevance of this taxonomy is to improve service knowledge.

The iterative sanity check technique was useful for the practical task carried out to audit and certify the current ontology. The same technique applies to the classification of newly identified concepts.

The case study focuses on the types of damage observed by service engineers and classified hierarchically in accordance with the predominant degradation mechanisms. The validation process can be used in the audit of information systems.

This research can help other service related application here access to historical information is essential, e.g. predicting system failure and spare parts planning.

The future work within this project involves developing a novel remaining useful life prediction using both current health information and history of a component.

The idea presented in this paper relies on the PLAN, DO, CHECK and ACT (PDCA) business process model (Q-BPM, 2014). The PDCA cycle involves continuous management activities to support decision making. It is an iterative operation observed as sanity checks. Techniques such as a workflow diagram can be used. With constant review and the addition of new degradation mechanisms, the efficiency, effectiveness and performance of the ontology is improved.

This paper discussed how the ‘TO-BE’ ontology structure was developed. Classification is on the basis of the most common or predominant type of degradation experienced by mechanical components. However, a significant difference in results between previous analysis, TO-BE’ ontology and the amount of keywords which were categorised into another ontology module. There is an improvement as the precise concepts captured were retrieved.

Furthermore, compound words (‘fire detector’, ‘fire wire’) can be excluded in the ontology in order to prevent redundancy of concepts. The iterative process this paper can be used in parallel with the ‘AS-IS’ and ‘TO-BE’ framework for effective and efficient execution of tasks in a sequential manner.

ACKNOWLEDGEMENTS

This work is funded by the EPSRC (Engineering and Physical Sciences Research Council, UK), grant number EP/I033246/1 delivered by EPSRC Centre for Innovative Manufacturing in Through-life Engineering Services. The authors also acknowledge Rolls-Royce PLC for their support in this research.

REFERENCES

Ahmad, M. N. and Colomb, R. M., 2007. Managing ontologies: a comparative study of ontology servers, *Proceedings of the eighteenth conference on Australasian database-Volume 63,* Australian Computer Society, Inc., Ballarat, Australia, pp.13-22.

Ameen, H. A., Hassan, K. S. and Mubarak, E. M. M., 2011. Effect of loads, sliding speeds and times on the wear rate for different materials, *American journal of scientific and industrial research,* vol. 2, pp. 99-106.

Boritz, J. E., 2005. IS practitioners' views on core concepts of information integrity, *International Journal of Accounting Information Systems,* vol. 6, no. 4, pp. 260-279.

Correia, J., Girardi, R. and de Faria, C. G., 2011. Extracting Ontology Hierarchies From Text(S), *SEKE, Knowledge Systems Institute Graduate School,* pp. 398-401.

Dadzie, A., Bhagdev, R., Chakravarthy, A., Chapman, S., Iria, J., Lanfranchi, V., Magalhães, J., Petrelli, D. and Ciravegna, F., 2009. Applying semantic web technologies to knowledge sharing in aerospace engineering, *Journal of Intelligent Manufacturing,* vol. 20, no. 5, pp. 611-623.

Dale, R., Moisl, H. and Somers, H., 2000. *Handbook of natural language processing,* CRC Press. Boca Raton, USA.

Dellschaft, K. and Staab, S., 2006. On how to perform a gold standard based evaluation of ontology learning, *The Semantic Web-ISWC 2006,* Springer, Berlin Heidelberg, pp. 228-241.

del-Rey-Chamorro, F. M., Roy, R., van Wegen, B. and Steele, A., 2003. A framework to create key performance indicators for knowledge management solutions, *Journal of Knowledge management,* vol. 7, no. 2, pp. 46-62.

Doultsinou, A., Roy, R., Baxter, D., Gao, J. and Mann, A., 2009. Developing a service knowledge reuse framework for engineering design, *Journal of Engineering Design,* vol. 20, no. 4, pp. 389-411.
Gruber, T. R., 1993. A translation approach to portable ontology specifications, *Knowledge acquisition,* vol. 5, no. 2, pp. 199-220.
Guarino, N., Oberle, D. and Staab, S., 2009. What is an Ontology?, in *Handbook on ontologies,* Springer, Springer, Berlin Heidelberg, pp. 1-17.
Jasper, R. and Uschold, M., 1999. A framework for understanding and classifying ontology applications, *Proceedings 12th Int. Workshop on Knowledge Acquisition, Modelling, and Management KAW,* Stockholm, Sweden, vol. 99, pp. 16-27.
Jiang, J., 2012. Information extraction from text, in *Mining text data,* Springer, Berlin Heidelberg, pp. 11-41.
Kim, S., Aurisicchio, M. and Wallace, K., 2009. Towards automatic causality boundary identification from root cause analysis reports, *Journal of Intelligent Manufacturing,* vol. 20, no. 5, pp. 581-591.
Lanfranchi, V., Bhagdev, R., Chapman, S., Ciravegna, F. and Petrelli, D., 2007. Extracting and searching knowledge for the aerospace industry, *Proceedings of 1st European Semantic Technology Conference,* .
Liu, Y., Zhang, J., Tjondronegoro, D. and Geve, S., 2007. A shape ontology framework for bird classification, *Digital Image Computing Techniques and Applications, 9th Biennial Conference of the Australian Pattern Recognition Society,* IEEE, pp. 478-484.
Maedche, A. and Volz, R., 2001. The ontology extraction & maintenance framework Text-To-Onto, *Proc. Workshop on Integrating Data Mining and Knowledge Management,* California, USA.
Medjaher, K., Tobon-Mejia, D. A. and Zerhouni, N., 2012. Remaining Useful Life Estimation of Critical Components With Application to Bearings, *IEEE Transactions on Reliability,* vol. 61, no. 2, pp. 292-302.
Norman, E. D., 2013. *Mechanical Behavior of Materials,* 4th ed, Pearson, England.
Okoh, C., Roy, R., Mehnen, J. and Redding, L., 2014. Overview of Remaining Useful Life Prediction Techniques in Through-Life Engineering Services, *6th CIRP Conference on Industrial Product-Service Systems (IPSS 2014), Windsor, Ontario, Canada,* 1-2 May.
Pomeroy, M. J., 2005. Coatings for gas turbine materials and long term stability issues, *Materials & Design,* vol. 26, no. 3, pp. 223-231.
Q-BPM, 2014. *AS-IS and TO-BE Process model*, available at: http://en.q-bpm.org/mediawiki/index.php/As-is#Overview (accessed March/18).
Rausand, M. and Høyland, A., 2004. *System reliability theory: models, statistical methods, and applications,* New Jersey, 2nd edition, John Wiley & Sons.
Roy, R., Shaw, A., Erkoyuncu, J. A. and Redding, L., 2013, Through-Life Engineering Services, *Journal of Measurement and Control,* vol. 46, no. 6, pp. 172-175.
Ryu, P. and Choi, K., 2006. Taxonomy learning using term specificity and similarity, *Proceedings of the 2nd Workshop on Ontology Learning and Population: Bridging the Gap between Text and Knowledge,* Syndey, Australia, pp. 41-48.
Sabou, M., Wroe, C., Goble, C. and Mishne, G., 2005. Learning domain ontologies for web service descriptions: an experiment in bioinformatics, *Proceedings of the 14th international conference on World Wide Web,* ACM, Chiba, Japan, pp. 190-198.
Saleem, K. and Bellahsene, Z., 2008. Automatic extraction of structurally coherent mini-taxonomies, *Conceptual Modeling-ER 2008,* Springer, Berlin Heidelberg, pp. 341-354.
Serra, I., Girardi, R. and Novais, P., 2013. PARNT: A Statistic based Approach to Extract Non-Taxonomic Relationships of Ontologies from Text, *Information Technology: New Generations (ITNG), 2013 Tenth International Conference,* Las Vagas, Navada, pp. 561-566,
UNAS, S. M., 2000. Ontological engineering: common approaches and visualisation capabilities, *Informatica,* Vilnius, Lithuania, vol. 11, no. 1, pp. 41-48.
Uschold, M. and Gruninger, M., 2004. Ontologies and semantics for seamless connectivity, *ACM SIGMod Record,* vol. 33, no. 4, pp. 58-64.
Wang, T., Li, Y., Bontcheva, K., Cunningham, H. and Wang, J., 2006. Automatic extraction of hierarchical relations from text, *The Semantic Web: Research and Applications,* Springer, Berlin Heidelberg, pp. 215-229.
Zhang, Z., Gentile, A. L. and Ciravegna, F., 2011. Harnessing different knowledge sources to measure semantic relatedness under a uniform model, *Proceedings of the Conference on Empirical Methods in Natural Language Processing,* Association for Computational Linguistics, pp. 991-1002.