Process Safety and Environmental Protection 177 (2023) 1485–1493

Contents lists available at ScienceDirect

Process Safety and Environmental Protection

journal homepage: www.journals.elsevier.com/process-safety-and-environmental-protection

Towards standardized reporting and failure classification of safety equipment: Semi-automated classification of failure data for safety equipment in the operating phase

Shenae Lee [a,*], Maria Vatshaug Ottermo [a], Stein Hauge [a], Mary Ann Lundteigen [b]

[a] Dept. Software Engineering, Safety and Security, SINTEF Digital, Trondheim, Norway
[b] Dept. Engineering Cybernetics, NTNU, Trondheim, Norway

ARTICLE INFO

Keywords:
Reliability of safety equipment
Failure classification
Reliability data
Maintenance
Operational experience
Technical language processing (TLP)
Entity recognition
Text annotation
Industry 4.0

ABSTRACT

Safety instrumented systems (SISs) are installed on process plants to protect against undesired events like e.g., gas leakage and overpressure. A SIS has reliability requirements that are determined during design, and conformance to these requirements should be verified during operation. It is therefore important that all SIS failures are recorded and classified according to their impact on the SIS reliability. Failures of SIS equipment classified as dangerous undetected are of particular interest because they are dormant (undetected) and will prevent the execution of the safety function (dangerous). Analysis of the failure mode and detection method is essential when deciding if a failure is dangerous and undetected. Such information is often provided as unstructured text in notifications registered into the maintenance management system. Therefore, the work of classifying failures requires considerable manual effort in reading and analyzing the texts. Approaches within natural language processing, like technical language processing, have the potential to be deployed more actively for this purpose. However, successful adoption relies on groundwork where classification rules are derived from international standards and commonly agreed industry practice. This paper presents a semi-automated process that incorporates classification rules and gives examples that indicate some of the capabilities of technical language processing for failure classification. The paper also elaborates on how the work relates to Industry 4.0 in creating digital representations to monitor the performance of safety instrumented systems. This work has been carried out as part of the APOS project (Automated process for follow-up of safety instrumented systems). The APOS project has developed knowledge and specifications that simplify and automate the design and operation of safety equipment and investigated how the failure classification process can be made more efficient.

1. Introduction

Safety instrumented systems (SISs) are installed on process plants to protect against undesired events like e.g., gas leakage and overpressure. Each SIS performs one or more safety instrumented functions (SIFs) whose role is to reduce the accident risk. The SIFs monitor for process deviations during normal operation but perform no functions unless the threshold values are exceeded. The average frequency of demands for SIF execution is typically low, and when demanded less than once per year, the SIFs are classified as operating in the low-demand mode. Failures of the kind that prevent the execution of the SIF once demanded which are not detectable during normal operation are classified as dangerous undetected (DU).

Regular testing of the SIF is essential to reveal DU failures early enough to reduce the risk of accidents. The aggregation of DU failures into updated DU failure rates is carried out to update the estimate of the SIF performance measures by the probability of failure on demand (PFD). The specific requirements for failure data collection, performance measures, operational follow-up, and documentation follow standards like IEC 61508 (IEC 61508, 2010) and IEC 61511 (IEC 61511, 2016). Additional requirements are also specified in national regulations (e.g., the Norwegian Petroleum Safety Authority (PSA) regulations), internal company governing documents as well as facility-specific requirements.

IEC 61508 specifies ranges of PFD into four safety integrity levels (SILs). The SIL (and PDF) requirements of the SIFs are determined by performing risk analysis, while the design of the SIFs (configuration, initial reliability of components) and the operational environment influence the

* Corresponding author.
E-mail address: shenae.lee@sintef.no (S. Lee).

https://doi.org/10.1016/j.psep.2023.07.061
Received 7 May 2023; Received in revised form 6 July 2023; Accepted 18 July 2023
Available online 21 July 2023

Nomenclature

APOS	Automated process for monitoring of safety instrumented systems.
CMMS	Computerised maintenance management system.
DD	Dangerous detected.
DU	Dangerous undetected.
EX	Explosion.
IEC	International Electrotechnical Commission.
IMS	Information management system.
ISO	International Standards Organization.
LOC	Loss of Containment.
NONC	Non critical.
NLP	Natural language processing.
PDS	Reliability of safety instrumented systems (In Norwegian).
PFD	Probability of failure on demand.
PM	Preventive maintenance.
PSA	Petroleum safety authority Norway.
SAS	Safety automation system.
SIF	Safety instrumented function.
SIL	Safety integrity level.
SIS	Safety instrumented system.
TLP	Technical language processing.

SIL (and PFD) performance in operation. The mentioned IEC standards require that PFD performance is estimated in the design phase and then subject to regular verification in the operational phase (Håbrekke et al., 2023; Lee et al., 2021). The PFD of a SIF is calculated from the DU failure rates and the proof test interval of the involved SIS equipment. Experience indicates that failure rates provided by manufacturers are generally lower than what is experienced in operation. Using too optimistic failure rates means that the SIF could seem more reliable than it is in reality. Therefore, the second edition of IEC 61511–1 (sub-clause 11.9.3) (IEC 61511, 2016) emphasizes the need to apply reliability data that are credible, traceable, documented, and justified. It is also stressed, that when applying reliability data collected during operation, the data must be based on feedback from similar devices used in similar operating environments.

To fulfil these requirements from IEC 61511–1 (IEC 61511, 2016), high quality in failure reporting and high precision in failure classification are essential (Ottermo et al., 2021). In particular, identification of all DU failures is critical as the DU failure rate is so influential on the PFD, and experience shows that manual checks are required to properly classify these failures. The adequacy of information in notifications registered into the maintenance system by technicians or process operators can vary. For instance, there is not always sufficient information to decide afterward if a failure mode is dangerous or safe and how the failure was revealed. Notifications often include a drop-down list for pre-defined failure modes and detection methods, from which the maintenance technician selects a failure mode and detection method, respectively. However, selections are not always correctly applied, and it is often necessary to investigate other sources such as event logs and condition monitoring systems or have discussions with technicians and equipment experts to be able to classify the reported failure. Notifications allow technicians and process technicians to add free (long) text, but analyzing this at a later stage is time-consuming. For example, the update of failure rates in the PDS data handbook required manual review of some thirty thousand notifications that corresponded to four man-hour years of review time (Ottermo et al., 2021). Failures classified during a manual review can also be subjective and vary from person to person, depending on the skills and background of the personnel involved. Hence, there is a strong motivation to explore the potential to replace some of the efforts with methods like Technical Language Processing (TLP) techniques.

TLP is a methodology that combines natural language processing (NLP) resources with engineering knowledge within the domain (Brundage et al., 2021; Gao et al., 2020). NLP refers to a research area within computer science to comprehend human-generated natural language in text or in speech. The need for TLP arises because the existing NLP pipelines are more suitable for non-engineering texts, and it is important to incorporate domain knowledge in NLP applications (Sharp et al., 2017), namely 'human-in-the-loop' and 'iterative' aspects affirmed by (Brundage et al., 2021). A typical example of engineering text is maintenance log for equipment. Previous studies have demonstrated how TLP techniques can be utilized for structuring of text data for obtaining information to fit the purposes like identifying underlying failure causes and estimates for mean time to failure (Sharp et al., 2017; Sexton et al., 2018; Bikaun and Hodkiewicz, 2021).

A contribution in the present paper is the application of a human-based and semi-automated failure classification for SIS which builds upon previous work (Ottermo et al., 2021) where the web-based annotation tool Redcoat (Stewart et al., 2019) was used for annotating free text fields in SIS notifications. The hypothesis was that the free text field includes vocabularies that can be associated with failure modes, and that these texts can be labelled (annotated) so that machines can use them. The words used to annotate the notifications in (Ottermo et al., 2021) were assembled into a dictionary with the purpose of annotating a larger dataset. This dictionary was then refined and tested for the semi-automated failure classification algorithm developed in (Lee et al., 2023). The main objective of this paper is to report the recent developments of the work in (Lee et al., 2023) and (Lee et al., 2022) and also to address how this approach relates to Industry 4.0 platforms for the seamless exchange of information and digital representations of assets. The paper contributes with examples of the groundwork needed to transform failure and maintenance data into machine-readable formats. This work has been part of an ongoing research and innovation project named APOS (Automated process for follow-up of safety instrumented systems) which has been a collaboration with 11 industry partners and the PDS forum, an organization having liaison status in IEC 61511.

The outline of the paper is as follows: Section 2 describes how failure reporting and classification for some relevant parameters are performed today. Section 3 presents the approach used in this work. In Section 4 application of the approach is described for two cases. Section 5 investigates the relevance of this work in the digitalization context, while Section 6 provides a discussion and suggestions for further work. Finally, Section 7 presents some concluding remarks.

2. Failure reporting and classification

Failures revealed during operation, testing, and maintenance are usually reported and classified in a Computerized Maintenance Management System (CMMS) by a technician. Each failure is registered as a notification with a unique notification ID that is linked to the tag number of the failed equipment. The notification consists of a combination of fixed data fields, drop-down fields, and free text fields (i.e. short text and long text), as shown in Table 1. The drop-down fields provide a list of alternatives for detection methods and failure modes that the technician can choose from.

The **detection method** characterizes how the failure was discovered. If the failure was revealed immediately (i.e. alarmed upon occurrence), it is classified as *detected.* On the other hand, if the failure was revealed during a functional test, periodic maintenance, causal observation, or demand (i.e. latent until test or demand), it is classified as *undetected.*

The **failure mode** characterizes how it was observed that a function has been fully or partially lost. For instance, for a valve, the safety function will typically be to open or close on demand, and to keep tight in the closed position. Hence, a failure to close or open on demand or an internal leakage in the closed position are both examples of failure modes. By combining the failure mode with information about the equipment group and its related safety function, it can be determined whether the failure is

Table 1
Detection method taxonomy for shutdown valves.

Tag	Tag description	Loca-tion	Short text	Problem code text	Failure mode	Cause code	Long text	Detection method
xx.yyy.zz -##-####	Import pipeline A	xxxx	Partial stroke test not OK	Valve failure to function on demand	FTC	Design related	* We ran a partial stroke on 10 valves last night. 8 of these failed! * We ran all the valves at least 2 times according to procedure, still the test is not approved. *Some valves were run many times while field operator observed the valve. At first the valve did not engage all within the stipulated time for partial stroke. After several try it started to go a few percent. We assume that this problem is due to cold hydraulic oil, as we have had a long period of cold weather. The temperature was -6 degrees when the test was run.	Preventive maintenance

dangerous or *safe*. Hence, correct reporting of failure modes is essential for determining the severity of the failure. The combination of the detection method and failure mode is the minimum information required to determine the failure class (Hauge et al., 2023a). CMMS systems used by operating companies in Norway commonly implement drop-down lists according to the standardization of failure modes and detection methods in ISO 14224 (ISO 14224, 2016). ISO 14224 provides a basis for the collection of reliability data for oil and gas equipment including equipment groups and lists of possible detection methods, failure modes, failure mechanisms, and failure causes for the equipment groups. The technicians that write failure notifications are not always able to select the right alternative, particularly for the failure mode, as there are often a lot of possible choices. In that case, the technician can select the ’other’ or ’unknown’ categories in the drop-down menu and provide additional explanations in the free text field. A manual review of notifications from a Norwegian offshore facility showed that more than 50% of notifications were registered as ‘other’ or ‘unknown’. This implies that the most relevant information about failure mode and detection method is often found in the free text field(s) in the notifications (Ottermo et al., 2021).

As mentioned above, an important foundation for enabling the collection of high-quality failure data is to ensure that the failures are reported consistently with a high level of precision about detection methods and failure modes. This requires standardized taxonomies for equipment hierarchies in conjunction with standardized taxonomies for failure parameters (e.g. failure mode hierarchy) since this will ensure that failure data for similar equipment can be collected and aggregated One of the goals of the APOS project has therefore been to align common practices from operating companies in the Norwegian oil and gas industry, both for equipment hierarchies, and the taxonomies for detection methods and failure modes, respectively (Hauge et al., 2023a). We define an equipment group as a collection of equipment types with some common characteristics, such as comparable functionality, design, failure rate, etc. Examples of equipment groups are smoke detectors or shutdown valves (Hauge et al., 2023a). The suggested detection method taxonomy for shutdown valves, which has been modified from ISO 14224 (ISO 14224, 2016) is shown in Table 2. The corresponding ISO 14224 categories (from Table B.4 in the standard) are listed in the rightmost column.

The suggested failure mode taxonomy is equipment group specific because equipment fails in different ways. In addition, the criticality of the failure mode will often depend on the functionality of the equipment. For some equipment groups, some subgroups of components can have different functional requirements. For instance, some shutdown valves have leakage requirements, which implies that leakage in closed position (LCP) is defined as ’dangerous’. However, for a shutdown valve with no leakage requirement, a leakage in closed position is classified as ’degraded’ or ’non-critical’. The suggested failure mode taxonomy for shutdown valves is shown in Table 3 (Hauge et al., 2023a).

3. Semi-automated approach to failure classification

An approach was suggested to analyze a set of free texts for a specified SIS equipment group, emergency shutdown (ESD) valve. The main aim was to identify DU failures from the dataset. Considering that the initial classification made by different technicians (using drop-down list) can be subjective, a secondary means for classifying failures using TLP could be useful for comparison and verification. A rule-based method was used to annotate each free text with detection methods

Table 3
Failure mode taxonomy for shutdown valves.

Failure mode class	Failure mode
Dangerous failure	Fail to close (FTC)
	Delayed operation (DOP)
	Leakage in closed position (LCP)
Safe failure	Fail to open (FTO)
Spurious failure	Spurious operation (SPO)
Degraded failure	Delayed operation (DOP)
	Leakage in closed position (LCP)
	Structural deficiency (STD)
No-effect failure	Noise (NOI)
	Abnormal instrument reading (AIR)
	Minor in-service problems (SER)
Loss of containment (LOC)	External leakage – utility medium (ELU)
	External leakage – process medium (ELP)
Loss of explosion (EX) protection	Loss of EX protection (LEX)

Table 2
Detection method taxonomy for shutdown valves.

Detection method class	Detection method	Corresponding ISO 14224 categories
1. Undetected	1.1 Functional test	02 Functional testing
	1.2 Other periodic maintenance (PM) activity	01 PM
		03 Inspection
		04 Periodic condition monitoring
	1.3 Demand	07 Production interference
		10 On demand
	1.4 Casual observation	05 Pressure testing
		08 Causal observation
		09 Corrective maintenance
2. Detected	2.1 Diagnosed / immediately detected event	06 Continuous condition monitoring

and failure modes. Then, the notifications annotated with dangerous failure mode(s) and undetected detection method(s) were classified as DU failures. The dictionary deployed for the annotation consisted of keywords indexed by a failure mode or a detection method. The dictionary was manually created based on human-based knowledge and manual checks, while the process of extracting information from the text was automated without involving human annotators. Hence, the suggested approach can be considered semi-automatic.

The text annotation procedure resembles Named Entity Recognition (NER), which is one of the important tasks in NLP for various applications like information extraction and machine translation (Zitouni, 2014). When using NER in NLP, an entity is a representative category of nouns in text, for example, person and location, which appear frequently in natural language texts. A named entity represents an instance (i.e. a single or a chunk of words) that can be annotated with an entity, for example, 'Jane' for the entity 'Person', and 'New York' for the entity 'city ' (Mikheev et al., 1999; Nadeau and Sekine, 2007). However, NER for domain-specific texts is demanding due to the required resources like lexicon and training of the models (Tomori et al., 2016; Zitouni, 2014). In the domain of electronic health records data, NER deploys commonly used existing ontologies like Medical Subject Heading and Unified Medical Language System (Govindarajan et al., 2023; Jimeno et al., 2008; Kundeti et al., 2016). In the domain of engineering data, in particular maintenance work orders, TLP tools and models have been developed for numerical representation of texts, pre-processing, and the NER models based on machine learning (Bikaun and Hodkiewicz, 2021; Gao et al., 2020; Naqvi et al., 2022; Stewart et al., 2022).

As mentioned above, a NER-*like* process was suggested to annotate the text with a hierarchical structure of entities. Entity hierarchies have been adopted in several NLP and TLP studies to better reflect domain knowledge (Stewart et al., 2019; Murty et al., 2017). The entity structure for ESD valves was established as illustrated in Fig. 1, and it was aligned with the APOS project taxonomies for shutdown valves shown in Table 2 and Table 3. The failure modes and the detection methods represent entities at the lowest level in the hierarchy (in the yellow box in Fig. 1). Each failure mode (e.g. FTC) belonged to a failure class (e.g. Dangerous) and each detection method (e.g. Functional test) belong to a detection method class (e.g. Undetected). The failure mode entities were specific for the ESD valves, while detection methods are common for all equipment groups.

In the suggested approach, it was possible to annotate the keywords with multiple entities, a feature that is also offered by existing annotation tools like Redcoat and BRAT (Ohta et al., 2011; Stewart et al., 2019). The possibility of multiple labels is particularly important for SIS notification data because a long free text may include both words implying dangerous failure modes and other less critical failure modes. For this reason, the approach was to annotate with all possible failure modes and then select the most critical one (Step 5). This approach was inspired by (Sexton et al., 2018), where maintenance work orders were tagged with multiple equipment labels with the intention to enable estimation of mean time to failure (MTTF). In this study, it was demonstrated that labelling with a single representative system will result in under-estimation of MTTF, because the failure in the other parts attached to the system will not be explicitly tagged and therefore discarded, although this failure can in fact indicate the system failure.

The data applied to the semi-automated approach was compiled from notifications exported into Excel from the maintenance management system of several oil and gas process plants in Norway. Depending

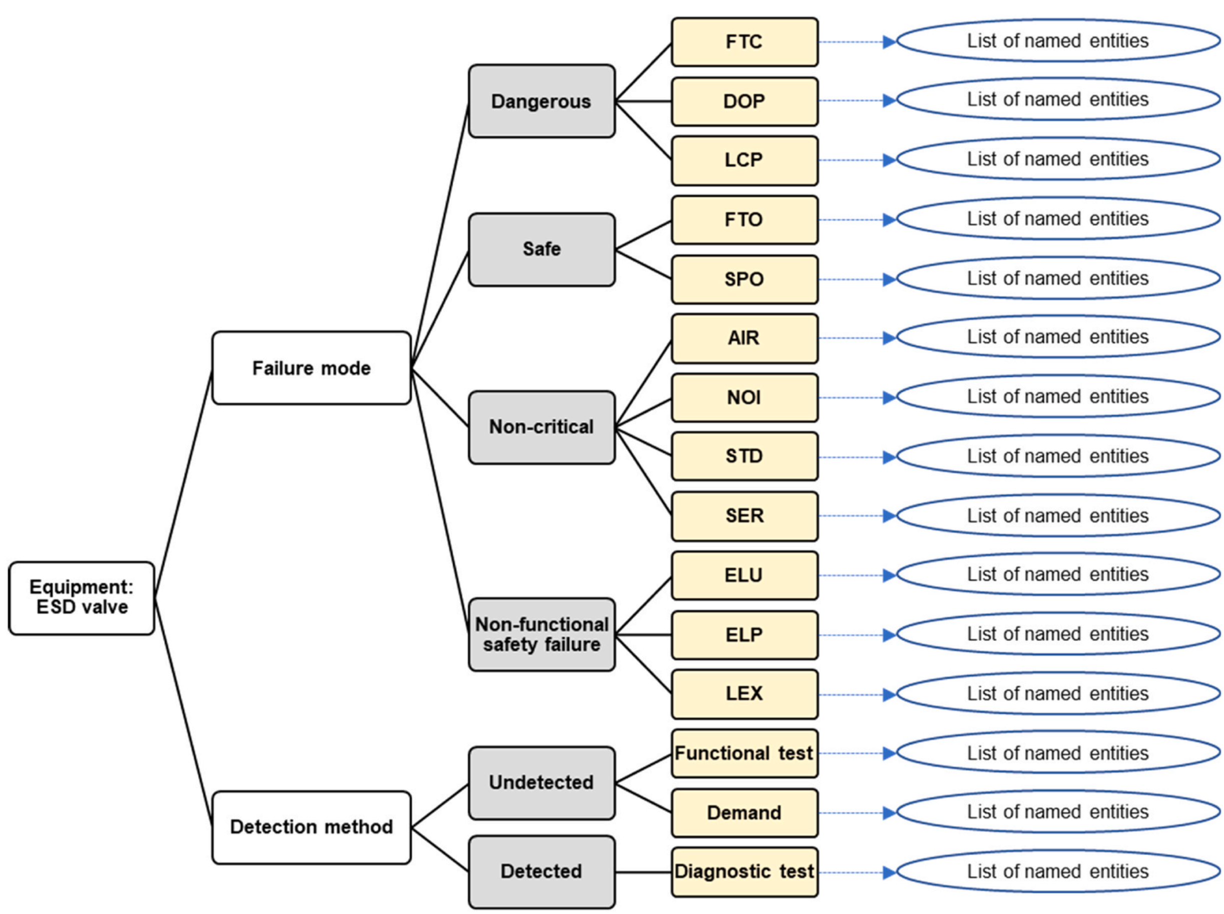

Fig. 1. The entity categories defined for ESD valves. The entities (in the yellow box) represent the failure modes and the detection methods. Each entity belongs to the entity category one level above (in the grey box), for example, FTC (failure mode) belongs to the Dangerous (failure mode class). Each entity is associated with the list of named entities or keywords

on the operating company, the formats of notifications varied slightly, but it was possible to route the information correctly into the predefined set of data fields. An example of selected data fields is presented in Table 1. Data fields like “Tag” are fixed (static) data, while fields like “Cause code”, “Failure mode”, and “Problem code text” have been selected from drop-down lists. For example, in the data field “failure mode”, the technicians have selected the failure mode that seems most relevant for the observed failure.

In this work, the commonly used safety equipment ESD valve, was selected as a representative equipment group. The most critical failure mode for ESD valves is FTC, which denotes the inability of the valve to close on command. The suggested approach consists of five main steps, as illustrated in Fig. 2.

Step 1 Pre-process the notification text.

Many of the notification texts for ESD valves include words and phrases describing the observed valve movement during testing, which can be pieces of information about detection methods and failure modes that are needed for classifying failures. Most of the free texts in the datasets were written in Norwegian, and a few of them were written in other languages (i.e., English, Swedish, Danish).

An example of a sentence in a free text for ESD valve is ‘Ventilen var treg og brukte for lang tid på lukking og gikk derfor i feil’ (English: The valve was stuck and used too long time to close, and thus failed). The word ‘lukking’ (English: closing) in this sentence is one keyword that describes the valve movement. ‘Lukking’ (English: closing) is a variant of the basic form ‘lukke’ (English: close) with the suffix ‘ing’. Other forms of this word can be e.g., ‘lukker’ (English: is closing), ‘lukket’ (English: closed) and ‘har lukket’ (English: has closed). Typical suffixes to the verbs such as ‘ing’, ‘er’, ‘et’ (English: clos) have been removed, while keeping ‘lukk’ (Lane et al., 2019). In addition, nouns and adjectives that have the same root word were reduced to their stem, for example, ‘lekkasje’ (English: leakage), ‘lekker’ (English: leaking) can be steamed to ‘lekk’ (English: leak). As there were relatively few words that needed stemming, it was manageable using a simple search-replace in Excel, instead of deploying existing NLP algorithms like Snowball which could have been used since it does support Norwegian language (Snowball, 2022).

Step 2 List keywords for failure mode and detection method entities.

As shown in Fig. 1, each entity was associated with a list of named entities or keywords. For instance, FTC had a list of keywords including ’Not closing’, ’Fail to close’, or ’Problem with closing’. For the entity ‘functional test’, the associated keywords were ’partial stroke testing’ and ’periodic function test’. Such lists were created from APOS project taxonomy and expert knowledge from the experience of manual reviews over several decades. Fig. 3 illustrates the process of establishing dictionaries that can be updated and refined whenever necessary. Non-exhaustive examples of named entities for ESD valves are shown in Table 4.

Step 3 Post-processing of dictionary.

This step was necessary to modify the dictionaries made in Step 2 to enable more efficient NER-*like* process by reducing the number of entities included in the dictionary. As seen in Table 4, named entities can be written in various ways, for instance, ‘fail to close’, ‘failure with closing’, and ‘closing failure’. This means that the dictionary will be large if we include all such variations. For this reason, we can cut a sentence into word units and words such as ‘with’, ‘in’, ‘to’ that are typically identified as ‘stop words’ in NLP that commonly appear in a language but are of little use in text analysis (Cambria and White, 2014; Lane et al., 2019; Riadsolh et al., 2020). For example, ‘fail to open’ can be split into {fail, to, open}, which can be reduced to {failed, open} by removing ‘to’.

In addition, it was quickly realized that the description for many failure modes of ESD valves include negative words like “not”, “problem”, and “fail”, together with a verb describing the function of the equipment like “open”, “close”, and “go”. Common negative words can be placed both before and after the verb, e.g., “failed to close” or “closing failed”, which results in many possible combinations of two words “fail” (negative word) and “close” (verb). For this reason, negative words were searched first, and then the verbs and nouns were looked up. In this way, the number of named entities could be reduced. The list of named entities for ‘FTC’ that are split into two parts is shown in Table 5.

Step 4 Annotate notifications with entities for failure modes and detection method.

A simple algorithm using an Excel formula with built-in index and match functions was used to annotate the free text with the predefined entities, as illustrated in Fig. 4.

Step 5 Classify failures by tagging entity categories.

Once failure modes were annotated in Step 4, the corresponding failure mode class(es) were tagged according to the entity hierarchy in Fig. 1. In the same way, the detection method class was labeled. By using the tagged entities, the failure class can be determined, as illustrated in Fig. 5.

In case the long text is annotated with more than one entity for failure modes, the algorithm using if-then statements was used to choose the most safety-critical failure mode according to the criteria in (Hauge et al., 2023a), as illustrated in Fig. 6. This task was added to avoid that dangerous failure modes not ruled out due to non-dangerous failures. For example, a notification text from functional testing includes ‘The valve is not closing. At the closing signal, the valve fails and stays put for approx. 20% open. May seem like a mechanical fault. ’. In this example, words ‘not’ and ‘closing’ are included in one sentence, and ‘open’ in another sentence. The text was tagged with ‘FTC’ due to the combination of ‘not’ and ‘closing’ in Step 4. In addition, ‘FTO’ was also annotated because ‘not’ and ‘open’ are used. In such a case, the algorithm can correctly select ‘FTC’ which is the most safety-critical (severe) failure mode.

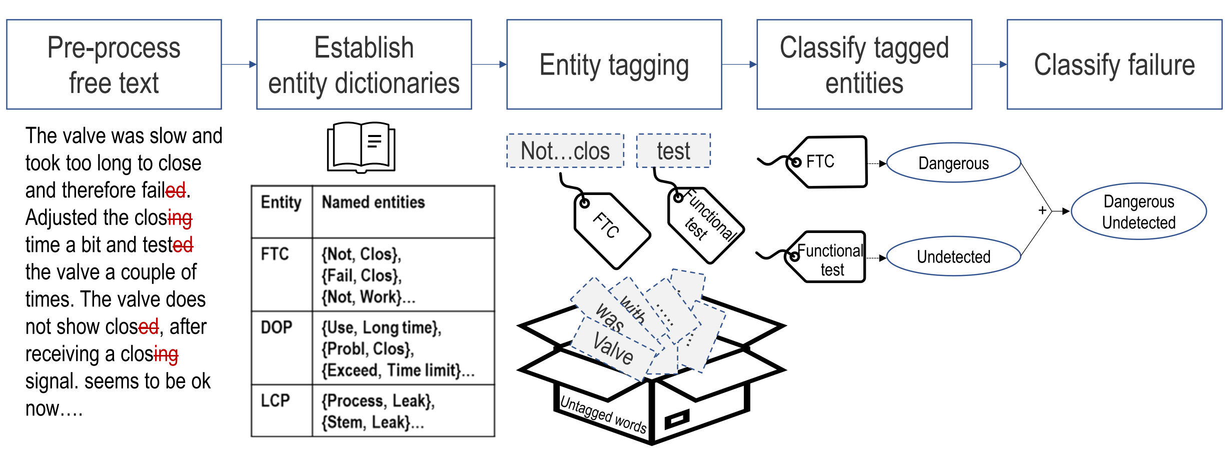

Fig. 2. A simplified illustration of the suggested approach applied for ESD valves.

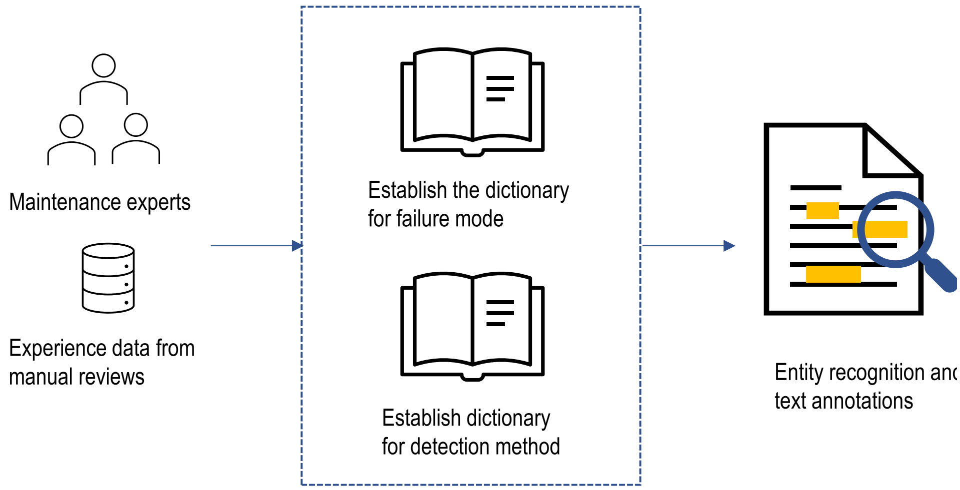

Fig. 3. Process of establishing the dictionaries for failure modes and detection methods.

Table 4
An example of named entities for failure modes and detection methods of ESD valves.

Entity category	Entity	Named entities (Example)
Failure mode	FTC	Not closing, Fail to close, Not moving, Didn't function, Problem with closing, Difficult to close, Not go, Not working, Didn't work, Failed to move
	FTO	Not open, Not opened, Fail to open, Problem with opening
	DOP	Used long time, Exceeded time limit, Long closing time
	ELP	Leaks, Leaking
Detection method	Functional test	Proof test, partial stroke test, partial stroke
	Observation	Random observation, Unplanned walk around
	Demand	With demand, Reset after shutdown
	Diagnose	Alarm, Diagnostic Alarm, Control room

Table 5
List of named entities (in Norwegian) for entity 'FTC'.

Entity	FTC	
	Step 1 (Negative word)	Step 2 (Verb or Noun)
Named entities	Ikk	lukk
	Feil	naar
	Vansk	steng
	Treg	virk
	Lang	roer
	Mellom	kjoe
	Skad	gaa
	Blokk	gikk
	Gnisning	bytt
	Tett	bruk
	Fast	Staa
	….	sitt
		stengetid
		gangtid

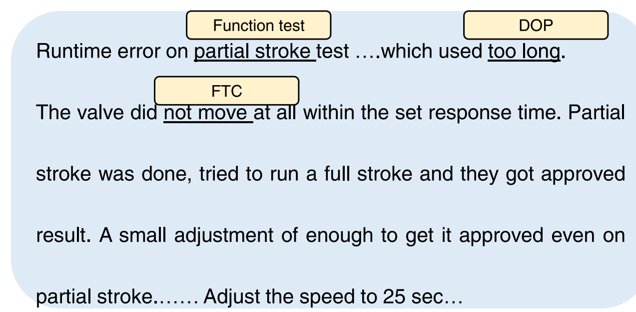

Fig. 4. Text annotation with failure mode entity and detection method entity.

4. Case study

The suggested approach was applied to two datasets. The first dataset consisted of 96 notifications for ESD valves from an onshore plant, and the second dataset was 50 notifications for ESD valves obtained from an onshore plant. The dictionary used for the text annotation was updated from previous work (Lee et al., 2022). The algorithm described in Section 3 was applied for text annotation and failure classification.

For the first dataset, the manual review (correct classification) gave the following distribution: 19 DU failures, 15 safe failures, 40 non-applicable (NA) failures, while 22 were not classified. On the other hand, the results from the semi-automated classification were as follows: 37 DU failures, 4 Dangerous detected (DD) failures, 1 Non-critical (NONC) failures, 10 Non-functional safety failures, 13 safe failures, 0 NA failures, and 31 that were not classified. Fig. 7 presents the distribution of failure modes obtained from manual review and the semi-automated approach, respectively.

To verify the functionality of the algorithm, it was checked if the 19 DU failures identified in the manual review were included in the 37 DU failures from the semi-automated classification. 15 out of the 19 DU failures were classified as DU failures, and four notifications that were unmatched were manually checked afterward. Three of the four unmatched notifications were written in English. As the dictionary for annotation was in Norwegian, English words in the text cannot be tagged with any entity. The last mismatched notification was classified as DD failure by the algorithm, while the correct failure class was DU. The notification was annotated with FTC, FTO, SPO, and SER, and conservatively classified with FTC (dangerous failure mode). However, the notification was annotated with 'Detected' due to the entity 'Alarm' that was included in the free text, but in reality the failure was detected during a 'Demand'. The occurrence of this demand was not found in the free text field but was registered into the detection method field, which means that if information from the detection method field had been combined with the free text, the DU failure would have been correctly classified.

The second dataset consisted of 50 notifications for ESD valves registered from functional tests. The same dictionary used for the first dataset (case 1) was used for the second dataset. The manual review (correct classification) gave the following distribution of the 50

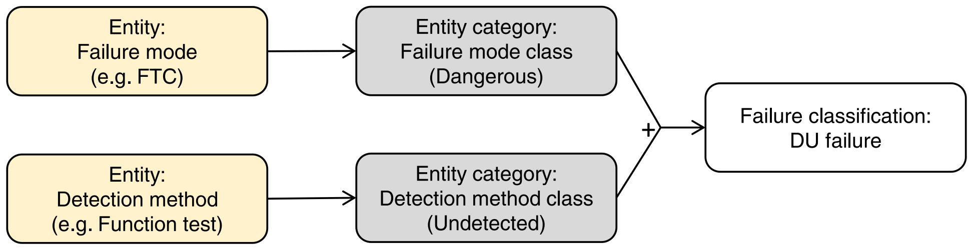

Fig. 5. Failure classification by using annotated entities.

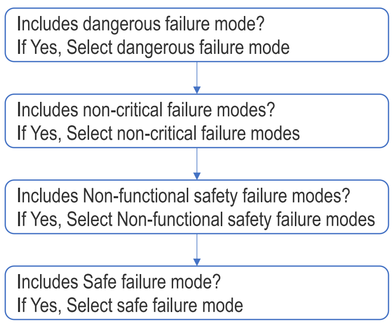

Fig. 6. Selection of failure mode class in case of multiple entity tags.

notifications: 15 DU failures, 16 safe failures, 18 non-applicable (NA) failures, while 1 failure was not classified. On the other hand, the results from the semi-automated classification were as follows: 20 DU failures, 6 Non-functional safety failures, 17 safe failures, and 7 that were not classified. Fig. 8 presents the distribution of failure modes obtained from the manual review and the semi-automated approach, respectively.

The 15 DU failures from the manual review were compared with the 20 DUs identified from the algorithm, which showed 13 DU matches. The 2 unmatched notifications were manually checked. The first one had the blank free text field. The second notification was annotated as FTO and classified as safe failure, while it was classified as safe or DU in the manual review.

5. Relevance to the Industry 4.0 context

Failure classification is preceded by the collection of failure data, and as of today, such data are usually found in the CMMS. However, relevant information can also be collected from other data sources, such as the safety and automation system (SAS), condition monitoring systems, and the information management system (IMS). The reason for this is that many notifications in CMMS may not necessarily provide all the information needed for the classification. For example, the response time for a valve is a type of information that may not be available in the CMMS notification but could be useful for classification. Today, the process of collecting information from different sources is time-consuming and one way to overcome this limitation is to enable interoperability between different source systems. A solution to achieve this is to take advantage of industry 4.0 technologies that enable the digital representation of assets. This means that digital twins for the SIF and the associated SIF components can be established such that failure data needed for the failure classification are reported in interoperable formats, as opposed to the different formats as per today (Hauge et al., 2023b). This will facilitate the aggregation of all the relevant data for automatic failure classification by a TLP-based tool, as illustrated in Fig. 9.

We have already produced a guideline on how to report and classify failure and maintenance data for safety equipment. In this guideline, standardized equipment grouping, equipment properties, and simplified failure taxonomies are suggested. For future digitalization, the guideline also identifies standardized equipment properties and associated property values to enable the establishment of a complete information model for functional safety. The suggested taxonomies and properties have therefore been compared and mapped against recognized standards and relevant electronic equipment libraries (Hauge et al., 2023a). This work will also be pursued in our new research project, APOS 2.0, which will develop and test new digital solutions for functional safety pilots that facilitate sharing of information between systems and companies.

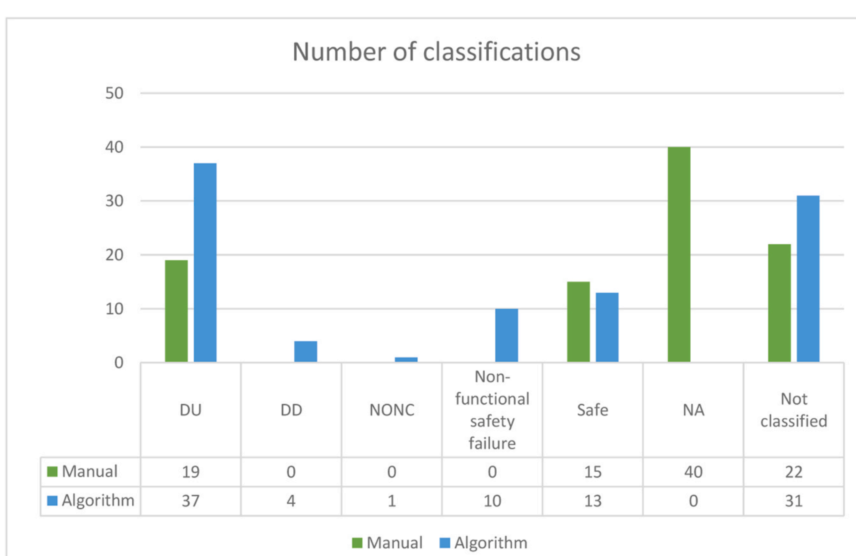

Fig. 7. Number of classified failures for 96 notifications from an onshore plant.

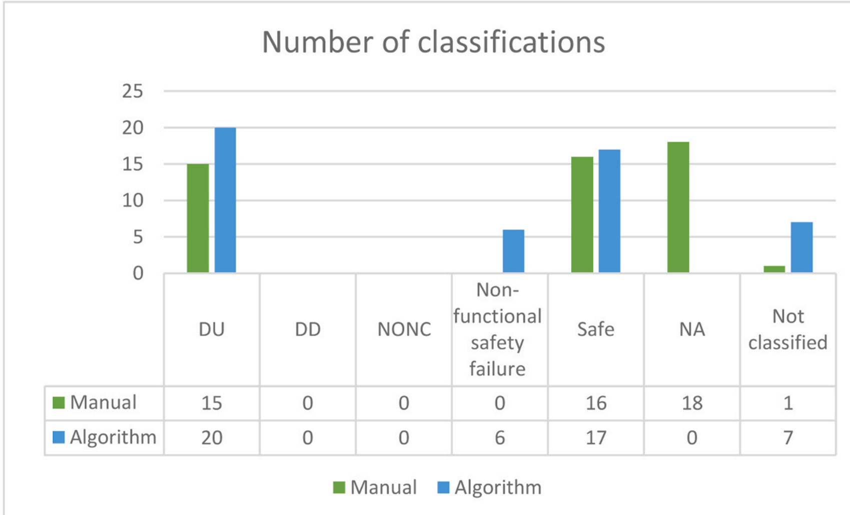

Fig. 8. Number of classified failures for 50 notifications from an offshore plant.

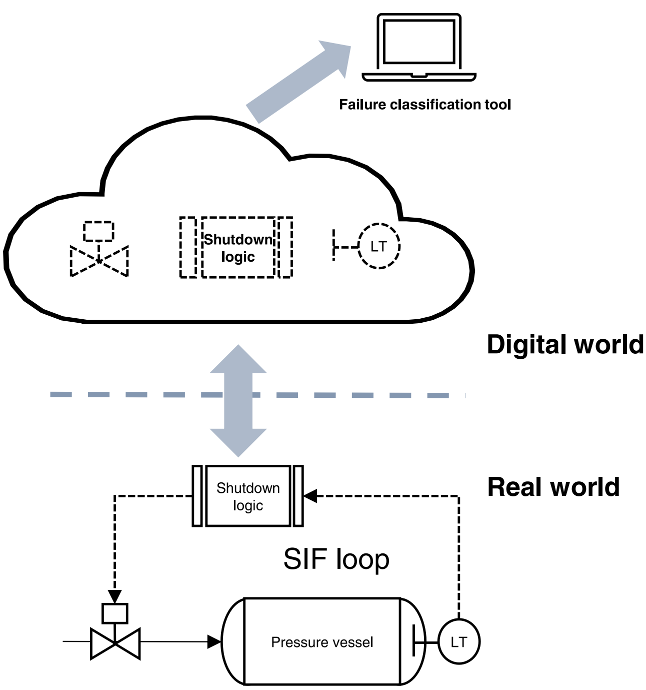

Fig. 9. Automatic failure classification facilitated by SIF digital twins.

6. Discussion and further work

The two case studies in Section 4 show that the semi-automated approach sufficiently support the classification of DU failures. However, the results from the two datasets indicate that the suggested approach gives very conservative results. Especially for the first dataset, the automated approach gives about twice as many DU failures (37 DUs) as the actual DUs identified from the manual (operational) review (19 DUs). One reason is that the suggested approach is limited to extract keywords about failure mode and detection method and is not capable of interpreting the failure described in the text using context information. For example, a free text field including ‘Valve runs full stroke on partial stroke. I believe that the problem here is that the valve closes too quickly, not partial stroke times. Partial stroke time is currently set to 15 s′ notification was classified as ‘safe’ from the manual review. However, the algorithm, by recognizing the combination of ‘problem’ and ‘partial stroke’, tagged the text with FTC and DOP and then classified with dangerous failure mode. However, the sentence ‘Valve runs full stroke on partial stroke’ indicates the valve was functioning as required, while the algorithm was not capable of comprehending the full sentence. Another example is the text ‘Valve closed in 3 s, should use 16 s. Adjusted the run time to 16 s′ This was classified as ‘safe’ according to the operational review. In the semi-automated approach, this text is not tagged with any failure mode entity, because there are no negative words like ‘not’ or ‘problem’ that can be combined with the word ‘close’. For this reason, it is sometimes necessary to understand the sentences (context) rather than extracting keywords, which implies that rule-based entity tagging may not be sufficient on its own, and that it may be necessary to apply deep learning based model for text annotation (Stewart et al., 2022; Usuga-Cadavid et al., 2022).

7. Conclusions

The suggested approach represents a semi-automated process for failure classification, which can supplement the manual classification of notifications from oil and gas plants. The case study demonstrates that the approach is well suited to filter out possible DU failures, and can therefore be used to verify that all the DU failures are identified. Another advantage is that the approach is transparent in a way that the practitioners are able to understand the process of semi-automated classification and also contribute to creating and updating the dictionaries, which is good for the communication aspect. Moreover, the case study showed that the semi-automated failure classification using the APOS project taxonomies matched sufficiently with manual classification based on the reviews. This may showcase that the failure modes and detection method taxonomies for each equipment group in the APOS project, which are also aligned with international standards like ISO 14224, are comprehensive enough to reflect the reality for SIS. Hence, APOS taxonomies represent groundwork that is needed to train TLP tools to interpret SIS component (equipment) failures. Today, the process of collecting failure data and information from different sources is time-consuming and the paper briefly discusses how we can take advantage of the Industry 4.0 framework to enable interoperability between different source systems.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

The paper presents results from APOS, a joint industry project on automated processes for follow-up of safety instrumented systems. The project is supported by the Norwegian Research Council (Project no. 295902), 11 industry partners representing oil companies, engineering, consultants, and vendors of control and safety systems and PDS forum.

References

Bikaun, T., Hodkiewicz, M., 2021. Semi-automated estimation of reliability measures from maintenancework order records. no. 1, Art. no. 1 PHM Soc. Eur. Conf. vol. 6. https://doi.org/10.36001/phme.2021.v6i1.2950.

Brundage, M.P., Sexton, T., Hodkiewicz, M., Dima, A., Lukens, S., 2021. Technical language processing: unlocking maintenance knowledge. Manuf. Lett. vol. 27, 42–46.

Cambria, E., White, B., 2014. 'Jumping NLP curves: a review of natural language processing research [Review Article]. IEEE Comput. Intell. Mag. vol. 9 (2), 48–57. https://doi.org/10.1109/MCI.2014.2307227.

Gao, Y., Woods, C., Liu, W., French, T., Hodkiewicz, M., 2020. Pipeline for machine reading of unstructured maintenance work order records. Proceedings of the 30th European Safety and Reliability Conference and 15th Probabilistic Safety Assessment and Management Conference (ESREL).

Govindarajan, S., Mustafa, M.A., Kiyosov, S., Duong, N.D., Naga Raju, M., Gola, K.K., 2023. An optimization based feature extraction and machine learning techniques for named entity identification. Optik vol. 272, 170348. https://doi.org/10.1016/j.ijleo.2022.170348.

Håbrekke, S., Hauge, S., Lundteigen, M.A., 2023. Guideline for follow-up of Safety Instrumented Systems (SIS) in the operating phase, Ed. 3 (APOS H3)', SINTEF, SINTEF Rep. 00107, 2023.

Hauge, S., Håbrekke, S., Lundteigen, M.A., Lee, S., Ottermo, M.V., 2023a. Guidelines for standardised failure reporting and classification of safety equipment failures in the petroleum industry, Ed. 1 (open version) (APOS H1). SINTEF Rep. 00108, 2023.

Hauge, S., Lundteigen, M.A., Ottermo, M.V., Lee, S., Petersen, S., 2023b. Information model for functional safety (APOS H5). SINTEF Rep. 00109, 2023.

IEC 61511, 2016. Functional safety - Safety instrumented systems for the process industry sector. International Electrotechnical Commission, Geneva.

IEC 61508, 2010. Functional safety of electrical/electronic/programmable electronic safety-related systems. International Electrotechnical Commission, Geneva.

ISO 14224, 2016. Petroleum, petrochemical and natural gas industries: Collection and exchange of reliability and maintenance data for equipment. International Organization for Standardization, Geneva.

Jimeno, A., Jimenez-Ruiz, E., Lee, V., Gaudan, S., Berlanga, R., Rebholz-Schuhmann, D., 2008. Assessment of disease named entity recognition on a corpus of annotated sentences. BMC Bioinforma. vol. 9 (3), S3. https://doi.org/10.1186/1471-2105-9-S3-S3.

Kundeti, S.R., Vijayananda, J., Mujjiga, S., Kalyan, M., 2016. Clinical named entity recognition: challenges and opportunities. 2016 IEEE Int. Conf. Big Data (Big Data) 1937–1945. https://doi.org/10.1109/BigData.2016.7840814.

Lane, L., Howard, C., Hapke, H.M., 2019. Natural language processing in action. Mann Publ.

Lee, S., Lundteigen, M.A., Paltrinieri, N., 2021. Dynamic risk analysis from the perspective of life cycle approach in Iec 61508 and Iec 61511. Chem. Eng. Trans. vol. 86, 265–270. https://doi.org/10.3303/CET2186045.

Lee, S., Ottermo, M.V., Hauge, S., Lundteigen, M.A., 2022. Automated classification of failure data for safety equipment for follow-up in the operations phase. Presented in 25th MKOPSC Int. Process Saf. Symp.

Lee, S., Ottermo, M.V., Hauge, S., Håbrekke, S., Lundteigen, M.A., 2023. Potential for automated follow-up of safety equipment. SINTEF Rep. 00110, 2023.

Mikheev, A., Moens, M., Grover, C., 1999. Named Entity Recognition without Gazetteers. Ninth Conference of the European Chapter of the Association for Computational Linguistics, Bergen, Norway. Association for Computational Linguistics, pp. 1–8. Accessed: Jun. 30, 2023. [Online]. Available: https://aclanthology.org/E99-1001.

Murty, S., Verga, P., Vilnis, L. and McCallum, A., 2017. Finer grained entity typing with typenet. arXiv preprint arXiv:1711.05795.

Nadeau, D., Sekine, S., 2007, 91634 A Surv. Named Entity Recognit. Classif. vol. 30 (1), 3–26. https://doi.org/10.1075/li.30.1.03nad.

Naqvi, S.M.R., Ghufran, M., Meraghni, S., Varnier, C., Nicod, J.-M., Zerhouni, N., 2022. Generating semantic matches between maintenance work orders for diagnostic decision support. Annu. Conf. PHM Soc. vol. 14 (1) https://doi.org/10.36001/phmconf.2022.v14i1.3241.

Ohta, T., Pyysalo, S., Tsujii, J., 2011. Overview of the Epigenetics and Post-translational Modifications (EPI) task of BioNLP Shared Task 2011. Proceedings of BioNLP Shared Task 2011 Workshop. Portland, Oregon, USA: Association for Computational Linguistics,, pp. 16–25. Accessed: Jul. 04, 2023. [Online]. Available: https://aclanthology.org/W11-1803.

Ottermo, M.V., Hauge, S., Håbrekke, S., 2021. Reliability data for safety equipment. PDS data Handb.

Ottermo, M.V., Håbrekke, S., Hauge, S., Bodsberg, S., Technical, L., 2021. Language processing for efficient classification of failure events for safety critical equipment. Presente PHM Soc. Eur. Conf. https://doi.org/10.36001/phme.2021.v6i1.2792.

Riadsolh, A., Lasri, I., ElBelkacemi, M., 2020. Cloud-based sentiment analysis for measuring customer satisfaction in the Moroccan banking sector using naïve bayes and stanford NLP. vol. Vol. 14 J. Autom. Mob. Robot. Intell. Syst. (No. 4). https://doi.org/10.14313/JAMRIS/4-2020/47.

Sexton, T., Hodkiewicz, M., Brundage, M.P., Smoker, T., 2018. Benchmarking for keyword extraction methodologies in maintenance work orders. PHM_CONF vol. 10 (1). https://doi.org/10.36001/phmconf.2018.v10i1.541.

Sharp, M., Sexton, T., Brundage, M.P., 2017. 'Toward Semi-autonomous Information', in Advances in Production Management Systems. The Path to Intelligent, Collaborative and Sustainable Manufacturing. In: Lödding, H., Riedel, R., Thoben, K.-D., von Cieminski, G., Kiritsis, D. (Eds.), in IFIP Advances in Information and Communication Technology. Springer International Publishing, Cham, pp. 425–432. https://doi.org/10.1007/978-3-319-66923-6_50.

Snowball. https://snowballstem.org/. Accessed 14 December 2022.

Stewart, M., Liu, W., Cardell-Oliver, R., 2019. 'Redcoat: A Collaborative Annotation Tool for Hierarchical Entity Typing', in Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP). System Demonstrations. Association for Computational Linguistics, Hong Kong, China, pp. 193–198. https://doi.org/10.18653/v1/D19-3033.

Stewart, M., Hodkiewicz, M., Liu, W., French, T., 2022. MWO2KG and Echidna: constructing and exploring knowledge graphs from maintenance data, 1748006×221131128 Proc. Inst. Mech. Eng., Part O: J. Risk Reliab.. https://doi.org/10.1177/1748006×221131128.

Tomori, S., Ninomiya, T., Mori, S., 2016. Domain Specific Named Entity Recognition Referring to the Real World by Deep Neural Networks. In: Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics, Volume 2. Association for Computational Linguistics, Berlin, Germany, pp. 236–242. https://doi.org/10.18653/v1/P16-2039.

Usuga-Cadavid, J.P., Lamouri, S., Grabot, B., Fortin, A., 2022. Using deep learning to value free-form text data for predictive maintenance. Int. J. Prod. Res. vol. 60 (14), 4548–4575. https://doi.org/10.1080/00207543.2021.1951868.

Zitouni, I. (Ed.), 2014. Natural Language Processing of Semitic Languages. in Theory and Applications of Natural Language Processing. Berlin, Heidelberg: Springer. 10.1007/978-3-642-45358-8.